DIFFERENTIAL GAME-BASED CONTROL METHODS FOR UNCERTAIN CONTINUOUS-TIME NONLINEAR SYSTEMS

By

MARCUS A. JOHNSON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To my wife, *Mirna*, my parents, *Carolyn* and *Keith Johnson*, and my sisters, *Christine* and *Michele*, for their unwavering support and constant encouragement

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor, Warren E. Dixon, whose experience and motivation have been instrumental in the successful completion of my PhD. The guidance and the patience he has shown over the years have helped me mature in my research and as a professional. I would also like to extend my gratitude to my committee members Norman Fitz-Coy, Prabir Barooah, and Pramod Khargonekar for the time and help they provided. I would like to thank my wife for her love and patience. Also, I would like to thank my family for believing in me.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DIFFERENTIAL GAME-BASED CONTROL METHODS FOR UNCERTAIN CONTINUOUS-TIME NONLINEAR SYSTEMS

By

Marcus A. Johnson

August 2011

Chair: Warren E. Dixon
Major: Aerospace Engineering

Game theory methods have been instrumental in the advancement of various disciplines such as social science, economics, biology, and engineering. The focus of this dissertation is to develop techniques for approximating solutions to zero-sum and nonzero-sum noncooperative differential games and using these solutions to stabilize some classes of parametrically uncertain and disturbed nonlinear systems. One contribution of this work is the development of a robust (sub)optimal controller that stabilizes an uncertain Euler-Lagrange system with additive disturbances and yields a solution to the feedback Nash differential game. The control formulation utilizes the Robust Integral Sign of the Error (RISE) control technique to asymptotically identify nonlinearities in the dynamics and converge to a residual linearized system, then the solution to the Nash game is used to derive the stabilizing feedback control laws.

Furthermore the Nash optimal control technique is improved when one player has additional information about the other player, in the case of the Stackelberg differential game. Another contribution of this work is a (sub)optimal open-loop Stackelberg-based controller with a leader-follower structure, which both players act as inputs to a parametrically uncertain and disturbed nonlinear system.

Another contribution of this work is a technique for solving a two player zero-sum infinite horizon game subject to continuous-time unknown nonlinear dynamics. The technique involves a generalization of an actor-critic-identifier (ACI) structure which

is used to implement Hamilton-Jacobi-Isaac (HJI) approximation algorithm. The HJI approximation uses, two neural network (NN) actor structures and one NN critic structure to approximate the optimal control laws and value function, respectively.

Using the ACI technique, another contribution of this work is deriving an approximate solution to a N-player nonzero-sum game. The technique expands the ACI structure to solve a multi-player differential game problem, wherein N-actor and N-critic neural network structures are used to approximate the optimal control laws and the optimal value functions, respectively. Simulations and Lyapunov stability analysis are provided in each section to demonstrate the performance of the control designs.

CHAPTER 1
INTRODUCTION

1.1 Motivation

Numerous technical, economical or biological processes are often governed by ordinary differential equations, where the state of the system is a function of time and can be influenced with input parameters and exogenous environmental disturbances. The field of control theory studies methods of defining the input parameters such that state of the system behaves in an *acceptable* manner; where *acceptable* performance is generally defined in terms of time history and frequency response criteria. The field of optimal control theory was developed as an approach to analytically determine input parameters that will satisfy the physical constraints of the system, while also minimizing a performance criteria. Optimal control have been extensively investigated as a means to derive analytic proofs for classes of systems where a single input parameter influences the system, particularly for linear dynamics. However, some interesting questions arise when multiple input parameters are considered in a dynamic system, for example:

- How can systems be described when more than one input provides influence?
- How do the input parameters influence the system? or each other?
- What criteria demonstrates the behavior of the system is *acceptable*?
- Given optimality constraints, how can optimal controllers be determined?
- How can the criteria for optimality be determined?

Game theory is one approach to address concerns raised from the synthesis of controllers for complex dynamic systems. Game theory deals with strategic interactions among multiple input parameters, called players (and in some context agents), with each player's objectives captured in a value (or objective) function which the player either tries to maximize or minimize. For a non-trivial game, the value function of a player depends on the choices (actions, or equivalently decision variable) of at least one other player, and generally of all the players; hence, players cannot simply optimize their own objective

functions independent of the choices of the other players. This incorporates coupling between the actions of the players and binds them together in decision making even in a non-cooperative environment. This dissertation explores the utility of game theory (particularly differential game theory) in deriving controllers that stabilize nonlinear dynamic systems.

1.2 Background

The basic problem in optimization theory is to find the minimum value of a function:

$$\min_{x \in \mathcal{X}} V(x).$$

Typically $V(x)$ is a continuous function and the minimum is sought over a closed, possibly unbounded domain $\mathcal{X} \subseteq \mathbb{R}^n$. Extensive research has investigated the existence of the minimum, necessary and sufficient conditions for optimality, and computational methods for approximating a solution. Optimization theory launched the study of optimal control theory, where the state $x(t) \in \mathbb{R}^n$ evolves over time. For the standard optimal control problem, $x(t)$ evolves based on an ordinary differential equation (ODE) as

$$\dot{x} = f(t, x(t), u(t)) \qquad x(0) = x_0, t \in [0, T],$$

where $t \mapsto u(t) \in \mathcal{U}$ is the control strategy, ranging within the set of admissible control strategies $\mathcal{U}$. Given an initial condition x_0, the optimal control problem is to determine a control strategy $u(\cdot)$ which minimizes a value (or cost) function $J(t, x(t), u(t)) \in \mathbb{R}$

$$J = \Psi(x(T)) + \int_0^T L(\tau, x(\tau), u(\tau))\, d\tau, \tag{1–1}$$

where $\Psi \in \mathbb{R}$ is the terminal cost and $L \in \mathbb{R}$ is the local (or running) cost. Techniques to determine solutions to the optimal control problem have largely been based on two different fundamental ideas: Bellman's optimality principle and Pontryagin's maximum principle. Dynamic programming and the associated optimality principle, which is a sufficient condition for optimality, was introduced by Bellman in the United States,

whereas the maximum principle, which is a necessary condition for optimality, was introduced by Pontryagin in the Soviet Union. Bellman's approach to optimality considers that if a given state-action sequence is optimal, and the initial state and action are removed, the remaining sequence is also optimal. In contrast, Pontryagin's maximum principle involves finding an admissible control input that minimizes a Hamiltonian function. Pontryagin's maximum principle can only be applied to deterministic problems, but yields the same solutions as the dynamic programming approach. However, converse to the dynamic programming approach the maximum principle avoids the curse of dimensionality. Under certain conditions, the maximum principle and the Bellman principle can be reduced to the Hamilton-Jacobi-Bellman (HJB) equation. For nonlinear systems, the solution to the HJB equation can often be intractable or may not exist, thus various numeric and analytic techniques have been developed to approximate the solution [1–10] or indirectly determine a solution using inverse methods [11–18].

For systems with multiple players, game theory offers a natural extension to the optimal control problem. The work by Von Neumann and Morgenstern [19], widely regarded as the preliminary work on game theory, focused primarily on two-player, zero-sum games. Nash [20] provided a solution approach for a class of general N-player non-cooperative games. Motivated by the analysis of market economy, the monograph by Stackelberg [21] on hierarchical relationships among players provided further contribution to the theory of games. In the early 1950s Rufus Isaacs [22] pioneered differential game theory, which enabled a natural multi-player extension of the dynamic programming solution to the optimal control problem. In the case of two players, a differential game considers a system whose state $x(t) \in \mathbb{R}^n$ evolves according to the ODE

$$\dot{x} = f(t, x(t), u_1(t), u_2(t)) \qquad x(0) = x_0, t \in [0, T],$$

where $t \mapsto u_i(t) \in \mathcal{U}_i$, $i = 1, 2$ is the control strategy, ranging within the set of admissible control strategies $\mathcal{U}_i$. Given the initial condition x_0, the objective of the i-th player is to

minimize the value function $J_i\left(t, x\left(t\right), u_1\left(t\right), u_2\left(t\right)\right) \in \mathbb{R}$

$$J_i = \Psi\left(x\left(T\right)\right) + \int_0^T L\left(\tau, x\left(\tau\right), u_1\left(\tau\right), u_2\left(\tau\right)\right) d\tau.$$

While the optimal control problem has a clear definition of optimality that is tied to the minimization of the value function, the concept of optimality in game theory is not as well-defined. Specific forms of *optimality* for differential games can be defined in terms of equilibrium solutions (e.g. Nash, Pareto, Bayesian, Stackelberg). In this construct, the structure and information sets of differential games can vastly change the games objective. The amount of cooperation that occurs between players in a game is one of the key differences between different branches of game theory literature. If the players act in unison, but each player has a different objective (cost function) then multi-objective optimization is obtained [23–25]. In a situation in which there is a common value function and all players act cooperatively, then team theory is obtained [26–28] and if some subset of the players can make their decisions in unison, such that a mutual beneficial outcome can be obtained then a cooperative game is achieved [29, 30]. Cooperative game theory implies players are able to form binding commitments such that the best result occurs for the game at large, whereas noncooperative games are when each player pursues individual interests that are partly conflicting with others. The most extreme case of conflicting interest is a zero-sum game, in which players are diametrically opposed. Previous research has exploited noncooperative game theory [31–46] to provide numerous solution techniques to a wide range of control engineering applications. Solutions to noncooperative games are referred to as an equilibrium due to the fact that the solution represents control strategies that provide balance between independent interests of each player. A zero-sum game formulation that has been thoroughly explored in control theory is the two-player min-max H_∞ control optimization problem [41], where the controller is a minimizing player and the disturbance is a maximizing player yielding a Nash equilibria. In a zero-sum game with linear dynamics and an infinite horizon quadratic cost function, the Nash equilibrium

solution is equivalent to solving the generalized game algebraic Riccati equation (GARE). However, for nonlinear dynamics developing an analytical solution is complicated by solving a Hamilton-Jacobi-Isaacs (HJI) partial differential equation, where a solution may be non-smooth or intractable.

In a differential game formulation, the controlled system is influenced by a number of different inputs, computed by different players that individually try to optimize their respective performance functions. The control objective is to determine a set of policies that are admissible [47], i.e. control policies that guarantee the stability of the dynamic system and minimize individual performance functions to yield an equilibrium. The subsequent sections will introduce common definitions of equilibrium in differential games.

The Pareto Equilibrium. Pareto optimality is a genre of cooperative game theory. The so-called Pareto efficient solutions are based on the premise that the cost any one specific player incurs is not uniquely determined, rather the solution is determined when the cost incurred by all players simultaneously cannot be improved. Formally, the cost function $J_i\left(t, x, u_1, \ldots, u_N\right)$ is defined as

$$J_i = \int_0^T L\left(\tau, x\left(\tau\right), u_1\left(\tau\right), \ldots, u_N\left(\tau\right)\right) d\tau, \quad i = 1, \ldots, N \tag{1–2}$$

where $x\left(t\right)$ is the solution to

$$\dot{x} = f\left(t, x\left(t\right), u_1\left(t\right), \ldots, u_N\left(t\right)\right), \quad x\left(0\right) = x_0. \tag{1–3}$$

A set of control actions $\hat{u}$ is called Pareto efficient if the set of inequalities

$$J_i\left(u\right) \leq J_i\left(\hat{u}\right), \; i = 1, \ldots, N,$$

does not permit a solution $u \in \mathcal{U}$, where at least one of the inequalities is strict. The corresponding point $\left(J_1\left(\hat{u}\right), \ldots, J_N\left(\hat{u}\right)\right) \in \mathbb{R}^{\mathbb{N}}$ is called a Pareto solution. The set of all Pareto solutions is called the Pareto frontier. A well-known way to determine Pareto solutions is to solve a parameterized optimal control problem [48–50], however,

in general, it is unclear whether this approach yields all Pareto solutions. The Pareto efficient solution is included in this introduction for completeness, however the techniques developed in this work will not utilize optimality defined in this structure.

The Nash Equilibrium. A Nash differential game consists of multiple players making simultaneous decisions where each player has an outcome that cannot be unilaterally improved from a change in strategy. Players are committed to following a predetermined strategy based on knowledge of the initial state, the system model and the cost functional to be minimized. Formally, for the cost function in Eq. 1–2, subject to the state dynamics in Eq. 1–3, a set of control actions $(u_i^*, \ldots, u_N^*)$ is a Nash equilibrium solution for the N-player game, if the following N inequalities are satisfied for all $u_i^* \in \mathcal{U}_i, i \in N$:

$$\left.\begin{aligned} J_1^* \triangleq J_1\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq J_1\left(x\left(t\right), u_1, u_2^*, \ldots u_N^*\right) \\ J_2^* \triangleq J_2\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq J_2\left(x\left(t\right), u_1^*, u_2, \ldots u_N^*\right) \\ \cdots \\ \cdots \\ J_N^* \triangleq J_N\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq J_N\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N\right) \end{aligned}\right\}.$$

Solution techniques to the Nash equilibrium can be classified in various ways depending on the amount of information available to the players (open-loop, closed-loop, feedback, etc.), the objectives of each player (zero-sum and nonzero-sum), the planning horizon (infinite horizon and finite horizon), and the nature of the dynamic constraints (continuous, discrete, linear, nonlinear, etc). A large body of research has focused on linear quadratic games on a finite time horizon. One issue that the Nash equilibria poses, is that in general a unique Nash equilibrium is not expected. Non-uniqueness issues with Nash equilibria were discussed for a nonzero-sum differential game in [51]. In the case of the open-loop nonzero-sum game, where every player knows at time $t \in [0, T]$ the initial state x_0, conditions for the existence of a unique Nash equilibrium can be given by [52]. In the case of closed-loop perfect state information, where every player knows at time $t \in [0, T]$ the complete history of the state, it has been shown that infinitely many Nash equilibria may

exist. In this case, it is possible to restrict the Nash equilibrium to a subset of feedback solutions, which is known as the (sub)game perfect Nash equilibria (or feedback Nash equilibria). The work of Case [53] and Friedman [54], was shown that (sub)game perfect Nash equilibria are (at least heuristically) given by feedback strategies and that their corresponding value functions are the solution to a system of Hamilton–Jacobi equations. These concepts have been successfully applied to linear-quadratic (LQ) differential games [48, 53]. A special case of the Nash game is the min-max saddle point equilibrium, which is widely used in control theory to minimize control effort under a *worst-case* level of uncertainty. The saddle point equilibrium has been heavily exploited in H_∞ control theory [55], which considers finding the smallest gain $\gamma \geq 0$ under which the upper value of the cost function

$$J_\gamma(u, v) = \int_0^\infty Q(x) + u(x)^2 - \gamma^2 \|v(x)\|^2 \, d\tau, \tag{1–4}$$

is bounded and finding the corresponding controller that achieves this upper bound. H_∞ control theory relates to LQ dynamic games in the sense that the worst-case H_∞ design problems have equal upper and lower bounds of the objective function in Eq. 1–4, which results in the saddle-point solution to the LQ game problem. In both the H_∞ control problem and the LQ game problem, the underlying dynamic optimization is for a two player zero-sum game with the controller being the minimizing player and the disturbance being the maximizing player.

The Stackelberg Equilibrium. A hierarchical nonzero-sum technique was derived by Von Stackelberg [21], where an equilibrium solution can be determined when one player's strategy has influence over another player's strategy. The Stackelberg technique has been accepted as the solution to a broad class of hierarchical decision making problems where one decision maker (called the leader) announces a strategy prior to the announcement of the second decision maker's (called the follower) strategy. For the two

player game, consider the cost function defined in Eq. 1–2, where $N = 2$, with the dynamic constraints in Eq. 1–3. The optimal reaction set for player 1 (the follower $u_1 \in \mathcal{U}_1$) to a control $u_2 \in \mathcal{U}_2$ is

$$\mathcal{R}_1(u_2) = \{\gamma \in \mathcal{U}_1 \,|\, J_1(\gamma, u_2) \leq J_1(u_1, u_2)\ \forall u_1 \in \mathcal{U}_1\}.$$

If player 2 is leading then $u_2^* \in \mathcal{U}_2$ is called a Stackelberg equilibrium for player 2. If for all $u_2 \in \mathcal{U}_2$

$$\sup_{\gamma \in \mathcal{R}_1(u_2^*)} J_2(\gamma, u_2^*) \leq \sup_{\gamma \in \mathcal{R}_1(u_2^*)} J_2(\gamma, u_2),$$

then $u_1^* \in \mathcal{R}_1(u_2^*)$ is an optimal Stackelberg strategy for the follower. An important motivation for the use of the Stackelberg strategy by the leader lies in the reduced value of the cost function as compared to the Nash strategy; thus it can be shown that a Stackelberg strategy is at least as good as any Nash strategy for the leader [56]. The Stackelberg strategy can be divided into three essential types: 1) open-loop strategies [37, 57, 58], 2) closed-loop strategies [35, 36, 45], and 3) feedback strategies [38, 58–61]. In [37], an N-player nonzero-sum Stackelberg differential game is generalized for a multi-input linear system, where the players are divided into a group of leaders that use a Stackelberg policy and a group of followers that use a Nash policy. Furthermore, hierarchical control problems for closed-loop Stackelberg solutions are presented in [35,36,45]. In [36] necessary conditions are developed for a closed-loop two-player Stackelberg game with a linear quadratic cost constrained by linear dynamics. Whereas in [35], sufficient conditions are derived for a closed-loop two player solution subject to a discrete linear system. The novelty of [35] is that the requirement for an a priori restriction on the structure of the player's strategies is removed. The technique in [35] is extended in [45] to derive the team optimal solution for the two-person continuous-time linear differential game problems under quadratic cost functions for both players. Feedback strategies are presented in [59–61] which focus on the solution to the dynamic Stackelberg problem and are characteristic of containing memoryless (pure feedback) features in the control strategies.

A solution technique for a class of nonclassical dynamics is presented in [38]. Previous research of the open-loop Stackelberg game in control theory has mainly focused on deriving an analytic solution and providing gain constraints; however, these results are limited to linear systems with known plants and don't demonstrate stabilization properties of the control law.

1.3 Problem Statement and Contributions

It is widely known that minimizing the cost function in Eq. 1–1 is equivalent to minimizing the HJB equation given by

$$\begin{aligned} 0 &= \min_{u_i} \left[\nabla V_i^* \left(f\left(t, x\left(t\right), u_1\left(t\right), \ldots, u_N\left(t\right)\right)\right) + L\left(x, u_1, \ldots, u_N\right)\right], \\ & V_i^*\left(0\right) = 0,\ i \in N \end{aligned}$$

which is a partial differential equation. For linear systems, solving this equation reduces to finding the solution of a generalized game algebraic Riccati, equation, however for a nonlinear system, finding analytic solutions to the HJB may be burdensome. Furthermore, for certain classes of multi-player games (particularly nonzero-sum games), minimization of a cost function results in a set of coupled HJB equations. In linear systems, these coupled HJB equations reduce to a coupled set of Differential Riccati Equations (DRE), and various techniques have been proposed for establishing necessary and sufficient conditions for the existence of a solution to DREs. A body of research (e.g. [41, 52, 62–64]) has also been dedicated to determining the conditions for uniqueness of differential games with linear dynamics, particularly in the area of games with linear quadratic cost functions. For nonlinear dynamics, the coupled HJB equations are nonlinear and derivations of existence and uniqueness are often sparse. While it is known that analytic control laws can be derived for these systems, often the control laws are dependent on the solution to the coupled HJB equations, which can be impractical for real-time implementation on engineering systems. Two approaches are investigated in this dissertation to address these

limitations for uncertain nonlinear continuous-time systems: robust feedback linearization and approximate dynamic programming.

Analytic Optimal Control Solutions: A common technique used in control applications to derive an analytic optimal control solution to a nonlinear system is the nonlinear H_∞ control solution [55, 65–68]. However, the infinite horizon formulation of the nonlinear H_∞ control problem requires a significant computational effort for nonlinear systems thereby making its application to real systems often nearly impossible. In particular, the challenge is that the nonlinear H_∞ control problem requires the solution to the HJB. Inverse optimal control (IOC) [11, 12, 15–17, 69–74] is an approach to develop optimal controllers for systems without solving the HJB equation. The objective of IOC is to develop a controller that is optimal with respect to a stability analysis-derived cost functional. Previous research in IOCs has focused on finding a control Lyapunov function (CLF) and a controller that stabilizes the system, then determining if the controller is optimal for a meaningful cost. IOC differs from other analytic optimal control techniques by requiring the local cost to be posteriori determined by the stabilizing feedback, rather than a priori by the designer. When parametric uncertainty exists in the system, several inverse optimal adaptive controllers (IOACs) [74–79] have been developed to compensate for uncertainties that are linear in the parameters (LP).

The use of neural networks (NNs) is another approach to approximate unknown dynamics as a means of developing approximate analytic optimal controllers. Specifically, results such as [47, 80–84] find a one-player optimal control law for a given cost function constrained by a partial feedback linearized system, and then modify the optimal control law with a NN to approximate the unknown dynamics. The tracking errors for the NN methods in [47, 80–84] are proven to be uniformly ultimately bounded (UUB) and the resulting state space system, for which the analytic optimal controller is developed, is approximated. In [85], an optimal controller is derived for an Euler-Lagrange system using a RISE feedback technique combined with an optimal controller that minimizes an

objective function. The RISE controller partially feedback linearizes the system, leaving a residue nonlinear system that can be manipulated to form a linear quadratic optimal control problem. The work using a partial feedback linearized system to determine an optimal controller is the basis for Chapters 2 and 3.

Approximation of Optimal Control Solutions: Due to the difficulty involved in determining a solution to the HJB a branch of research is dedicated to approximating a solution to the optimal control problem via dynamic programming [86–90]. Reinforcement learning (RL) is a method wherein appropriate actions are learned based on evaluative feedback from the environment. A widely used RL method is based on the actor-critic (AC) architecture, where an actor performs certain actions by interacting with its environment, the critic evaluates the actions and gives feedback to the actor, leading to an improvement in the performance of subsequent actions. AC algorithms are pervasive in machine learning and are used to learn the optimal policy online for finite-space discrete-time Markov decision problems [1, 2, 91]. Previous research on RL using adaptive critics in the machine learning community [1–5] provides an approach to determining the solution of an optimal control problem using Approximate Dynamic Programming (ADP) [86–90]. The discrete/iterative nature of the ADP formulation lends itself naturally to the design of discrete-time optimal controllers [89, 92–96]. Baird [97] proposed Advantage Updating, an extension of the Q-learning algorithm which could be implemented in continuous-time and provided fast convergence. A HJB-based framework is used in [98] and [99], and Galerkin's spectral method is used to approximate the generalized HJB solution in [7]. All of the aforementioned approaches for continuous-time nonlinear systems are computed offline and/or require complete knowledge of system dynamics. A contribution in [100] is the requirement of only partial knowledge of the system and a hybrid continuous-time/discrete-time sampled data controller is developed based on policy iteration (PI), where the feedback control operation of the actor occurs at faster time scale than the learning process of the critic. Vamvoudakis and Lewis [101] extended the idea by designing

a hybrid model-based online algorithm called synchronous PI which involved synchronous continuous-time adaptation of both actor and critic neural networks. Bhasin et. al [102] developed a continuous actor-critic-identifier (ACI) technique to solve the infinite horizon optimal single player optimal control problem, by using a robust dynamic neural network (DNN) to identify the dynamics and a critic NN to approximate the value function. This technique removes the requirement of complete knowledge of the system drift dynamics and incorporates an indirect adaptive control technique for a RL problem. Most of the previous research on continuous-time reinforcement learning algorithms that provide an online approach to the solution of optimal control problems, assumed that the dynamical system is affected by a single control strategy. Previous research has also investigated the generalization of RL controllers to differential game problems [101, 103–109]. Techniques utilizing Q-learning algorithms have been developed for a zero-sum game in [110]. An ADP procedure that provides a solution to the HJI equation associated with the two-player zero-sum nonlinear differential game is introduced in [103]. The ADP algorithm involves two iterative cost functions finding the upper and lower performance indices as sequences that converge to the saddle point solution of the game. The AC structure required for learning the saddle point solution is composed of four action networks and two critic networks. An iterative ADP solution was presented in [104], where it considers solving zero-sum differential games under the condition that the saddle point does not exist, and a mixed optimal performance index function is obtained under a deterministic mixed optimal control scheme when the saddle point does not exist. Another ADP iteration technique is presented in [105], in which the non-affine nonlinear quadratic zero-sum game is transformed into an equivalent sequence of linear quadratic zero-sum games to approximate an optimal saddle point solution. In [106], an integral RL method is used to determine an online solution to the two player nonzero-sum game for a linear system without complete knowledge of the dynamics. The synchronous PI method in [101] was then further generalized to solve the two-player zero-sum game problem in [108] and a

multi-player nonzero-sum game in [109] for nonlinear continuous-time systems with known dynamics. The work from [101, 102, 108, 109] provides the foundation for Chapters 4 and 5, where the two player zero-sum game and multi-player nonzero sum game are solved using an ACI technique where the controllers are implemented online and without the requirement of complete knowledge of the system drift dynamics.

This dissertation focuses on developing differential-game based controllers for specific classes of uncertain continuous-time nonlinear systems. The contributions of Chapters 2-5 are as follows:

Asymptotic Nash Optimal Control Design for an Uncertain Euler-Lagrange System: The main contribution of Chapter 2 is the development of robust (sub)optimal Nash-based feedback control laws. This chapter combines the Robust Integral Sign of the Error (RISE) [111] controller with an optimal Nash strategy to stabilize an uncertain Euler-Lagrange system with additive disturbances. One advantage of this method over previous techniques is that the controller accounts for uncertainty in a state-varying mass inertia matrix. This chapter illustrates the development of the RISE controller which is used to asymptotically identify the nonlinearities in the dynamics. By applying the RISE controller, the nonlinear dynamics converge to a residual partially linearized system, and the solution to the feedback Nash game for the residual system is used to derive the stabilizing control laws. The (sub)optimal feedback controllers minimize a cost functional in the presence of unknown bounded disturbances. A Lyapunov-based analysis is used to prove asymptotic tracking for the combined RISE and Nash-based strategy. Existence of the feedback Nash solution is discussed and simulation results demonstrate the control performance.

Asymptotic Stackelberg Optimal Control Design for an Uncertain Euler-Lagrange System: Chapter 3 derives a stabilizing set of controllers for a system in which one control input has additional information about the other control input. This scenario is representative of many engineering applications, where interactions among a

leader and a follower are observed (e.g. formation control, autonomous docking, aerial refueling, etc.). In comparison, the Nash game in Chapter 2 considers both players to have no a priori knowledge about each other, which can lead to an overly conservative control design. The main contribution of this work is the development of robust (sub)optimal open-loop Stackelberg-based controllers for the leader and follower, which both act as inputs to an uncertain nonlinear system. A RISE controller is used in conjunction with the derived Stackelberg strategy. The RISE controller enables the dynamics to be written in a residual form, which allows for the Stackelberg differential game formulation. The controller accounts for a state-varying mass inertia matrix, as well as, additive exogenous disturbances and parametric uncertainties in the dynamics. One novelty of the techniques in Chapter 2 and 3 is the use of the Skew Symmetric property to reduce the coupled differential Riccati equations to algebraic Riccati equations which allows for conditions to be established for the solution to the Nash and Stackelberg nonzero-sum games. The control formulation utilizes the solution to the hierarchical open-loop Stackelberg nonzero-sum game to derive the feedback control laws. A Lyapunov-based stability analysis to prove asymptotic tracking, and a brief discussion on existence of solution is provided. Simulation results are included to illustrate the performance of the developed controller.

Nonlinear two-player zero-sum game approximate solution using an HJI approximation algorithm: In contrast to the approaches in Chapters 2 and 3, which are largely based on Pontryagin's maximum principle, the techniques in Chapters 4 and 5 seek to approximate the solution to the HJI and HJB. This approximation is based on Bellman's optimality principle and dynamic programming. The main contribution of Chapter 4 is solving a two player zero-sum infinite horizon game subject to continuous-time unknown nonlinear dynamics that are affine in the input. In the developed method, two actor and one critic NNs use gradient and least squares-based update laws, respectively, to minimize the Bellman error, which is the difference between the exact and the approximate HJI equations. The identifier DNN is a combination of a

Hopfield-type [112] component, in parallel configuration with the system [113], and a RISE component. The Hopfield component of the DNN learns the system dynamics based on online gradient-based weight tuning laws, while the RISE term robustly accounts for the function reconstruction errors, guaranteeing asymptotic estimation of the state and the state derivative. The online estimation of the state derivative allows the ACI architecture to be implemented without knowledge of system drift dynamics; however, knowledge of the input gain matrix is required to implement the control policy. While the design of the actor and critic are coupled through a HJI equation, the design of the identifier is decoupled from the actor-critic and can be considered as a modular component in the ACI architecture. Convergence of the ACI-based algorithm and stability of the closed-loop system are analyzed using Lyapunov-based adaptive control methods and a persistence of excitation (PE) condition is used to guarantee convergence to within a bounded region of the optimal control and UUB stability of the closed-loop system. The main advantage of this ACI approach consists in the fact that neither of the two participants in the game makes use of explicit knowledge of the model of the drift dynamics of the system that they influence through their control strategies. This means that the two players will learn online the most effective control strategies that correspond to the Nash equilibrium while using no explicit knowledge on the drift dynamics of the differential game. In addition, this technique converges to the approximate solution of the Nash equilibrium, without the need for iterative techniques or offline training, and it incorporates theory from adaptive control, making it an approximate indirect adaptive solution to a two player zero-sum differential game.

Nonlinear N-player nonzero-sum game approximate solution using a HJB approximation algorithm: Nonzero-sum games present different challenges when compared to zero-sum games. For nonlinear dynamics, the HJI for zero-sum games is equivalently a coupled set of nonlinear HJB equations for nonzero-sum games. Research in nonzero-sum games for nonlinear systems is sparse and there are many open research

challenges. Chapter 5 considers a N-player nonzero-sum infinite horizon game subject to continuous-time uncertain nonlinear dynamics. Using the ACI technique, the main contribution of this work is deriving an approximate solution to a N-player nonzero-sum game with a technique that is continuous, online and based on adaptive control theory. Previous research in the area has focused on scalar nonlinear systems or implemented iterative/hybrid techniques that required complete knowledge of the drift dynamics. The technique developed in Chapter 5 expands the ACI structure to solve a differential game problem, wherein N-actor and N-critic neural network structures are used to approximate the optimal control laws and the optimal value function set, respectively. The main traits of this online algorithm involve the use of ADP techniques and adaptive theory to determine the Nash equilibrium solution of the game in manner that does not require full knowledge of the system dynamics and the approximately solves the underlying set of coupled HJB equations of the game problem. For an equivalent nonlinear system, previous research makes use of offline procedures or requires full knowledge of the system dynamics to determine the Nash equilibrium. A Lyapunov-based stability analysis shows that UUB tracking for the closed-loop system is guaranteed and a convergence analysis demonstrates that the approximate control policies converge to a neighborhood of the optimal solutions.

CHAPTER 2
ASYMPTOTIC NASH OPTIMAL CONTROL DESIGN FOR AN UNCERTAIN EULER-LAGRANGE SYSTEM

A zero-sum game formulation that has received notable interest in control theory is the two-player min-max H_∞ control problem [41], where the controller is a minimizing player and the disturbance is a maximizing player in yielding a Nash equilibria. The H_∞ formulation is well suited for disturbance rejection problems where the controller and *worst-case* disturbance are derived for a Nash equilibrium. A Nash strategy is one such that the outcome of each player's input cannot unilaterally improve by changing the player's strategy. Previous Nash games focus on zero-sum solution techniques for linear systems with infinite horizon quadratic cost functionals. In this chapter, a general framework is developed for feedback control of an Euler-Lagrange system using a feedback nonzero-sum Nash differential game. A RISE controller is used to compensate for some uncertain nonlinearities so that Nash optimal controllers can be derived for the general tracking problem. A Lyapunov-based stability analysis and numerical simulations are provided to examine the stability and performance of the developed controllers.

2.1 Dynamic Model and Properties

The class of nonlinear dynamic systems considered in this chapter are assumed to be modeled by the following Euler-Lagrange formulation:

$$M(q)\ddot{q} + V_m(q,\dot{q})\dot{q} + G(q) + F\left(\dot{q}\right) + \tau_d(t) = \tau_1(t) + \tau_2(t). \tag{2–1}$$

In Eq. 2–1, $M(q) \in \mathbb{R}^{n\times n}$ denotes the generalized inertia matrix, $V_m(q,\dot{q}) \in \mathbb{R}^{n\times n}$ denotes the generalized centripetal-Coriolis matrix, $G(q) \in \mathbb{R}^n$ denotes the generalized gravity vector, $F\left(\dot{q}\right) \in \mathbb{R}^n$ denotes the generalized friction vector, $\tau_d(t) \in \mathbb{R}^n$ is a general bounded disturbance, $\tau_1(t), \tau_2(t) \in \mathbb{R}^n$ represents input control vectors, and $q(t)$, $\dot{q}(t)$, $\ddot{q}(t) \in \mathbb{R}^n$ denote the generalized position, velocity, and acceleration vectors, respectively. The subsequent development is based on the assumption that $q(t)$ and $\dot{q}(t)$ are measurable, and

$M(q)$, $V_m(q,\dot{q})$, $G(q)$, $F(\dot{q})$, and $\tau_d(t)$ are unknown. Moreover, the following properties and assumptions are exploited in the subsequent development.

Assumption 2.1. *The inertia matrix $M(q)$ is symmetric, positive definite, and satisfies the following inequality $\forall y(t) \in \mathbb{R}^n$:*

$$m_1 \|\xi\|^2 \leq \xi^T M(q) \xi \leq \bar{m}(q) \|\xi\|^2, \tag{2–2}$$

where $m_1 \in \mathbb{R}$ is a known positive constant, $\bar{m}(q) \in \mathbb{R}$ is a known positive function, and $\|\cdot\|$ denotes the standard Euclidean norm.

Assumption 2.2. *The following skew-symmetric relationships are satisfied:*

$$\xi^T \left(\dot{M}(q) - 2V_m(q,\dot{q}) \right) \xi = 0 \qquad \forall \xi \in \mathbb{R}^n \tag{2–3}$$

$$-\left(\dot{M}(q) - 2V_m(q,\dot{q}) \right)^T = \dot{M}(q) - 2V_m(q,\dot{q}) \tag{2–4}$$

$$-\left(\dot{M}(q) - \left(V_m(q,\dot{q}) + V_m^T(q,\dot{q}) \right) \right)^T = \dot{M}(q) - \left(V_m(q,\dot{q}) + V_m^T(q,\dot{q}) \right). \tag{2–5}$$

Assumption 2.3. *If $q(t), \dot{q}(t) \in \mathcal{L}_\infty$, then $V_m(q,\dot{q})$, $F(\dot{q})$ and $G(q)$ are bounded. Moreover, if $q(t), \dot{q}(t) \in \mathcal{L}_\infty$, then the first and second partial derivatives of the elements of $M(q)$, $V_m(q,\dot{q})$, $G(q)$ with respect to $q(t)$ exist and are bounded, and the first and second partial derivatives of the elements of $V_m(q,\dot{q})$, $F(\dot{q})$ with respect to $\dot{q}(t)$ exist and are bounded.*

Assumption 2.4. *The desired trajectory is assumed to be designed such that $q_d(t)$, $\dot{q}_d(t)$, $\ddot{q}_d(t)$, $\dddot{q}_d(t)$, $\ddddot{q}_d(t) \in \mathbb{R}^n$ exist, and are bounded.*

Assumption 2.5. *The disturbance term and its first two time derivatives, i.e. $\tau_d(t)$, $\dot{\tau}_d(t)$, $\ddot{\tau}_d(t)$ are bounded by known constants.*

2.2 Error System Development

The control objective is to ensure that the generalized coordinates track a desired time-varying trajectory, denoted by $q_d(t) \in \mathbb{R}^n$, despite uncertainties in the dynamic model, while minimizing a given performance index. To quantify the tracking objective, a position

tracking error, denoted by $e_1(q,t) \in \mathbb{R}^n$, is defined as

$$e_1 \triangleq q_d - q. \tag{2–6}$$

To facilitate the subsequent analysis, filtered tracking errors, denoted by $e_2(q,\dot{q},t)$ and $r(q,\dot{q},\ddot{q},t) \in \mathbb{R}^n$, are also defined as

$$e_2 \triangleq \dot{e}_1 + \alpha_1 e_1, \tag{2–7}$$

$$r \triangleq \dot{e}_2 + \alpha_2 e_2, \tag{2–8}$$

where $\alpha_1, \alpha_2 \in \mathbb{R}^{n\times n}$ are positive definite, constant, gain matrices. The filtered tracking error $r(q,\dot{q},\ddot{q},t)$ is not measurable due to the functional dependence on $\ddot{q}(t)$. The error systems are based on the assumption that the generalized coordinates of the Euler-Lagrange dynamics allow additive and not multiplicative errors. A state-space model can be developed based on the tracking errors in Eqs. 2–6 and 2–7. Based on this model, a controller is derived that minimizes a quadratic performance index under the (temporary) assumption that the dynamics in Eq. 2–1 are known. A control term is developed as the solution to a nonzero-sum Nash differential game. The Nash-derived control term is then combined with a robust controller to identify the unknown dynamics and additive disturbance, thereby relaxing the temporary assumption that these dynamics are known. To develop a state-space model for the tracking errors in Eqs. 2–6 and 2–7, the inertia matrix is pre-multiplied to the time derivative of Eq. 2–7, and substitutions are made from Eqs. 2–1 and 2–6 to obtain

$$M\dot{e}_2 = -V_m e_2 + h + \tau_d - (\tau_1 + \tau_2), \tag{2–9}$$

where the nonlinear function $h(q,\dot{q},t) \in \mathbb{R}^n$ is defined as

$$h \triangleq M(\ddot{q}_d + \alpha_1 \dot{e}_1) + V_m(\dot{q}_d + \alpha_1 e_1) + G + F. \tag{2–10}$$

Under the (temporary) assumption that the dynamics in Eq. 2–1 are known, the control inputs can be designed as

$$\tau_1 + \tau_2 \triangleq h + \tau_d - (u_1 + u_2) \tag{2–11}$$

where $u_1(t), u_2(t) \in \mathbb{R}^n$ are auxiliary control inputs that will be designed to minimize a desired performance index. By substituting Eq. 2–11 into Eq. 2–9 the closed-loop error system for $e_2(t)$ can be obtained as

$$M\dot{e}_2 = -V_m e_2 + u_1 + u_2. \tag{2–12}$$

A state-space model for Eqs. 2–7 and 2–12 can now be developed as

$$\dot{z} = A(q, \dot{q})\, z + B_1(q)\, u_1 + B_2(q) u_2 \tag{2–13}$$

where $A(q, \dot{q}) \in \mathbb{R}^{2n \times 2n}$, $B_1(q), B_2(q) \in \mathbb{R}^{2n \times n}$, $z(t) \in \mathbb{R}^{2n}$ and are defined as

$$A \triangleq \begin{bmatrix} -\alpha_1 & I_{n\times n} \\ 0_{n\times n} & -M^{-1}V_m \end{bmatrix}, \qquad B_1 = B_2 \triangleq \begin{bmatrix} 0_{n\times n} & M^{-1} \end{bmatrix}^T, \qquad z \triangleq \begin{bmatrix} e_1^T & e_2^T \end{bmatrix},$$

where $I_{n\times n}$ and $0_{n\times n}$ denote a $n \times n$ identity matrix and matrix of zeros, respectively.

2.3 Two Player Feedback Nash Nonzero-Sum Differential Game

The Nash solution is characterized by an equilibria in which each player has an outcome that cannot be improved by a unilateral change of strategy. The Nash strategy safeguards against a single player deviating from the equilibrium strategy and is well suited for problems where cooperation between players cannot be guaranteed. To formulate the feedback Nash solution, consider the system in Eq. 2–13 in terms of players of a Nash equilibrium game (u_{N1},u_{N2}) given as

$$\dot{z} = A(q, \dot{q})\, z + B_1(q)\, u_{N1} + B_2(q) u_{N2}, \tag{2–14}$$

Each player has a cost functional $J_{N1}(z, u_{N1}, u_{N2}), J_{N2}(z, u_{N1}, u_{N2}) \in \mathbb{R}$ defined as

$$J_{N1} = \frac{1}{2}\int_{t_0}^{\infty} \left(z^T Q_N z + u_{N1}^T R_{N11} u_{N1} + u_{N2}^T R_{N12} u_{N2}\right) dt \tag{2–15}$$

$$J_{N2} = \frac{1}{2}\int_{t_0}^{\infty} \left(z^T L_N z + u_{N2}^T R_{N22} u_{N2} + u_{N1}^T R_{N21} u_{N1}\right) dt, \tag{2–16}$$

where $t_0 \in \mathbb{R}$ is the initial time, $Q_N, L_N \in \mathbb{R}^{2n\times 2n}$ are symmetric constant matrices defined as

$$Q_N = \begin{bmatrix} Q_{N11} & Q_{N12} \\ Q_{N12}^T & Q_{N22} \end{bmatrix} \qquad L_N = \begin{bmatrix} L_{N11} & L_{N12} \\ L_{N12}^T & L_{N22} \end{bmatrix},$$

where Q_{Nij}, and $L_{Nij} \in \mathbb{R}^{n\times n}$ are symmetric semi-definite constant matrices, and $R_{Nij} \in \mathbb{R}^{n\times n}$ is positive definite for $i, j = 1, 2$. This chapter focuses on a game with memoryless perfect state information, therefore the controller information set contains only the initial conditions z_0 and the current state estimates $z(t)$ at time t. In this context, the actions of the players are completely determined by the relations $(u_{N1}, u_{N2}) = (\gamma_1(z_0, z), \gamma_2(z_0, z))$. A pair of strategies (γ_1^*, γ_2^*) is called a Nash equilibrium set for the differential game if for all strategies (γ_1, γ_2) the following inequalities hold

$$\begin{aligned} J_{N1}(\gamma_1, \gamma_2^*) &\geq J_{N1}(\gamma_1^*, \gamma_2^*) \\ J_{N2}(\gamma_1^*, \gamma_2) &\geq J_{N2}(\gamma_1^*, \gamma_2^*). \end{aligned}$$

Based on the minimum principal [114], the Hamiltonians $H_{N1}(z, u_{N1}, u_{N2}), H_{N2}(z, u_{N1}, u_{N2}) \in \mathbb{R}$ of the control inputs u_{N1} and u_{N2} are defined as,

$$\begin{aligned} H_{N1} = {} & \frac{1}{2}\left(z^T Q_N z + u_{N1}^T R_{N11} u_{N1} + u_{N2}^T R_{N12} u_{N2}\right) \\ & + \lambda_{N1}^T \left(Az + B_1 u_{N1} + B_2 u_{N2}\right) \end{aligned} \tag{2–17}$$

$$\begin{aligned} H_{N2} = {} & \frac{1}{2}\left(z^T L_N z + u_{N2}^T R_{N22} u_{N2} + u_{N1}^T R_{N21} u_{N1}\right) \\ & + \lambda_{N2}^T \left(Az + B_1 u_{N1} + B_2 u_{N2}\right) \end{aligned} \tag{2–18}$$

respectively. Two previous results from [33] and [62] utilizing the memoryless perfect information structure are given in the following theorems.

Theorem 2.1. *Let the strategies* (γ_1^*, γ_2^*) *be such that there exists solutions* (λ_1, λ_2) *to the differential equations*

$$\dot{\lambda}_1 = -\frac{\partial}{\partial z} H_{N1}(z, \gamma_1^*, \gamma_2^*, \lambda_1) - \frac{\partial}{\partial u_2} H_{N1}(z, \gamma_1^*, \gamma_2^*, \lambda_1) \cdot \frac{\partial}{\partial z} \gamma_2^*(z_0, z) \tag{2–19}$$

$$\dot{\lambda}_2 = -\frac{\partial}{\partial z} H_{N2}(z, \gamma_1^*, \gamma_2^*, \lambda_1) - \frac{\partial}{\partial u_1} H_{N2}(z, \gamma_1^*, \gamma_2^*, \lambda_1) \cdot \frac{\partial}{\partial z} \gamma_1^*(z_0, z), \tag{2–20}$$

where H_{N1} *and* H_{N2} *are defined in Eqs. 2–17 and 2–18 and such that*

$$\frac{\partial}{\partial u_1} H_{N1} = 0 \qquad \frac{\partial}{\partial u_2} H_{N2} = 0,$$

and z *satisfies*

$$\dot{z} = Az + B_1\gamma_1^*, + B_2\gamma_2^*, \quad z(0) = z_0.$$

Then (γ_1^*, γ_2^*) *is a Nash equilibrium with respect to the memoryless perfect state information structure and the following equalities hold*

$$u_{1N}^* = \gamma_1^* = -R_{N11}^{-1} B_1^T \lambda_{N1} \tag{2–21}$$

$$u_{2N}^* = \gamma_2^* = -R_{N22}^{-1} B_2^T \lambda_{N2} \tag{2–22}$$

Proof. Refer to [33]. □

Remark 2.1. From Theorem 2-1, it can easily be shown that the open loop Nash equilibrium is also a Nash equilibrium with respect to the memoryless perfect state information structure.

In [33] it was shown that if the admissible strategies are restricted to a class of (possibly time varying) linear feedback strategies, then there exists an analytic linear feedback for the Nash equilibrium. The following theorem summarizes that result for the current system.

Theorem 2.2. *Suppose* (K_N, P_N) *satisfy the coupled Differential Riccati equations (DRE), given by*

$$\begin{aligned} 0 &= \dot{K_N} + K_N A + A^T K_N - K_N B_1 R_{N11}^{-1} B_1^T K_N \\ &\quad - K_N B_2 R_{N22}^{-1} B_2^T P_N + Q_N^T - P_N B_2 R_{N22}^{-1} B_2^T K_N \\ &\quad + P_N B_2 R_{N22}^{-T} R_{N12} R_{N22}^{-1} B_2^T P_N \end{aligned} \tag{2–23}$$

$$\begin{aligned} 0 &= \dot{P_N} + P_N A + A^T P_N - P_N B_1 R_{N11}^{-1} B_1^T K_N \\ &\quad - P_N B_2 R_{N22}^{-1} B_2^T P_N + L_N^T - K_N B_1 R_{N11}^{-1} B_1^T P_N \\ &\quad + K_N B_1 R_{N11}^{-T} R_{N21} R_{N11}^{-1} B_1^T K_N. \end{aligned} \tag{2–24}$$

Then the pair of strategies $(\gamma_1^*, \gamma_2^*) \triangleq \left(-R_{N11}^{-1} B_1^T K_N z, -R_{N22}^{-1} B_2^T P_N z\right)$ *is a linear feedback Nash equilibrium and the solutions to the costate equations defined in Eqs. 2–19 and 2–20 are linear feedbacks given as*

$$\lambda_{N1} = K_N z \tag{2–25}$$

$$\lambda_{N2} = P_N z \tag{2–26}$$

Proof. Refer to [62]. □

Remark 2.2. It was shown in [115] that for more general (i.e. nonlinear feedback) strategies there may exist infinitely many feedback Nash equilibria for the memoryless perfect state information structure.

The subsequent analysis utilizes the feedback structure from Theorem 2-2 and the skew-symmetric property for Euler-Lagrange systems to reduce the DREs defined in Eqs. 2–23 and 2–24 to algebraic Riccati equations, thereby deriving an analytic solution for (K_N, P_N), based on the control gains. Assume that $K_N(t), P_N(t) \in \mathbb{R}^{2n \times 2n}$ are time-varying positive definite diagonal matrices defined as

$$K_N = \begin{bmatrix} K_{N11} & 0_{n\times n} \\ 0_{n\times n} & K_{N22} \end{bmatrix} \qquad P = \begin{bmatrix} P_{N11} & 0_{n\times n} \\ 0_{n\times n} & P_{N22} \end{bmatrix}.$$

The two DREs in Eqs. 2–23 and 2–24 must be solved simultaneously to yield a control strategy for the both players. Substituting Eqs. 2–14, 2–25, and 2–26 into Eq. 2–23 yields four simultaneous equations as

$$0 = \dot{K}_{N11} - K_{N11}\alpha_1 - \alpha_1^T K_{N11} + Q_{N11} \tag{2–27}$$

$$0 = K_{N11} + Q_{N12} \tag{2–28}$$

$$0 = K_{N11} + Q_{N12}^T \tag{2–29}$$

$$\begin{aligned} 0 = {} & \dot{K}_{N22} - K_{N22}M^{-1}V_m - V_m^T M^{-1}K_{N22} + Q_{N22} \\ & - K_{N22}M^{-1}R_{N11}^{-1}M^{-1}K_{N22} - K_{N22}M^{-1}R_{N22}^{-1}M^{-1}P_{22} \\ & - P_{N22}M^{-1}R_{N22}^{-1}M^{-1}K_{N22} + P_{N22}M^{-1}R_{N22}^{-T}R_{N12}R_{N22}^{-1}M^{-1}P_{22}. \end{aligned} \tag{2–30}$$

Likewise, from Eq. 2–24, four similar simultaneous equations are generated as

$$0 = \dot{P}_{N11} - P_{N11}\alpha_1 - \alpha_1^T P_{N11} + L_{N11} \tag{2–31}$$

$$0 = P_{N11} + L_{N12} \tag{2–32}$$

$$0 = P_{N11} + L_{N12}^T \tag{2–33}$$

$$\begin{aligned} 0 = {} & \dot{P}_{N22} - P_{N22}M^{-1}V_m - V_m^T M^{-1}P_{N22} + L_{N22} \\ & - P_{N22}M^{-1}R_{N11}^{-1}M^{-1}K_{N22} - P_{N22}M^{-1}R_{N22}^{-1}M^{-1}P_{N22} \\ & - K_{N22}M^{-1}R_{N11}^{-1}M^{-1}P_{N22} + K_{N22}M^{-1}R_{N11}^{-T}R_{N21}R_{N11}^{-1}M^{-1}K_{22}. \end{aligned} \tag{2–34}$$

If $P_{N22}(t)$ and $K_{N22}(t)$ are selected as

$$P_{N22} = K_{N22} = M(q) \tag{2–35}$$

then the skew symmetry properties in Assumption 2-2 can be applied to Eqs. 2–30 and 2–34 to determine two constraints on the control gains

$$-R_{N11}^{-1} - 2R_{N22}^{-1} + R_{N22}^{-T}R_{N12}R_{N22}^{-1} + Q_{N22} = 0, \tag{2–36}$$

$$-2R_{N11}^{-1} - R_{N22}^{-1} + R_{N11}^{-T}R_{N21}R_{N11}^{-1} + L_{N22} = 0, \tag{2–37}$$

Since Q_N and L_N are constant matrices then K_{N11} and P_{N11} from Eqs. 2–28 and 2–32, respectively, must also be constant matrices (i.e. $\dot{P}_{N11} = \dot{K}_{N11} = 0$). It is evident from Eqs. 2–28, 2–29, 2–32, and 2–33 that the following relationships can be established

$$K_{N11} = -\frac{1}{2}\left(Q_{N12} + Q_{N12}^T\right), \tag{2–38}$$

$$P_{N11} = -\frac{1}{2}\left(L_{N12} + L_{N12}^T\right). \tag{2–39}$$

Two more constraints can be established by substituting Eqs. 2–38 and 2–39 into 2–27 and 2–31, respectively, then reducing the equations as

$$\begin{aligned} 0 &= Q_{N11} + \frac{1}{2}\left[\left(Q_{N12} + Q_{N12}^T\right)\alpha_1 + \alpha_1^T\left(Q_{N12} + Q_{N12}^T\right)\right], \\ 0 &= L_{N11} + \frac{1}{2}\left[\left(L_{N12} + L_{N12}^T\right)\alpha_1 + \alpha_1^T\left(L_{N12} + L_{N12}^T\right)\right]. \end{aligned}$$

Substituting Eqs. 2–14, 2–25, and 2–35 into Eq. 2–21 yields the Nash derived controller

$$u_{N1} = -R_{N11}^{-1}B_1^T K_N z = -R_{N11}^{-1}e_2. \tag{2–40}$$

The controller in Eq. 2–40 is subject to the other player's input, derived by substituting Eqs. 2–14, 2–26 , and 2–35 into Eq. 2–22, as

$$u_{N2} = -R_{N22}^{-1}B_2^T L_N z = -R_{N22}^{-1}e_{2.} \tag{2–41}$$

The weights (Q_N, L_N) imposed a penalty on the state vectors in the cost functions Eqs. 2–15 and 2–16 and the gain matrices R_{N11}^{-1} and R_{N22}^{-1} are subject to the following constraints

$$0 = -R_{N11}^{-1} - 2R_{N22}^{-1} + R_{N22}^{-T}R_{N12}R_{N22}^{-1} + Q_{N22} \tag{2–42}$$

$$0 = -2R_{N11}^{-1} - R_{N22}^{-1} + R_{N11}^{-T}R_{N21}R_{N11}^{-1} + L_{N22} \tag{2–43}$$

$$0 = \frac{1}{2}\left[\left(Q_{N12} + Q_{N12}^T\right)\alpha_1 + \alpha_1^T\left(Q_{12} + Q_{12}^T\right)\right] + Q_{N11} \tag{2–44}$$

$$0 = \frac{1}{2}\left[\left(L_{N12} + L_{N12}^T\right)\alpha_1 + \alpha_1^T\left(L_{N12} + L_{N12}^T\right)\right] + L_{N11}. \tag{2–45}$$

Based on the feedback Nash strategy, the derived controller in Eq. 2–40 minimizes the cost functional given by Eq. 2–15 and is subject to the second player's control input in Eq. 2–41 that minimizes the cost functional given by Eq. 2–16. To demonstrate optimality of the proposed controller, Hamiltonians were constructed in Eqs. 2–17 and 2–18 and the optimal control problem was formulated. The costate variables in Eqs. 2–25 and 2–26 were assumed to be solutions of Eqs. 2–19 and 2–20 and gain constraints were developed. If all constraints in Eqs. 2–42-2–45 are satisfied then the assumed solutions in Eqs. 2–25 and 2–26 satisfy Eqs. 2–18-2–20, and hence, are (sub)optimal.

Existence and Uniqueness of Nash Equilibrium Solution. Theorem 2-2 exploits an existence and uniqueness proof for the feedback Nash equilibrium solution that is well known in literature, however this proof demonstrates the existence and uniqueness for a non-stationary Nash feedback. If the mass inertia matrix $M(q)$ is constant then the prior analysis demonstrates one possible analytic solution for the DREs that would yielded a *stationary* feedback strategy for the players u_{N1} and u_{N2}, therefore existence and uniqueness of the Nash solution needs to be further investigated. For the game with an infinite-planning horizon, Proposition 3.6 in [62] gives sufficient conditions for the existence of a Nash equilibrium. The following definitions and theorems summarize the results.

Definition 2.1. Consider the system

$$\begin{aligned} \dot{z} &= Az + Bu \\ y &= Cz + Du. \end{aligned}$$

The system is called output stabilizable if there exists a state feedback $u = Fx$ such that the corresponding output $y_F = (C + DF)\, x$ converges to zero as $t \to \infty$.

Theorem 2.3. *Suppose that* (C_i, D_i) *are such that* $\left(C_1^T C_1, C_2^T C_2\right) = (Q_N, L_N)$, $C_1^T D_1 = C_2^T D_2 = 0$, *and* $\left(D_1^T D_1, D_2^T D_2\right) = (R_{11}, R_{22})$. *If there exists a pair of solutions* (K_N, P_N)

that satisfy the coupled algebraic Riccati equations

$$\begin{aligned} 0 &= K_N A + A^T K_N - K_N B_1 R_{N11}^{-1} B_1^T K_N - K_N B_2 R_{N22}^{-1} B_2^T P_N \\ &\quad + Q_N^T - P_N B_2 R_{N22}^{-1} B_2^T K_N, \end{aligned} \tag{2–46}$$

$$\begin{aligned} 0 &= \dot{P}_N + P_N A + A^T P_N - P_N B_1 R_{N11}^{-1} B_1^T K_N - P_N B_2 R_{N22}^{-1} B_2^T P_N \\ &\quad + L_N^T - K_N B_1 R_{N11}^{-1} B_1^T P_N, \end{aligned} \tag{2–47}$$

such that K_N is the smallest real positive semi-definite solution of Eq. 2–46 for a given P_N and P_N is the smallest real positive semi-definite solution of Eq. 2–47 for a given K_N, and if (K_N, P_N) are such that the systems

$$\left(A - B_2 R_{N22}^{-1} B_2^T P_N, B_1, C_1, D_1\right) \quad \textit{and} \quad \left(A - B_1 R_{N11}^{-1} B_1^T K_N, B_2, C_2, D_2\right)$$

are both output stabilizable, then the strategies (γ_1^*, γ_2^*) given by

$$\begin{aligned} u_{1N}^* &= -R_{N11}^{-1} B_1^T K_N z, \\ u_{2N}^* &= -R_{N22}^{-1} B_2^T P_N z, \end{aligned}$$

constitutes a feedback Nash equilibrium in a linear stationary strategy.

Proof. Similar to proof of Proposition 3.6 in [62]. □

Remark 2.3. Theorem 2-3 requires small real positive semi-definite solutions of the coupled Riccati equations; however, the theorem does not imply uniqueness of the equilibrium. Furthermore [62] uses a scalar example to illustrate the possible non-uniqueness of the linear stationary feedback Nash equilibria. A proof for the uniqueness of a linear stationary feedback Nash equilibrium remains an open problem.

2.4 RISE Feedback Control Development

In general, the bounded disturbance $\tau_d(t)$ and the nonlinear dynamics given in Eq. 2–10 are unknown, so the controller given in Eq. 2–11 can not be implemented. However, if the control input contains a method to identify and cancel these effects, then $z(t)$ will

converge to the state space model in Eq. 2–13 so that $u_1(t)$ and $u_2(t)$ each minimize a performance index. In this section, a control input is developed that exploits RISE feedback to identify the nonlinear effects and bounded disturbances to enable the system to asymptotically converge to the state space model $z(t)$.

To develop the control input, the error system in Eq. 2–8 is pre-multiplied by $M(q)$ and the expressions in Eqs. 2–1, 2–6, and 2–7 are used to obtain

$$Mr = -V_m e_2 + h + \tau_d + \alpha_2 M e_2 - (\tau_1 + \tau_2). \tag{2–48}$$

Based on the open-loop error system in Eq. 2–48, the control inputs are composed of the optimal controllers developed in Eqs. 2–40 and 2–41, plus a subsequently designed auxiliary control term $\mu(t) \in \mathbb{R}^n$ as

$$\tau_1 + \tau_2 \triangleq \mu - (u_{N1} + u_{N2}). \tag{2–49}$$

The closed-loop tracking error system can be developed by substituting Eq. 2–49 into Eq. 2–48 as

$$Mr = -V_m e_2 + h + \tau_d + \alpha_2 M e_2 + (u_{N1} + u_{N2}) - \mu. \tag{2–50}$$

To facilitate the subsequent stability analysis the auxiliary function $f_d(q_d, \dot{q}_d, \ddot{q}_d) \in \mathbb{R}^n$, which is defined as

$$f_d \triangleq M(q_d)\ddot{q}_d + V_m(q_d, \dot{q}_d)\dot{q}_d + G(q_d) + F(\dot{q}_d), \tag{2–51}$$

is added and subtracted to Eq. 2–50 to yield

$$Mr = -V_m e_2 + \bar{h} + f_d + \tau_d + u_{N1} + u_{N2} - \mu + \alpha_2 M e_2, \tag{2–52}$$

where $\bar{h}(t) \in \mathbb{R}^n$ is defined as

$$\bar{h} \triangleq h - f_d. \tag{2–53}$$

Substituting Eqs. 2–40 and 2–41 into Eq. 2–52, taking a time derivative and using the relationship in Eq. 2–8 yields

$$M\dot{r} = -\frac{1}{2}\dot{M}r + \tilde{N} + N_D - e_2 - (R_{N11}^{-1} + R_{N22}^{-1})r - \dot{\mu} \tag{2–54}$$

after strategically grouping specific terms. In Eq. 2–54, the unmeasurable auxiliary terms $\tilde{N}(q, \dot{q}, \ddot{q}, e_1, e_2, r)$, $N_D\left(q_d, \dot{q}_d, \ddot{q}_d, \dddot{q}_d\right) \in \mathbb{R}^n$ are defined as

$$\tilde{N} \triangleq -\dot{V}_m e_2 - V_m \dot{e}_2 - \frac{1}{2}\dot{M}r + \dot{\bar{h}} + \alpha_2 \dot{M} e_2 + \alpha_2 M \dot{e}_2 + e_2 + (R_{N11}^{-1} + R_{N22}^{-1})\alpha_2 e_2.$$

$$N_D \triangleq \dot{f}_d + \dot{\tau}_d.$$

Motivation for grouping terms into $\tilde{N}\left(\cdot\right)$ and $N_D\left(\cdot\right)$ comes from the subsequent stability analysis and the fact that the Mean Value Theorem, Assumption 2-3, Assumption 2-4, and Assumption 2-5 can be used to upper bound the auxiliary terms as

$$\left\|\tilde{N}(t)\right\| \leq \rho\left(\|y\|\right)\|y\|, \tag{2–55}$$

$$\|N_D\| \leq \zeta_1, \qquad \left\|\dot{N}_D\right\| \leq \zeta_2, \tag{2–56}$$

where $y(e_1, e_2, r) \in \mathbb{R}^{3n}$ is defined as

$$y \triangleq [e_1^T \quad e_2^T \quad r^T]^T, \tag{2–57}$$

the bounding function $\rho(\|y\|) \in \mathbb{R}$ is a positive globally invertible nondecreasing function, and $\zeta_i \in \mathbb{R}$ $(i = 1, 2)$ denote known positive constants. Based on Eq. 2–54, the control term $\mu(t)$ is designed as the generalized solution to

$$\dot{\mu}(t) \triangleq k_s r(t) + \beta_1 sgn(e_2), \tag{2–58}$$

where k_s, $\beta_1 \in \mathbb{R}$ are positive constant control gains. The closed loop error systems for $r(q, \dot{q}, \ddot{q}, t)$ can now be obtained by substituting Eq. 2–58 into Eq. 2–54 as

$$M\dot{r} = -\frac{1}{2}\dot{M}r + \tilde{N} + N_D - e_2 - (R_{N11}^{-1} + R_{N22}^{-1})r - k_s r - \beta_1 sgn(e_2). \tag{2–59}$$

2.5 Stability Analysis

Lemma 2.1. *Let* $O(e_2, r, t) \in \mathbb{R}$ denote the generalized solution to

$$\dot{O}(t) \triangleq -r^T(N_D(t) - \beta_1 sgn(e_2)), \qquad O(0) = \beta_1 \sum_{i=1}^{n} |e_{2i}(0)| - e_2(0)^T N_D(0) \tag{2–60}$$

where $e_{2i}(0)$ *denotes the* i*th element of the vector* $e_2(0)$*. Provided that* β_1 *is selected according to the sufficient conditions:*

$$\beta_1 > \zeta_1 + \frac{1}{\lambda_{min}(\alpha_2)}\zeta_2, \tag{2–61}$$

where ζ_1 *and* ζ_2 *are known positive constants defined in Eq. 2–56, then* $O(e_2, r, t) \geq 0$*.*

Proof. Refer to [111]. □

Theorem 2.4. *The controller given by Eqs. 2–40, 2–41, and 2–49 ensures that all system signals are bounded under closed-loop operation, and the tracking errors are semi-globally asymptotically regulated in the sense that*

$$\|e_1(t)\|, \|e_2(t)\|, \|r(t)\| \to 0 \qquad as \qquad t \to \infty, \tag{2–62}$$

provided the control gain k_s *in Eq. 2–58 is selected sufficiently large based on the initial conditions of the system,* β_1 *in Eq. 2–58 is selected sufficiently large, and* α_1*,* α_2 *are selected according to the sufficient conditions:*

$$\lambda_{min}(\alpha_1) > \frac{1}{2} \qquad \lambda_{min}(\alpha_2) > 1. \tag{2–63}$$

Furthermore, $u_{N1}(t)$ *and* $u_{N2}(t)$ *minimize Eqs. 2–15 and 2–16 subject to Eq. 2–13 provided the gain constraints given in Eqs. 2–42-2–45 are satisfied.*

Remark 2.4. The control gain α_1 can not be arbitrarily selected, rather it is calculated using a Lyapunov equation solver. Its value is determined based on the value of Q_N, L_N, R_{N11}, $R_{N21}$and R_{N22}. Therefore Q_N, L_N, R_{N11}, R_{N21}, and R_{N22} must be chosen such that Eq. 2–63 is satisfied.

Proof. Let $\mathcal{D} \subset \mathbb{R}^{3n+1}$ be a domain containing $\Phi(t) = 0$, where $\Phi(t) \in \mathbb{R}^{3n+1}$ is defined as

$$\Phi(t) \triangleq \left[y^T(t) \quad \sqrt{O(t)} \right]^T. \tag{2–64}$$

Let $V_L(\Phi, t) : \mathcal{D} \times (0, \infty) \to \mathbb{R}$ be a Lipschitz continuous regular positive definite function defined as

$$V_L(\Phi, t) \triangleq e_1^T e_1 + \frac{1}{2} e_2^T e_2 + \frac{1}{2} r^T M(q) r + O, \tag{2–65}$$

which satisfies the following inequalities:

$$U_1(\Phi) \leq V_L(\Phi, t) \leq U_2(\Phi) \tag{2–66}$$

provided the sufficient conditions introduced in Eqs. 2–63-2–61 are satisfied. In Eq. 2–66, the continuous positive definite functions $U_1(\Phi)$ and $U_2(\Phi) \in \mathbb{R}$ are defined as $U_1(\Phi) \triangleq \lambda_1 \|\Phi\|^2$ and $U_2(\Phi) \triangleq \lambda_2(q) \|\Phi\|^2$, where λ_1, $\lambda_2(q) \in \mathbb{R}$ are defined as

$$\lambda_1 \triangleq \frac{1}{2} \min\{1, m_1\} \qquad \lambda_2(q) \triangleq \max\left\{ \frac{1}{2} \bar{m}(q), 1 \right\},$$

where m_1, $\bar{m}(q)$ are introduced in Eq. 2–2. After taking the time derivative of Eq. 2–65, $\dot{V}_L(\Phi, t)$ can be expressed as

$$\dot{V}_L(\Phi, t) = 2 e_1^T \dot{e}_1 + e_2^T \dot{e}_2 + \frac{1}{2} r^T \dot{M}(q) r + r^T M(q) \dot{r} + \dot{O}.$$

From Eqs. 2–7, 2–59, 2–60, and 2–65, some of the differential equations describing the closed-loop system for which the stability analysis is being performed have discontinuous

right-hand sides

$$\dot{e}_1 = e_2 - \alpha_1 e_1 \quad \dot{e}_2 = r - \alpha_2 e_2 \tag{2–67}$$

$$M\dot{r} = -\frac{1}{2}\dot{M}(q)r + \tilde{N} + N_B - k_s r - \beta_1 sgn(e_2) - e_2 \tag{2–68}$$

$$\dot{O}(t) = -r^T(N_B - \beta_1 sgn(e_2)). \tag{2–69}$$

Let $f(\Phi, t) \in \mathbb{R}^{3n+1}$ denote the right-hand side of Eqs. 2–67-2–69. As described in [116–118], the existence of Filippov's generalized solution can be established for Eqs. 2–67-2–69. Note that $f(\Phi, t)$ is continuous except in the set $\{(\Phi, t) | e_2 = 0\}$. From [116–118], an absolute continuous Filippov solution $\Phi(t)$ exists almost everywhere (a.e.) so that $\dot{\Phi} \in K[f](y, t)$ *a.e.* Except for the points on the discontinuous surface $\{(\Phi, t) | e_2 = 0\}$, the Filippov set-valued map includes unique solutions. Under Filippov's framework, a generalized Lyapunov stability theory can be used ([119–121] for further details) to establish strong stability of the closed-loop system. The generalized time derivative of Eq. 2–65 exists (a.e.), and $\dot{V}_L(\Phi, t) \in^{a.e.} \dot{\tilde{V}}_L(\Phi, t)$ where

$$\dot{\tilde{V}}_L = \bigcap_{\xi \in \partial V_L(\Phi,t)} \xi^T K \begin{bmatrix} \dot{e}_1^T & \dot{e}_2^T & \dot{r}^T & \frac{1}{2}O^{-\frac{1}{2}}\dot{O} & 1 \end{bmatrix}^T,$$

where $\partial V_L(\Phi, t)$ is the generalized gradient of $V_L(\Phi, t)$ [119], and $K[\cdot]$ is defined as [120, 121]

$$K[f](\Phi) \triangleq \bigcap_{\delta>0} \bigcap_{\mu N=0} \bar{co} f(B(\Phi, \delta) - N),$$

where $\bigcap_{\mu N=0}$ denotes the intersection of all sets N of Lebesgue measure zero, $\bar{co}$ denotes convex closure, and $B(\Phi, \delta)$ represents a ball of radius δ around Φ. Since $V_L(\Phi, t)$ is a Lipschitz continuous regular function

$$\begin{aligned} \dot{\tilde{V}}_L &= \nabla V_L^T K \begin{bmatrix} \dot{e}_1^T & \dot{e}_2^T & \dot{r}^T & \frac{1}{2}O^{-\frac{1}{2}}\dot{O} & 1 \end{bmatrix}^T \\ &= \begin{bmatrix} 2e_1^T & e_2^T & r^T M & 2O^{\frac{1}{2}} & \frac{1}{2}r^T\dot{M}r \end{bmatrix} K \begin{bmatrix} \dot{e}_1^T & \dot{e}_2^T & \dot{r}^T & \frac{1}{2}O^{-\frac{1}{2}}\dot{O} & 1 \end{bmatrix}^T. \end{aligned}$$

Using calculus for $K[\cdot]$ from [121] and substituting the dynamics from Eqs. 2–7, 2–58, 2–59, and 2–69 yields

$$\dot{\tilde{V}}_L \leq r^T \tilde{N} - \left(k_s + \lambda_{\min}\left(R_{11}^{-1} + R_{22}^{-1}\right)\right)\|r\|^2 - \alpha_2 \|e_2\|^2 - 2\alpha_1 \|e_1\|^2 + 2e_2^T e_1,$$

where the fact that $(r^T(t) - r^T(t))_i SGN(e_{2i}) = 0$ is used (the subscript i denotes the i^{th} element), and $K[sgn(e_2)] = SGN(e_2)$ [121] such that $SGN(e_{2_i}) = 1$ if $e_{2_i}(t) > 0$, $SGN(e_{2_i}) = [1, -1]$ if $e_{2_i}(t) = 0$, $SGN(e_{2_i}) = -1$ if $e_{2_i}(t) < 0$. Based on the fact that $2e_2^T(t)e_1(t) \leq \|e_1(t)\|^2 + \|e_2(t)\|^2$, the expression in Eq. 2–55 can be used to upper bound $\dot{\tilde{V}}_L(t)$ using the squares of the components of $z(t)$ as

$$\dot{\tilde{V}}_L \leq -\lambda_3 \|y\|^2 - \left[k_s \|r\|^2 - \rho(\|y\|)\|r\|\|y\|\right] \tag{2–70}$$

where

$$\lambda_3 \triangleq \min\{2\alpha_1 - 1, \alpha_2 - 1, \lambda_{\min}\left(R_{N11}^{-1} + R_{N22}^{-1}\right)\};$$

hence, α_1, and α_2 must be chosen according to the sufficient condition in Eq. 2–63. After completing the squares for the terms inside the brackets in Eq. 2–70, the following expression is obtained:

$$\dot{\tilde{V}}_L \leq -\lambda_3 \|y\|^2 + \frac{\rho^2(\|y\|)\|y\|^2}{4k_s}. \tag{2–71}$$

The expression in Eq. 2–71 can be further upper bounded as

$$\dot{V}_L(\Phi, t) \leq -U(\Phi) = -c\|y\|^2 \qquad \forall \Phi \in \mathcal{D} \tag{2–72}$$

for some positive constant c, where

$$\mathcal{D} \triangleq \left\{\Phi \in \mathbb{R}^{3n+1} \mid \|\Phi\| \leq \rho^{-1}\left(2\sqrt{\lambda_3 k_s}\right)\right\},$$

where k_s is selected as $k_s > 0$ Larger values of k_s will expand the size of the domain $\mathcal{D}$. The result in Eq. 2–72 indicates that $\dot{V}_L(\Phi, t) \leq -U(\Phi)\ \forall\ \dot{V}_L(\Phi, t) \in \dot{\tilde{V}}_L(\Phi, t)$. The inequality in Eq. 2–72 can be used to show that $V_L(\Phi, t) \in \mathcal{L}_\infty$ in $\mathcal{D}$; hence, $e_1(t)$, $e_2(t)$, and $r(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$. Given that $e_1(t)$, $e_2(t)$, and $r(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$, standard analysis

methods can be used to prove that the control input and all closed-loop signals are bounded, and that $U(\Phi)$ is uniformly continuous in $\mathcal{D}$. Let $\mathcal{S} \subset \mathcal{D}$ denote the set defined as $\mathcal{S} \triangleq \left\{ \Phi(t) \in \mathcal{D} \mid U_2(\Phi(t)) < \lambda_1 \left(\rho^{-1} \left(2\sqrt{\lambda_3 k_s} \right) \right)^2 \right\}$. The region of attraction can be made arbitrarily large to include any initial conditions by increasing the control gain k_s (i.e., a semi-global type of stability result), and hence $c \left\| y(t) \right\|^2 \to 0$ and $\left\| e_1(t) \right\| \to 0$ as $t \to \infty$ $\forall y(0) \in \mathcal{S}$. Since $u_{N1}(t)$, $u_{N2}(t) \to 0$ as $e_2(t) \to 0$ (by Eq. 2–40), then Eq. 2–52 can be used to conclude that

$$\mu \to \bar{h} + f_d + \tau_{du} \quad \text{as} \quad r(t), e_2(t) \to 0. \tag{2–73}$$

The result in Eq. 2–73 indicates that the dynamics in Eq. 2–1 converge to the state-space system in Eq. 2–13. Hence, $u_{N1}(t)$, $u_{N2}(t)$ converge to optimal controllers that minimizes Eqs. 2–15 and 2–16, respectively, subject to Eq. 2–13 in the presence of structured disturbances; provided the gain constraints given in Eqs. 2–42-2–45 are satisfied. □

2.6 Simulation

To examine the performance of the Nash-derived controller developed in Eqs. 2–40, 2–41, and 2–49 a numerical simulation was performed. To illustrate the utility of the technique a model is described by the Euler-Lagrange dynamics as

$$\begin{aligned} \begin{bmatrix} \tau_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \tau_2 \end{bmatrix} = & \begin{bmatrix} p_1 + 2p_3c_2 & p_2 + p_3c_2 \\ p_2 + p_3c_2 & p_2 \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} \\ & + \begin{bmatrix} -p_3s_2\dot{q}_2 & -p_3s_2(\dot{q}_1 + \dot{q}_2) \\ p_3s_2\dot{q}_1 & 0 \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} \\ & + \begin{bmatrix} f_{d_1} & 0 \\ 0 & f_{d_2} \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} + \begin{bmatrix} \tau_{d_1} \\ \tau_{d_2} \end{bmatrix}, \end{aligned} \tag{2–74}$$

where $p_1 = 3.473\ kg \cdot m^2$, $p_2 = 0.196\ kg \cdot m^2$, $p_3 = 0.242\ kg \cdot m^2$, $f_{d_1} = 5.3\ Nm \cdot \sec$, $f_{d_2} = 1.1\ Nm \cdot \sec$, c_2 denotes $cos(q_2)$, s_2 denotes $sin(q_2)$ and τ_{d_1}, τ_{d_2} denote bounded

disturbances defined as

$$\tau_{d_1} = 10\ \sin(t) + 3.5\ \cos(3t)$$

$$\tau_{d_2} = 2.5\ \sin(2t) + 1.5\cos(t).$$

In the Nash strategy, player 1 is defined by $\tau_{N1} = \tau_1$ and player 2 is defied as $\tau_{N2} = \tau_2$. The objective of both players is to track a desired trajectory given as

$$q_{d_1} = q_{d_2} = 60\sin(2t)\left(1 - \exp\left(-0.01t^3\right)\right),$$

and the initial conditions were selected as

$$q_1(0) = q_2(0) = 14.3\ deg$$

$$\dot{q}_1(0) = \dot{q}_2(0) = 28.6\ deg/\sec.$$

The weighting matrices for both controllers were chosen as

$$Q_{N11} = diag\{5,5\} \qquad Q_{N12} = diag\{-5,-5\}$$

$$Q_{N22} = diag\{5,5\} \qquad L_{N12} = diag\{-5,-5\}$$

$$L_{N22} = diag\{10,10\} \qquad L_{N11} = diag\{5,5\},$$

which using the Nash constraints given in Eqs. 2–42-2–45 yield the Nash gains R_{N22}, R_{N11}, R_{N21}, and R_{N12} as

$$R_{N22} = diag\left\{\frac{1}{5}, \frac{1}{5}\right\} \qquad R_{N11} = diag\left\{\frac{1}{10}, \frac{1}{10}\right\}$$

$$R_{N12} = diag\left\{\frac{1}{25}, \frac{1}{25}\right\} \qquad R_{N21} = diag\left\{\frac{1}{100}, \frac{1}{100}\right\}.$$

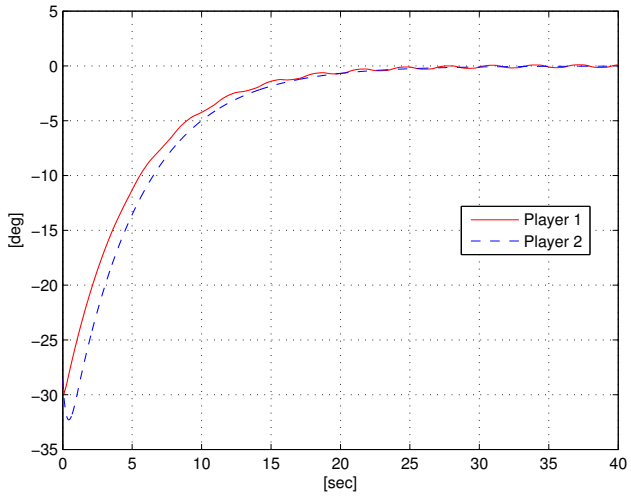

Figure 2-1. The simulated tracking errors for the RISE and Nash optimal controller.

The control gains for RISE control element were selected as

$$\alpha_1 = diag\{5, 5\}$$

$$\alpha_2 = diag\{15, 3.5\}$$

$$\beta_1 = diag\{15, 10\}$$

$$k_s = diag\{65, 25\}.$$

The tracking errors and the control inputs for the RISE and optimal controller are shown in Figure 2-1 and 2-2, respectively. To show that the RISE feedback identifies the nonlinear effects and bounded disturbances, a plot of the difference is shown in Figure 2-3. As this difference goes to zero, the dynamics in Eq. 2–1 converge to the state-space system in Eq. 2–13, and the controller solves the two player differential game. In addition, Figure 2-4 shows the convergence of the cost functionals for each player.

2.7 Summary

In this chapter, a novel approach for the design of a differential game-based feedback controller is developed for an Euler-Lagrange system subject to uncertainties and bounded disturbances. An optimal game-derived feedback component was used in conjunction with

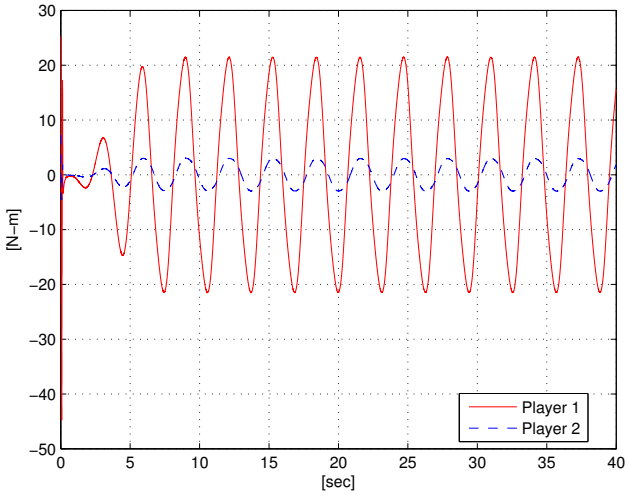

Figure 2-2. The simulated torques for the RISE and Nash optimal controller.

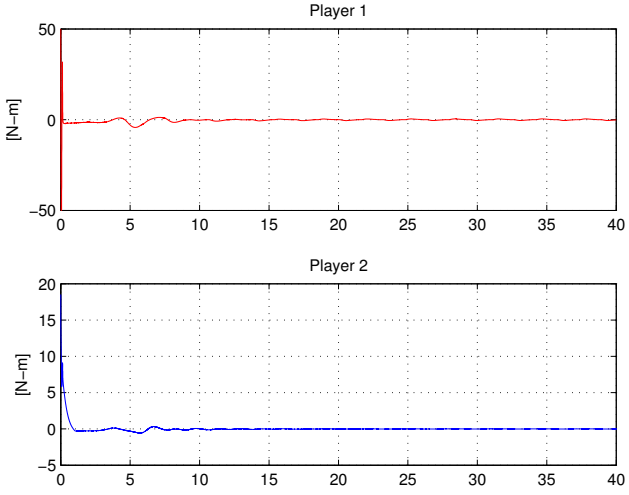

Figure 2-3. The difference between the RISE feedback and the nonlinear effect and bounded disturbances.

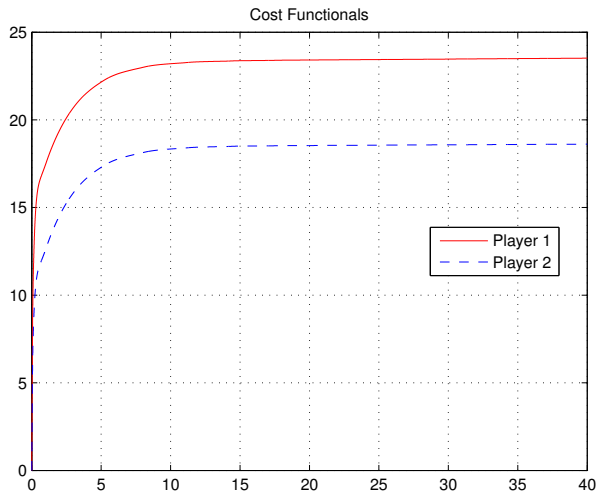

Figure 2-4. Cost functionals for u_1 and u_2 .

a RISE feedback component, which enables the generalized coordinates of the system to globally asymptotically track a desired time-varying trajectory despite uncertainty in the dynamics. Using a Lyapunov stability analysis and a feedback Nash game development, sufficient gain conditions were derived to ensure asymptotic stability while minimizing a cost function for the developed controllers.

CHAPTER 3
ASYMPTOTIC STACKELBERG OPTIMAL CONTROL DESIGN FOR AN UNCERTAIN EULER-LAGRANGE SYSTEM

Game theory establishes an optimal strategy for multiple players in either a cooperative or noncooperative manner where the objective is to reach an equilibrium state among the players. A Stackelberg game strategy involves a leader and a follower that follow a hierarchy relationship where the leader enforces its strategy on the follower. In this chapter, a general framework is developed for feedback control of an Euler Lagrange system using an open-loop non-zero sum Stackelberg differential game. A Robust Integral Sign of the Error (RISE) controller is used to cancel uncertain nonlinearities in the system and an open-loop Stackelberg game method is then applied to the residual uncertain nonlinear system to minimize cost functionals for each player. A Lyapunov analysis and simulation are provided to examine the stability and performance of the developed controllers.

3.1 Dynamic Model and Properties

This chapter considers the same dynamics presented in Chapter 2. In this formulation, player 1 is denoted as the follower $\tau_1 = \tau_F$ and player 2 is denoted as the leader $\tau_2 = \tau_L$, given as

$$M(q)\ddot{q} + V_m(q,\dot{q})\dot{q} + G(q) + F\left(\dot{q}\right) + \tau_d(t) = \tau_F(t) + \tau_L(t), \tag{3–1}$$

where the same assumption from Chapter 2 hold.

3.2 Error System Development

The control objective is to ensure that the generalized coordinates track a desired time-varying trajectory, denoted by $q_d(t) \in \mathbb{R}^n$, despite uncertainties in the dynamic model, while minimizing a given performance index. To quantify the tracking objective, a position tracking error, denoted by $e_1(q,t) \in \mathbb{R}^n$, is defined as

$$e_1 \triangleq q_d - q. \tag{3–2}$$

To facilitate the subsequent analysis, filtered tracking errors, denoted by $e_2(q, \dot{q}, t)$, $r(q, \dot{q}, \ddot{q}, t) \in \mathbb{R}^n$, are also defined as

$$e_2 \triangleq \dot{e}_1 + \alpha_1 e_1, \tag{3–3}$$

$$r \triangleq \dot{e}_2 + \alpha_2 e_2, \tag{3–4}$$

where $\alpha_1, \alpha_2 \in \mathbb{R}^{n \times n}$ are positive definite, constant, diagonal gain matrices. The filtered tracking error $r(q, \dot{q}, \ddot{q}, t)$ is not measurable since the expression in Eq. 3–4 depends on $\ddot{q}(t)$. The error systems are based on the assumption that the generalized coordinates of the Euler-Lagrange dynamics allow additive and not multiplicative errors. To develop a state-space model for the tracking errors in Eqs. 3–2 and 3–3, the inertia matrix is premultiplied to the time derivative of Eq.3–3, and substitutions are made from Eq. 3–1 and 3–2 to obtain

$$M\dot{e}_2 = -V_m e_2 + h + \tau_d - (\tau_L + \tau_F), \tag{3–5}$$

where the nonlinear function $h(q, \dot{q}, t) \in \mathbb{R}^n$ is defined as

$$h \triangleq M(\ddot{q}_d + \alpha_1 \dot{e}_1) + V_m(\dot{q}_d + \alpha_1 e_1) + G + F. \tag{3–6}$$

Under the (temporary) assumption that the dynamics in Eq. 3–1 are known, the control inputs can be designed as

$$\tau_L + \tau_F \triangleq h + \tau_d - (u_L + u_F) \tag{3–7}$$

where $u_F(t), u_L(t) \in \mathbb{R}^n$ are auxiliary control inputs for the follower and leader, respectively, that will be designed to minimize desired performance indices. Substituting Eq. 3–7 into Eq. 3–5 yields

$$M\dot{e}_2 = -V_m e_2 + u_L + u_F. \tag{3–8}$$

A state-space model for Eqs. 3–3 and 3–8 can now be developed as

$$\dot{z} = A(q, \dot{q})\, z + B_1(q)\, u_L + B_2(q) u_F, \tag{3–9}$$

where $A(q, \dot{q}) \in \mathbb{R}^{2n \times 2n}$, $B_1(q), B_2(q) \in \mathbb{R}^{2n \times n}$, and $z(e_i, e_2) \in \mathbb{R}^{2n}$ and are defined in Chapter 2. The state-space model in Eq. 3–9 is developed under the (temporary) assumption of exact knowledge of the dynamics. In the next section, cost functionals and controllers are developed for the residual uncertain nonlinear system in Eq. 3–9. The Stackelberg game-based controllers are then incorporated with the RISE control method that asymptotically reduces the original uncertain dynamics to the dynamics in Eq. 3–9.

3.3 Two Player Open-Loop Stackelberg Nonzero-Sum Differential Game

Stackelberg differential games provide a framework for systems that operate on different levels with a prescribed hierarchy of decisions. The two-player game is cast in two solution spaces: the leader and the follower. The follower tries to minimize its cost functional based on the decision from the leader, while the leader, who has insight into the follower's rationale, will define an input such that the leader and the follower's inputs will yield minimal cost functionals. The Stackelberg differential game for the system given in Eq. 3–9 can be formulated in an optimal control framework where the leader's input is $u_L(z)$, and the follower's input as $u_F(z)$. Each player in Eq. 3–9 has a cost functional $J_F(z, u_F, u_L), J_L(z, u_F, u_L) \in \mathbb{R}$ defined as

$$J_F = \frac{1}{2} \int_{t_0}^{\infty} \left(z^T Q z + u_F^T R_{11} u_F + u_L^T R_{12} u_L \right) dt \tag{3–10}$$

$$J_L = \frac{1}{2} \int_{t_0}^{\infty} \left(z^T N z + u_F^T R_{21} u_F + u_L^T R_{22} u_L \right) dt, \tag{3–11}$$

where $t_0 \in \mathbb{R}$ is the initial time, $Q, N \in \mathbb{R}^{2n \times 2n}$ are symmetric constant matrices defined as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix} \qquad N = \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^T & N_{22} \end{bmatrix},$$

where Q_{ij}, N_{ij}, $R_{ij} \in \mathbb{R}^{n \times n}$ are symmetric constant matrices for $i, j = 1, 2$. Based on the minimum principal [114], the Hamiltonian $H_F(z, u_F, u_L) \in \mathbb{R}$ of the follower is defined as

$$H_F = \frac{1}{2} \left(z^T Q z + u_F^T R_{11} u_F + u_L^T R_{12} u_L \right) + \lambda_F^T \left(A z + B_1 u_F + B_2 u_L \right), \tag{3–12}$$

where the optimal controller and costate equation of the follower are [114]

$$u_F = -\left(\frac{\partial H_F}{\partial u_F}\right)^T = -R_{11}^{-1}B_1^T\lambda_F \tag{3–13}$$

$$\dot{\lambda}_F = -\left(\frac{\partial H_F}{\partial z}\right)^T = -A^T\lambda_F - Q^T z. \tag{3–14}$$

Using Eq. 3–13, the leader's Hamiltonian $H_L(z, u_F, u_L) \in \mathbb{R}$ is defined as

$$\begin{aligned} H_L = &\frac{1}{2}\left(z^T N z + \lambda_F^T B_1 R_{11}^{-T} R_{21} R_{11}^{-1} B_1^T \lambda_F + u_L^T R_{22} u_L\right) \\ &+ \lambda_L^T\left(Az - B_1 R_{11}^{-1} B_1^T \lambda_F + B_2 u_L\right) + \psi^T \dot{\lambda}_F \end{aligned} \tag{3–15}$$

where the optimal controller and costate equations are defined as

$$u_L = -\left(\frac{\partial H_L}{\partial u_L}\right)^T = -R_{22}^{-1}B_2^T\lambda_L \tag{3–16}$$

$$\dot{\lambda}_L = -\left(\frac{\partial H_L}{\partial z}\right)^T = -N^T z - A^T\lambda_L + Q\psi \tag{3–17}$$

$$\begin{aligned} \dot{\psi} &= -\left(\frac{\partial H_L}{\partial \lambda_F}\right)^T \\ &= -B_1 R_{11}^{-T} R_{21} R_{11}^{-1} B_1^T \lambda_F + B_1 R_{11}^{-T} B_1^T \lambda_L + A\psi. \end{aligned} \tag{3–18}$$

The expressions derived in Eqs. 3–10-3–18 define the optimal control problem. The subsequent analysis aims at developing an expression for the costate variables $(\lambda_F(t), \lambda_L(t), \psi(t))$ which satisfy the costate equations $(\dot{\lambda}_F(t), \dot{\lambda}_L(t), \dot{\psi}(t))$ and can be implemented by the controllers $u_F(t)$ and $u_L(t)$. To this end, the subsequent development is based on the following assumed solutions for the costate variables

$$\lambda_F = Kz \tag{3–19}$$

$$\lambda_L = Pz \tag{3–20}$$

$$\psi = Sz, \tag{3–21}$$

where $K(t), P(t), S(t) \in \mathbb{R}^{2n \times 2n}$ are time-varying positive definite block diagonal matrices defined as

$$K = \begin{bmatrix} K_{11} & 0_{n\times n} \\ 0_{n\times n} & K_{22} \end{bmatrix} \qquad P = \begin{bmatrix} P_{11} & 0_{n\times n} \\ 0_{n\times n} & P_{22} \end{bmatrix} \qquad S = \begin{bmatrix} S_{11} & 0_{n\times n} \\ 0_{n\times n} & S_{22} \end{bmatrix}.$$

Given these assumed solutions, conditions/constraints are then developed to ensure these solutions satisfy Eqs. 3–12-3–18.

Substituting Eqs. 3–9, 3–13, 3–14, and 3–16-3–21 into the derivative of Eqs. 3–19-3–21 yields the three differential equations

$$0 = \dot{K} + KA + A^T K - KB_1 R_{11}^{-1} B_1^T K - KB_2 R_{22}^{-1} B_2^T P + Q^T \tag{3–22}$$

$$0 = \dot{P} + PA + A^T P - PB_1 R_{11}^{-1} B_1^T K - PB_2 R_{22}^{-1} B_2^T P + N^T - QS \tag{3–23}$$

$$\begin{aligned} 0 = {} & \dot{S} + SA - AS - SB_1 R_{11}^{-1} B_1^T K - SB_2 R_{22}^{-1} B_2^T P \\ & + B_1 \left(R_{11}^{-1}\right)^T R_{21} R_{11}^{-1} B_1^T K - B_1 \left(R_{11}^{-1}\right)^T B_1^T P, \end{aligned} \tag{3–24}$$

where Eqs. 3–22 and 3–23 are differential Riccati equations (DRE). Equations Eqs. 3–22-3–24 must be solved simultaneously to yield a control strategy for the leader and follower. The solutions to the DRE gains $K(t)$ and $P(t)$ correspond to $u_F(t)$ and $u_L(t)$ respectively, while $S(t)$ constrains the trajectory of $K(t)$ and $P(t)$. From the DRE in Eq. 3–23, four simultaneous equations are generated as

$$0 = \dot{P}_{11} - P_{11}\alpha_1 - \alpha_1^T P_{11} + N_{11} - Q_{11} S_{11} \tag{3–25}$$

$$0 = P_{11} + N_{12} - Q_{12} S_{22} \tag{3–26}$$

$$0 = P_{11} + N_{12}^T - Q_{12}^T S_{11} \tag{3–27}$$

$$\begin{aligned} 0 = {} & \dot{P}_{22} - P_{22} M^{-1} V_m - V_m^T M^{-1} P_{22} + N_{22} \\ & - Q_{22} S_{22} - P_{22} M^{-1} R_{11}^{-1} M^{-1} K_{22} - P_{22} M^{-1} R_{22}^{-1} M^{-1} P_{22}. \end{aligned} \tag{3–28}$$

If $P_{22}(t)$ and $K_{22}(t)$ are selected as

$$P_{22} = K_{22} = M(q), \tag{3–29}$$

then the skew symmetry properties in Assumption 2 can be applied to Eq. 3–28 to determine that

$$-R_{11}^{-1} - R_{22}^{-1} + N_{22} - Q_{22}S_{22} = 0, \tag{3–30}$$

which implies that S_{22} is a constant matrix; therefore, P_{11} must also be a constant matrix from Eq. 3–26. Four simultaneous equations are generated from the DRE in Eq. 3–22

$$0 = \dot{K}_{11} - K_{11}\alpha_1 - \alpha_1^T K_{11} + Q_{11} \tag{3–31}$$

$$0 = K_{11} + Q_{12} \tag{3–32}$$

$$0 = K_{11} + Q_{12}^T \tag{3–33}$$

$$\begin{aligned} 0 = \dot{K}_{22} - K_{22}M^{-1}V_m - V_m^T M^{-1}K_{22} + Q_{22} \\ - K_{22}M^{-1}R_{11}^{-1}M^{-1}K_{22} - K_{22}M^{-1}R_{22}^{-1}M^{-1}P_{22}. \end{aligned} \tag{3–34}$$

Substituting Eq. 3–29 into Eq. 3–34 yields

$$-R_{11}^{-1} - R_{22}^{-1} + Q_{22} = 0, \tag{3–35}$$

which when combined with Eq. 3–30 yields

$$Q_{22}(I_{n\times n} + S_{22}) - N_{22} = 0.$$

If N is chosen such that $N_{22} = -Q_{22}$, then S_{22} is constrained to be

$$S_{22} = -2I_{n\times n}. \tag{3–36}$$

From the Riccati equation given in Eq. 3–24 the following three simultaneous equations are generated

$$0 = \dot{S}_{11} - S_{11}\alpha_1 + \alpha_1 S_{11} \tag{3–37}$$

$$0 = S_{11} - S_{22} \tag{3–38}$$

$$\begin{aligned} 0 = {} & \dot{S}_{22} + M^{-1}V_m S_{22} - S_{22}M^{-1}R_{11}^{-1}M^{-1}K_{22} \\ & - S_{22}M^{-1}R_{22}^{-1}M^{-1}P_{22} + M^{-1}R_{11}^{-1}R_{21}R_{11}^{-1}K_{22} \\ & - M^{-1}R_{11}^{-1}M^{-1}P_{22} - S_{22}M^{-1}V_m. \end{aligned} \tag{3–39}$$

Substituting Eqs. 3–29 and 3–36 into Eq. 3–39 results in the constraint

$$R_{11}^{-1} + 2R_{22}^{-1} + R_{11}^{-1}R_{21}R_{11}^{-1} = 0. \tag{3–40}$$

In addition, substituting Eq. 3–36 into Eq. 3–38 yields

$$S_{11} = -2I_{n\times n}. \tag{3–41}$$

It is evident from Eqs. 3–26, 3–27, 3–32, 3–33, 3–36, and 3–41 that the following relationships can be established

$$K_{11} = -\frac{1}{2}\left(Q_{12} + Q_{12}^T\right) \tag{3–42}$$

$$P_{11} = -\frac{1}{2}\left(N_{12} + N_{12}^T\right) + 2K_{11}. \tag{3–43}$$

Another constraint can be established by substituting Eqs. 3–31, 3–41, and 3–43 into Eq. 3–25 and reducing the equation as

$$0 = N_{11} + \frac{1}{2}\left[\left(N_{12} + N_{12}^T\right)\alpha_1 + \alpha_1^T\left(N_{12} + N_{12}^T\right)\right].$$

Substituting Eqs. 3–9, 3–19, and 3–29 into Eq. 3–13 yields a Stackelberg derived controller given as

$$u_F = -R_{11}^{-1}B_1^T K z = -R_{11}^{-1}e_2. \tag{3–44}$$

The controller in Eq. 3–44 is subject to the leaders's input, derived by substituting Eqs. 3–9, 3–20, and 3–29 into Eq. 3–16, given as

$$u_L = -R_{22}^{-1} B_2^T P z = -R_{22}^{-1} e_2. \tag{3–45}$$

It is evident from Eqs. 3–35 and 3–40 that the gain matrices R_{11}^{-1} and R_{22}^{-1} are constrained by

$$Q_{22} + R_{22}^{-1} + R_{11}^{-1} R_{21} R_{11}^{-1} = 0. \tag{3–46}$$

In addition, the weights (Q, N) impose a penalty on the state vectors given in the cost functions Eqs. 3–10 and 3–11 and are subject to the following constraints

$$0 = N_{22} + Q_{22} \tag{3–47}$$

$$0 = \frac{1}{2}\left[\left(Q_{12} + Q_{12}^T\right)\alpha_1 + \alpha_1^T\left(Q_{12} + Q_{12}^T\right)\right] + Q_{11} \tag{3–48}$$

$$0 = \frac{1}{2}\left[\left(N_{12} + N_{12}^T\right)\alpha_1 + \alpha_1^T\left(N_{12} + N_{12}^T\right)\right] + N_{11}. \tag{3–49}$$

Based on the open-loop Stackelberg strategy, the derived controller in Eq. 3–44 minimizes the cost functional given by Eq. 3–10 and is subject to the leaders input in Eq. 3–45 that minimizes the cost functional given by Eq. 3–11. To demonstrate optimality of the proposed controller, Hamiltonians were constructed in Eqs. 3–12 and 3–15, and an optimal control problem was formulated. The costate variables in Eqs. 3–19-3–21 were assumed to be solutions to the costate equations Eqs. 3–14, 3–17 and 3–18 and gain constraints were developed. If all constraints in Eqs. 3–46-3–49 are satisfied then the assumed solutions in Eqs. 3–19-3–21 satisfy Eqs. 3–12-3–18, and hence are optimal with respect to the residual dynamics in Eq. 3–9.

Remark 3.1. Since the open-loop strategy for the leader is declared in advance for the entire game, if the follower minimizes its cost function, then it obtains the follower Stackelberg strategy which is the optimal reaction to the declared leader strategy. A drawback of any open-loop differential game approach is that, due to dynamic-inconsistency (also

called time-inconsistency [122]), the open-loop strategy does not satisfy the principle of optimality; i.e., if $u_1^{t_0}(x,t)$ has been found to be the open-loop strategy for the leader at $t = t_0$ and if after an interval of time $[t_0, t_1]$ a re-evaluation of the Stackelberg strategy is attempted it will, in general, be found that the resulting optimal Stackelberg strategy $u_1^{t_0}(x,t) \neq u_1^{t_1}(x,t)$. The open-loop Stackelberg strategy concept assumes a commitment by the leader to implement its announced strategy. This commitment is for the entire game, and if the actual interval of the game was different, the committed strategy generally would not coincide with the Stackelberg strategy for the new interval, but the leader would be obliged to use the non-optimal strategy (i.e., the game is subgame imperfect).

Existence and Uniqueness of Stackelberg Equilibrium Solution. The existence of unique Stackelberg equilibria was shown to be tied to the existence of solutions to certain non-symmetric Riccati equations, which are difficult to solve. In [63] a connection between solutions of a standard algebraic Riccati equation and a non-symmetric algebraic Riccati equation were given. The subsequent theorem, given as Theorem 3 in [63], utilizes the connection between the standard and non-symmetric Riccati equations to define existence and uniqueness.

Theorem 3.1. *If the convexity condition, given by*

$$\begin{aligned} R_{11} &\geq 0, \qquad R_{22} > 0, \qquad Q \geq 0, \\ R_{21} &\geq 0, \qquad N \geq 0, \end{aligned}$$

are satisfied and if there exists a stabilizing solution X to the non-symmetric algebraic Riccati equation

$$0 = X \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} + \begin{pmatrix} A^T & 0 \\ 0 & A^T \end{pmatrix} X + \begin{pmatrix} Q & 0 \\ N & -Q \end{pmatrix} - XGX \tag{3–50}$$

where

$$G \triangleq \begin{pmatrix} B_1 & B_2 & 0 \\ 0 & 0 & B_1 \end{pmatrix} \begin{pmatrix} R_{11} & 0 & 0 \\ 0 & R_{22} & 0 \\ R_{21} & 0 & R_{11} \end{pmatrix}^{-1} \begin{pmatrix} B_1^T & 0 \\ 0 & B_2^T \\ 0 & B_1^T \end{pmatrix},$$

then the unique Stackelberg exists and is given by

$$\begin{pmatrix} u_1^* \\ u_2^* \end{pmatrix} = - \begin{pmatrix} R_{11}^{-1} & 0 & 0 \\ 0 & R_{22}^{-1} & 0 \end{pmatrix} \begin{pmatrix} B_1^T & 0 \\ 0 & B_2^T \\ 0 & B_1^T \end{pmatrix} Xz, \tag{3–51}$$

where z is the solution of the closed-loop equation

$$\dot{z} = \left(\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} - GX \right) z, \quad z(0) = \begin{pmatrix} z_0 \\ 0 \end{pmatrix}.$$

Proof. See Theorem 3 in [63]. □

Its interesting to note that Theorem 3-1 does not depend on the solution to the Riccati equations, however given the existence of the stabilizing solution X to Eq. 3–50 at least one stabilizing solution (K, P, S) of the algebraic Riccati equations given by

$$0 = KA + A^T K - KB_1 R_{11}^{-1} B_1^T K - KB_2 R_{22}^{-1} B_2^T P + Q^T \tag{3–52}$$

$$0 = PA + A^T P - PB_1 R_{11}^{-1} B_1^T K - PB_2 R_{22}^{-1} B_2^T P + N^T - QS \tag{3–53}$$

$$\begin{aligned} 0 = {} & SA - AS - SB_1 R_{11}^{-1} B_1^T K - SB_2 R_{22}^{-1} B_2^T P \\ & + B_1 \left(R_{11}^{-1} \right)^T R_{21} R_{11}^{-1} B_1^T K - B_1 \left(R_{11}^{-1} \right)^T B_1^T P, \end{aligned} \tag{3–54}$$

exists. This follows from fact that $Im\begin{pmatrix} I_{n\times n} \\ X \end{pmatrix}$ is H_{st}-invariant, where $Im(\cdot)$ denotes the image operator, and H_{st} is the extended Hamiltonian defined as

$$H_{st} = \begin{pmatrix} A & -B_1 R_{11}^{-1} B_1^T & -B_2 R_{22}^{-1} B_2^T & 0 \\ -Q & -A^T & 0 & 0 \\ -N & 0 & -A^T & Q \\ 0 & -B_1 R_{11}^{-T} R_{21} R_{11}^{-1} B_1^T & B_1 R_{11}^{-1} B_1^T & A \end{pmatrix},$$

and contains an n-dimensional H_{st}-invariant subspace of the form $Im\left(I_{n\times n}, S^T, K^T, P^T\right)^T$, which defines the desired solution of Eqs. 3–52-3–54 [63]. If a stabilizing solution exists for the Riccati equations, imposing additional constraints such as: A is stable and every eigenvalue in A is (Q, A) unobservable, then the solution is unique. However, according to Chapter 2 of [123], satisfying the two constraints does not admit a stable solution to the algebraic Riccati equation for the leader and the non-symmetric coupled Riccati equation defined in Eq. 3–50.

3.4 RISE Feedback Control Development

In general, the bounded disturbance $\tau_d(t)$ and the nonlinear dynamics given in Eq. 3–6 are unknown, so the controller given in Eq. 3–7 can not be implemented. However, if the control input contains some method to identify and cancel these effects, then $z(t)$ will converge to the state space model in Eq. 3–9 so that $u_L(t)$ and $u_F(t)$ minimizes their respective performance index. In this section, a control input is developed that exploits RISE feedback to identify the nonlinear effects and bounded disturbances to enable $z(t)$ to asymptotically converge to the state space model.

The control input is defined the same as 2–49 in Chapter 2, however for this derivation player 1 is the follower $\tau_1 = \tau_F$ and player 2 is denoted as the leader $\tau_2 = \tau_L$. Using the control inputs, the closed loop error system can be derived as

$$M\dot{r} = -\frac{1}{2}\dot{M}r + \tilde{N} + N_D - e_2 - (R_{11}^{-1} + R_{22}^{-1})r - \dot{\mu} \tag{3–55}$$

In Eq. 3–55, the unmeasurable auxiliary terms $\tilde{N}(q, \dot{q}, \ddot{q}, e_1, e_2, r)$, $N_D\left(q_d, \dot{q}_d, \ddot{q}_d, \dddot{q}_d\right) \in \mathbb{R}^n$ are defined as

$$\tilde{N} \triangleq -\dot{V}_m e_2 - V_m \dot{e}_2 - \frac{1}{2}\dot{M} r + \dot{\bar{h}} + \alpha_2 \dot{M} e_2 + \alpha_2 M \dot{e}_2 + e_2 + (R_{11}^{-1} + R_{22}^{-1})\alpha_2 e_2,$$

$$N_D \triangleq \dot{f}_d + \dot{\tau}_d.$$

Motivation for grouping terms into $\tilde{N}$ and N_D comes from the subsequent stability analysis and the fact that the Mean Value Theorem, Assumption 3-3, Assumption 3-4, and Assumption 3-5 can be used to upper bound the auxiliary terms as

$$\left\| \tilde{N}(t) \right\| \leq \rho\left(\|y\|\right) \|y\|, \tag{3–56}$$

$$\|N_D\| \leq \zeta_1, \qquad \left\| \dot{N}_D \right\| \leq \zeta_2, \tag{3–57}$$

where $y(e_1, e_2, r) \in \mathbb{R}^{3n}$ is defined as

$$y(t) \triangleq [e_1^T \quad e_2^T \quad r^T]^T, \tag{3–58}$$

the bounding function $\rho(\|y\|) \in \mathbb{R}$ is a positive globally invertible nondecreasing function, and $\zeta_i \in \mathbb{R}$ $(i = 1, 2)$ denote known positive constants. Based on Eq. 3–55, the control term $\mu(t)$ is designed as the generalized solution to

$$\dot{\mu}(t) \triangleq k_s r(t) + \beta_1 sgn(e_2), \tag{3–59}$$

where k_s, $\beta_1 \in \mathbb{R}$ are positive constant control gains. The closed loop error systems for $r(t)$ can now be obtained by substituting Eq. 3–59 into Eq. 3–55 as

$$M\dot{r} = -\frac{1}{2}\dot{M} r + \tilde{N} + N_D - e_2 - (R_{11}^{-1} + R_{22}^{-1}) r - k_s r - \beta_1 sgn(e_2). \tag{3–60}$$

3.5 Stability Analysis

Lemma 3.1. *Let $O(e_2, r, t) \in \mathbb{R}$ denote the generalized solution to*

$$\dot{O} \triangleq -r^T(N_D - \beta_1 sgn(e_2)) \qquad O(0) = \beta_1 \sum_{i=1}^{n} |e_{2i}(0)| - e_2(0)^T N_D(0) \tag{3–61}$$

where $e_{2i}(0)$ denotes the i^{th} element of the vector $e_2(0)$. Provided β_1 is selected according to the following sufficient condition:

$$\beta_1 > \zeta_1 + \frac{1}{\lambda_{min}(\alpha_2)}\zeta_2, \tag{3–62}$$

where ζ_1 and ζ_2 are known positive constants defined in Eq. 3–57, then $O(e_2, r, t) \geq 0$.

Proof. See [111]. □

Theorem 3.2. *The controller given by Eqs. 3–44, 3–45, and 3–59 ensures that all system signals are bounded under closed-loop operation, and the tracking errors are semi-globally asymptotically regulated in the sense that*

$$\|e_1(t)\|, \|e_2(t)\|, \|r(t)\| \to 0 \qquad as \qquad t \to \infty \tag{3–63}$$

provided the control gain k_s in Eq. 3–59 is selected sufficiently large based on the initial conditions of the system, β_1 in Eq. 3–59 is selected according to Eq. 3–62, and α_1, α_2 are selected according to the sufficient conditions

$$\lambda_{min}(\alpha_1) > \frac{1}{2} \qquad \lambda_{min}(\alpha_2) > 1. \tag{3–64}$$

Furthermore, $u_F(t)$ and $u_L(t)$ minimize Eqs. 3–10 and 3–11 subject to 3–9 provided the gain constraints given in Eqs. 3–46-3–49 are satisfied.

Remark 3.2. The control gain α_1 can not be arbitrarily selected, rather it is calculated using a Lyapunov equation solver. Its value is determined based on the value of Q, N, R_{11}, R_{21} and R_{22}. Therefore Q, N, R_{11}, R_{21} and R_{22} must be chosen such that Eq. 3–64 is satisfied.

Proof. Refer to Theorem 2-4 from Chapter 2. □

Remark 3.3. Similar to LQR design, the state weights (Q, N) and input weight R are designed to regulate the states and inputs to a desired behavior, respectively. The gain constraints in Eqs. 3–46-3–49 provide a general framework for implementing the controller. The weights Q and N penalize the state $z(t)$ and can be chosen sufficiently large to yield desirable tracking error while the the leader's control input weight R_{22} can be chosen sufficiently large to yield desirable controller bandwidth. The follower's control input weight R_{11} is then generated using the chosen gains Q, N, and R_{22} and the constraints in Eqs. 3–46-3–49.

3.6 Simulation

To examine the performance of the Stackelberg-derived controller proposed in Eqs. 3–44, 3–45, and 3–59, a numerical simulation was performed. To illustrate the utility of the technique a model is described by the Euler-Lagrange dynamics, defined in Chapter 2, are considered. For the Stackelberg strategy, the follower input is $\tau_F = \tau_1$ and the leader input is $\tau_L = \tau_2$. In this framework the inertial and Coriolis effects of the leader are seen as a disturbance to the tracking objective of the follower. In both strategies, the objective of both players is to track a desired trajectory given as

$$q_{d_1} = q_{d_2} = 60 \sin(2t) \left(1 - \exp\left(-0.01t^3\right)\right),$$

and the initial conditions were selected as

$$q_1(0) = q_2(0) = 14.3 \; deg$$
$$\dot{q}_1(0) = \dot{q}_2(0) = 28.6 \; deg/\sec.$$

The weighting matrices for both controllers were chosen as

$$Q_{11} = diag\{5, 5\} \qquad Q_{12} = diag\{-5, -5\}$$
$$Q_{22} = diag\{-5, -5\} \qquad L_{12} = diag\{-5, -5\}$$
$$L_{22} = diag\{5, 5\} \qquad L_{11} = diag\{5, 5\},$$

which using the Stackelberg constraints given in Eqs. 3–46-3–48 yield the values Stackelberg gains R_{22}, R_{11} and R_{21}, as

$$R_{22} = diag\left\{\frac{4}{11}, \frac{4}{11}\right\} \qquad R_{11} = diag\left\{\frac{1}{15}, \frac{1}{15}\right\},$$
$$R_{21} = diag\left\{\frac{1}{100}, \frac{1}{100}\right\}.$$

The control gains for RISE control element were selected as

$$\alpha_1 = diag\{5, 5\}$$
$$\alpha_2 = diag\{15, 3.5\}$$
$$\beta_1 = diag\{15, 10\}$$
$$k_s = diag\{65, 25\}.$$

The tracking errors and the control inputs for the RISE and optimal controller are shown in Figure 3-1 and 3-2, respectively. To show that the RISE feedback identifies the nonlinear effects and bounded disturbances, a plot of the difference is shown in Figure 3-3. As this difference goes to zero, the dynamics in Eq. 3–1 converge to the state-space system in Eq. 3–9, and the controller solves the two player differential game. In addition, Figure 3-4 shows the convergence of the cost functionals for each player.

3.7 Summary

In this chapter, a novel approach for the design of a Stackelberg-based controller was proposed for a nonlinear Euler-Lagrange system subject to parametric uncertainty and bounded disturbances. Stackelberg game methods are used to develop tracking controllers

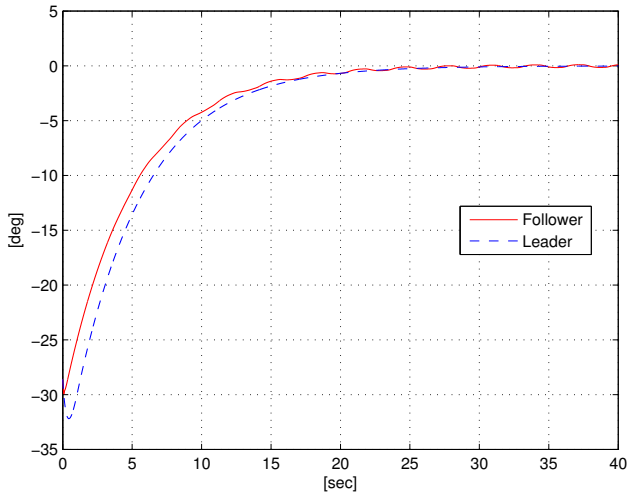

Figure 3-1. The simulated tracking errors for the RISE and Stackelberg optimal controller.

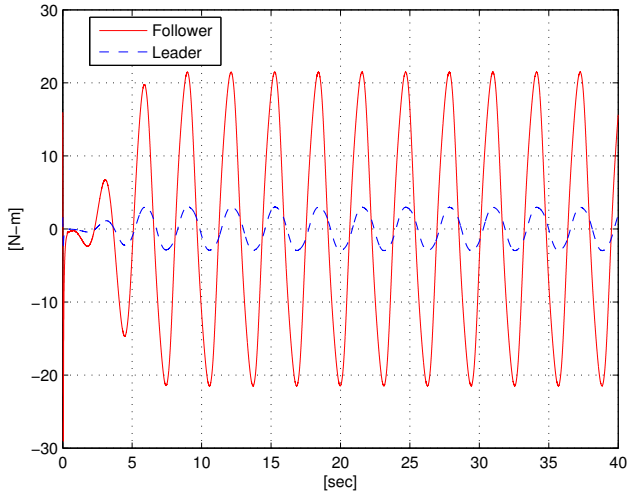

Figure 3-2. The simulated torques for the RISE and Stackelberg optimal controller.

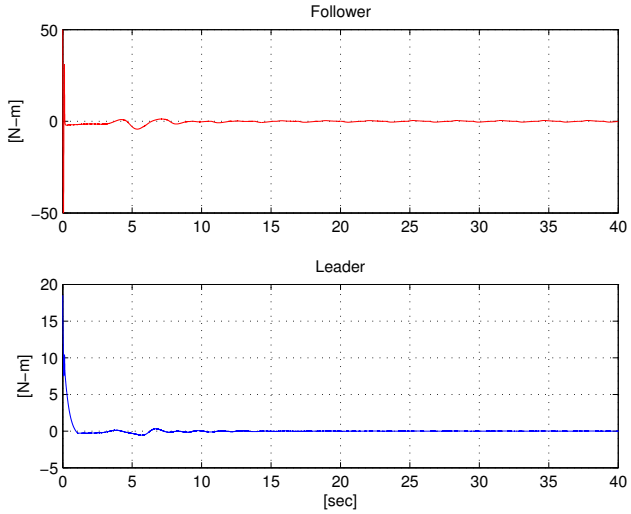

Figure 3-3. The difference between the RISE feedback and the nonlinear effect and bounded disturbances.

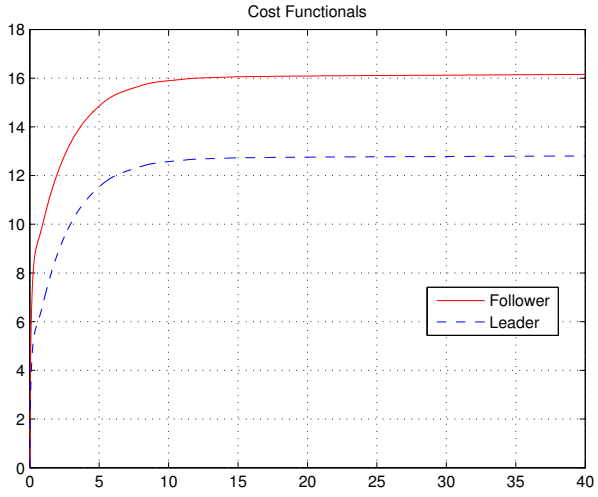

Figure 3-4. Cost functionals for the leader and follower.

that minimize cost functionals constrained by a residual uncertain nonlinear system with multiple inputs. A Lyapunov-based stability analysis is used to prove semi-global asymptotic tracking of the resulting controller. The inclusion of the RISE structure is an enabling method to allow the analytical development of a controller that asymptotically minimizes cost functionals in a Stackelberg game for the uncertain nonlinear continuous-time Euler-Lagrange system. However, the contribution of the implicit learning RISE structure is not included in the cost functional, yielding a (sub)optimal result.

CHAPTER 4
APPROXIMATE TWO PLAYER ZERO-SUM GAME SOLUTION FOR AN UNCERTAIN CONTINUOUS NONLINEAR SYSTEM

In recent work [102], an online approximate solution method is developed based on an approximation of the HJB for the (one player) infinite horizon optimal control problem of a continuous-time nonlinear systems with partially known dynamics. This approximate optimal adaptive controller uses two adaptive structures, a critic to approximate for the value (cost) function and an actor to approximate the control policy. In addition, a DNN is used to robustly identify the system parameters. The two adaptive structures are tuned simultaneously online to learn the solution of the HJB equation and the optimal policy. This chapter generalizes the method given in [102] and solves the two player zero-sum game problem for nonlinear continuous-time systems with partially known dynamics. This chapter presents an optimal adaptive control method that converges online to the solution to the two player differential game. The HJI approximation algorithm considers a two actor and one critic NNs architecture, where the actor and critic use gradient and least squares-based update laws, respectively, to minimize the Bellman error, which is the difference between the exact and the approximate HJI equations. An DNN identifier learns the system dynamics based on online gradient-based weight tuning laws, while a RISE term robustly accounts for the function reconstruction errors, guaranteeing asymptotic estimation of the state and the state derivative. The online estimation of the state derivative allows the ACI architecture to be implemented without knowledge of system drift dynamics. The parameter update laws tune the critic and actor neural networks online and simultaneously to converge to the solution to the HJI equation and the saddle point policies, while also guaranteeing closed-loop stability. These policies guarantee UUB tracking error for the closed-loop system.

4.1 Two Player Zero-Sum Differential Game

Consider the nonlinear time-invariant affine in the input dynamic system given by

$$\begin{aligned} \dot{x} &= f(x) + g_1(x)\,u_1(x) + g_2(x)\,u_2(x), \\ y &= h(x) \\ z &= \begin{bmatrix} y^T & u_1^T \end{bmatrix}^T \end{aligned} \tag{4–1}$$

where $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state vector, $u_1(x), u_2(x) \in \mathcal{U} \subseteq \mathbb{R}^m$ are the control inputs, and $f(x) \in \mathbb{R}^n$, and $g_1(x), g_2(x) \in \mathbb{R}^{n\times m}$ are the drift, and input matrices, respectively. Assume that $f(x)$ and $g_1(x)$ and $g_2(x)$ are Lipschitz continuous and that $f(0) = 0$ so that $x = 0$ is an equilibrium point for Eq. 4–1.

Bounded L_2 Gain Problem. The objective of the bounded L_2 gain control problem is to design a control input policy $u_1(x)$ such that

$$\int_0^\infty \|z\|^2 d\tau \equiv \int_0^\infty \left(h^T h + u_1 R u_1\right) d\tau \leq \gamma^2 \int_0^\infty \|u_2\|^2 d\tau,$$

for a given $\gamma > 0$; where $R = R^T > 0$, for all $u_2 \in L_2[0,\infty)$ when $x(0) = 0$. The H_∞ control problem is interested in determining the smallest $\gamma > 0$, known as γ^*, such that the bounded L_2 gain control problem has a solution [65]. This chapter is not interested in the H_∞ control objectives, rather it is assumed that γ is prescribed a priori such that $\gamma \geq \gamma^* \geq 0$. The value function [124] is given as

$$V(x, u_1, u_2) = \int_t^\infty \left(h^T h + u_1^T R u_1 - \gamma^2 ||u_2||^2\right) d\tau,$$

where $R = R^T \in \mathbb{R}^{m\times m}$ is positive definite. A differential equivalent to the value function is the nonlinear Lyapunov-like equation

$$0 = h^T h + u_1^T R u_1 - \gamma^2 ||u_2||^2 + \nabla V\left(f(x) + g_1(x)\,u_1(x) + g_2(x)\,u_2(x)\right), \tag{4–2}$$

where $\nabla V \triangleq \frac{\partial V}{\partial x} \in \mathbb{R}^{n\times 1}$ is the gradient of the value function. For admissible policies u_1, a solution $V(x) \geq 0$ to Eq. 4–2 is the value for a given $u_2 \in L_2[0,\infty)$.

Two Player Zero-Sum Game. For the two player zero-sum differential game, the infinite-horizon scalar value or cost functional $V(x(t), u_1, u_2)$ associated with the control policies $\{u_1 = u_1(x(s)); s \geq t\}$ and $\{u_2 = u_2(x(s)); s \geq t\}$ can be defined as

$$V(x) = \min_{u_1}\max_{u_2} \int_t^\infty r(x(s), u_1(s), u_2(s))\, ds, \tag{4–3}$$

where t is the initial time, and $r(x, u_1, u_2) \in \mathbb{R}$ is the local cost for the state, and controls, defined as

$$r = Q(x) + u_1^T R u_1 - \gamma^2 u_2^T u_2. \tag{4–4}$$

In this differential game, $u_1(x)$ is the minimizing player, and $u_2(x)$ is the maximizing player. This two player optimal control problem has a unique solution if the Nash condition holds

$$\min_{u_1}\max_{u_2} V(x(0), u_1, u_2) = \max_{u_2}\min_{u_1} V(x(0), u_1, u_2).$$

The objective of the optimal control problem is to find feedback policies [41] ($u_1^* = u_1(x)$ and $u_2^* = u_2(x)$), such that the cost in Eq. 4–3 associated with the system in Eq. 4–1 is minimized [114]. Assuming the value functional is continuously differentiable, Bellman's principle of optimality can be used to derive the following optimality condition

$$0 = \min_{u_1}\max_{u_2} \left[\nabla V (f(x) + g_1(x) u_1(x) + g_2(x) u_2(x)) + r(x, u_1, u_2)\right], \tag{4–5}$$

which is a nonlinear PDE, also called the HJI equation. Given a solution $V^*(x) \geq 0$ to the HJI, the local cost given in Eq. 4–4 can be used to form the algebraic expressions for the optimal control and disturbance inputs from Eq. 4–5 as

$$u_1^* = -\frac{1}{2} R^{-1} g_1^T(x) \nabla V^* \tag{4–6}$$

$$u_2^* = \frac{1}{2\gamma^2} g_2^T(x) \nabla V^*. \tag{4–7}$$

The closed form expression for the optimal control and disturbance in Eqs. 4–6 and 4–7, respectively, obviates the need to search for a feedback policy that minimize the value function; however, the solution $V^*(x)$ to the HJI equation given in Eq. 4–5 is required. The HJI equation in Eq. 4–5, can be rewritten by substituting for the local cost in Eq. 4–4 and the optimal control policies in Eqs. 4–6 and 4–7, as

$$\begin{aligned} 0 &= Q(x) + \nabla V^* f(x) - \frac{1}{4}\nabla V^* g_1(x) R^{-1} g_1^T(x) \nabla V^* \\ &\quad + \frac{1}{4\gamma^2}\nabla V^* g_2(x) g_2^T(x) \nabla V^* \qquad V^*(0) = 0. \end{aligned} \tag{4–8}$$

Since the HJI equation is troublesome to solve in general, this chapter considers an approximate solution. The HJI in Eq. 4–8 may have more than one nonnegative definite solution. A nonnegative definite solution V_a is such that there exists no other nonnegative definite solution V such that $V_a(x) \geq V(x) \geq 0$. In [41], the system is in Nash equilibrium with a value given as $V_a(x(0))$ and a saddle point equilibrium solution (u_1^*, u_2^*) among strategies in $L_2[0,\infty)$, if V_a is smooth and a minimal solution to the HJI and the system is zero state observable. Moreover, the closed-loop systems $f(x) + g_1 u_1^* + g_2 u_2^*$ and $f(x) + g_1 u_1^*$ are locally asymptotically stable. It is proven in [41], that the minimum nonnegative definite solution to the HJI is the unique solution for which the closed-loop system $f(x) + g_1 u_1^* + g_2 u_2^*$ is asymptotically stable. In [65] it was shown that the HJI equation has a local smooth solution $V(x)$ if the system $f(x) + g_1 u_1$ is locally asymptotically stable and $u_1(x)$ yields the L_2 gain of Eq. 4–1 $\leq \gamma$. From this it can be shown that $V_a(x)$ is also the minimal nonnegative solution to the HJI. The work in [65] also shows that for a given γ, where $V(x) \geq 0$ is smooth and is the solution to Eq. 4–8 and the system in Eq. 4–1 is zero state observable, then the system in Eq. 4–1 has a L_2 gain $\leq \gamma$ and the optimal control u_1^* in Eq. 4–6 solves the L_2 gain problem and yields the equilibrium point locally asymptotically stable. Moreover, yielding the optimal control as $u_1^*(t) \in L_2[0,\infty)$. It is evident that both the L_2 gain problem and the zero-sum game problem are dependent on the solution to the HJI.

Existence of solution to the HJI. While in general global solutions to the HJI in Eq. 4–8 may not exist, a local existence proof was given in [65]. The work in [65] proposes that for a given γ, if the system is zero state observable, and there exists a control policy $u_1(x)$ such that locally the system has a L_2 gain $\leq \gamma$ and the system is asymptotically stable, then there is a neighborhood $\Omega_x \in \mathbb{R}^n$ of the origin on which there exists a smooth solution $V(x) \geq 0$ to the HJI equation in Eq. 4–8. Furthermore, the control yields the L_2 gain $\leq \gamma$ for all trajectories originating at the origin and remaining inside Ω_x. Moreover, if they do, they may not be smooth. For a discussion on viscosity solutions to the HJI, see [55, 125].

4.2 HJI Approximation Algorithm

This chapter generalizes the ACI approximation architecture to solve the two player zero-sum game for Eq. 4–8. The ACI architecture eliminates the need for exact model knowledge and utilizes a DNN to robustly identify the system, a critic NN to approximate the value function, and an actor NN to find a control policy which minimizes the value functions. This section introduces the ACI architecture for the two player game, and subsequent sections give details of the design for the two player zero-sum game solution.

The Hamiltonian $H(x, \nabla V, u_1, u_2)$ of the system in Eq. 4–1 can be defined as

$$H = r + \nabla V F_u, \tag{4–9}$$

where ∇V is the Jacobian of the value function $V(x)$, $F_u(x, u_1, u_2) \triangleq f(x) + g_1 u_1 + g_2 u_2 \in \mathbb{R}^n$ denotes the system dynamics, and $r(x, u_1, u_2) \triangleq Q(x) + u_1^T R u_1 - \gamma^2 u_2^T u_2$ denotes the local cost. The optimal policy in Eq. 4–6 and the associated value function $V^*(x)$ satisfy the HJI equation

$$H(x, \nabla V^*, u_1^*, u_2^*) = r(x, u_1^*, u_2^*) + \nabla V^* F_{u^*} = 0. \tag{4–10}$$

Replacing the optimal Jacobian ∇V^* and optimal control policy u_1^* and disturbance input u_2^* by estimates $\nabla \hat{V}$, $\hat{u}_1$, and $\hat{u}_2$ respectively, yields the approximate HJI equation

$$H\left(x, \nabla \hat{V}, \hat{u}_1, \hat{u}_2\right) = r\left(x, \hat{u}_1, \hat{u}_2\right) + \nabla \hat{V} F_{\hat{u}}. \tag{4–11}$$

It is evident that the approximate HJI in Eq. 4–11 is dependent on the complete knowledge of the system. To overcome this limitation, an online system identifier replaces the system dynamics which modifies the approximate HJI in Eq. 4–11,

$$H\left(x, \hat{x}, \nabla \hat{V}, \hat{u}_1, \hat{u}_2\right) = r\left(x, \hat{u}_1, \hat{u}_2\right) + \nabla \hat{V} \hat{F}_{\hat{u}}, \tag{4–12}$$

where $\hat{F}_{\hat{u}}$ is an approximation of the system dynamics $F_{\hat{u}}$. The error between the optimal and approximate HJI equations in Eqs. 4–10 and 4–12, respectively, yields the Bellman residual error $\delta_{hjb}\left(x, \hat{x}, \hat{u}_1, \hat{u}_2, \nabla \hat{V}\right)$ defined as

$$\delta_{hjb} \triangleq H\left(x, \hat{x}, \nabla \hat{V}, \hat{u}_1, \hat{u}_2\right) - H\left(x, \nabla V^*, u_1^*, u_2^*\right). \tag{4–13}$$

However since $H\left(x, \nabla V^*, u_1^*, u_2^*\right) = 0$ then the Bellman residual error can be defined in a measurable form as

$$\delta_{hjb} = H\left(x, \hat{x}, \nabla \hat{V}, \hat{u}_1, \hat{u}_2\right).$$

The objective is to update both $\hat{u}_1$ and $\hat{u}_2$ (actors) and $\hat{V}$ (critic) simultaneously, based on the minimization of the Bellman residual error δ_{hjb}. All together the actors $\hat{u}_1$ and $\hat{u}_2$, the critic $\hat{V}$ and the identifier $\hat{F}_{\hat{u}}$ constitute the ACI architecture. To facilitate the subsequent analysis the following assumptions are given.

Assumption 4.1. *Given a continuous function $h : \mathbb{S} \to \mathbb{R}^n$, where $\mathbb{S}$ is a compact simply connected set, there exists ideal weights W,V such that the function can be represented by a NN as*

$$\hat{h}\left(x\right) = W^T \sigma\left(V^T x\right) + \varepsilon\left(x\right),$$

where $\sigma\left(\cdot\right)$ is the nonlinear activation function and $\varepsilon\left(x\right)$ is the function reconstruction error.

Assumption 4.2. *The NN activation function $\sigma(\cdot)$ and their time derivative $\sigma'(\cdot)$ with respect to its argument is bounded i.e. $||\sigma|| \leq \bar{\sigma}$ and $||\sigma'|| \leq \bar{\sigma'}$.*

Assumption 4.3. *The ideal NN weights are bounded by a known positive constant [126] i.e. $||W|| \leq \bar{W}$ and $||V|| \leq \bar{V}$.*

Assumption 4.4. *The NN function reconstruction errors are and its derivative is bounded [126], i.e. $||\varepsilon|| \leq \bar{\varepsilon}$ and $||\varepsilon'|| \leq \bar{\varepsilon'}$.*

4.3 System Identification

For the dynamics given in Eq. 4–1, the following assumptions about the system will be utilized in the subsequent development.

Assumption 4.5. *The input matrices $g_1(x)$ and $g_2(x)$ are known and bounded i.e. $\|g_1\| \leq \bar{g}_1$ and $\|g_2\| \leq \bar{g}_2$ where $\bar{g}_1$ and $\bar{g}_2$ are known constants.*

Assumption 4.6. *The inputs u_1 and u_2 are bounded i.e. $u_1, u_2 \in \mathcal{L}_\infty$.*

Using Assumption 4-1, the nonlinear system in Eq. 4–1 can be represented using a multi-layer NN as

$$\dot{x} = F_u(x, u_1, u_2) = W_f^T \sigma_f\left(V_f^T x\right) + \varepsilon_f(x) + g_1(x)\, u_1 + g_2(x)\, u_2, \tag{4–14}$$

where $W_f \in \mathbb{R}^{N_f+1\times n}$, $V_f \in \mathbb{R}^{n\times N_f}$ are unknown ideal NN weight matrices with N_f representing the neurons in the output layers. The activation function is given by $\sigma_f = \sigma\left(V_f^T x\right) \in \mathbb{R}^{N_f}$, and $\varepsilon_f(x) \in \mathbb{R}^n$ is the function reconstruction error in approximating the function $f(x)$. The proposed multi-layer dynamic neural network (MLDNN) used to identify the system in Eq. 4–1 is

$$\dot{\hat{x}} = \hat{F}_u(x, \hat{x}, u_1, u_2) = \hat{W}_f^T \hat{\sigma}_f + g_1(x)\, u_1 + g_2(x)\, u_2 + \mu, \tag{4–15}$$

where $\hat{x}(t) \in \mathbb{R}^n$ is the state of the MLDNN, $\hat{W}_f \in \mathbb{R}^{N_f+1\times n}$, $\hat{V}_f \in \mathbb{R}^{n\times N_f}$ are the estimates of the ideal weights of the NNs, and $\mu(t) \in \mathbb{R}^n$ denotes the RISE feedback term defined as

$$\mu \triangleq k\left(\tilde{x}(t) - \tilde{x}(0)\right) + \nu, \tag{4–16}$$

where measurable identification error $\tilde{x}(t) \in \mathbb{R}^n$ is defined as

$$\tilde{x} \triangleq x - \hat{x}, \tag{4–17}$$

and $\nu(t) \in \mathbb{R}^n$ is the generalized solution to

$$\dot{\nu} = (k\alpha + \gamma_f)\tilde{x} + \beta_1 sgn(\tilde{x}), \qquad \nu(0) = 0,$$

where k, α, γ_f $\beta_1 \in \mathbb{R}$ are positive constant gains, and $sgn(\cdot)$ denotes a vector signum function. The identification error dynamics are developed by taking the time derivative of Eq. 4–17 and substituting for Eqs. 4–14 and 4–15 as

$$\dot{\tilde{x}} = \tilde{F}_u(x, \hat{x}, u_1, u_2) = W_f^T \sigma_f - \hat{W}_f^T \hat{\sigma}_f + \varepsilon_f(x) - \mu, \tag{4–18}$$

where $\tilde{F}_u(x, \hat{x}, u_1, u_2) = F_u(x, u_1, u_2) - \hat{F}_u(x, \hat{x}, u_1, u_2) \in \mathbb{R}^n$. A filtered identification error is defined as

$$r \triangleq \dot{\tilde{x}} + \alpha\tilde{x}. \tag{4–19}$$

Taking the time derivative of Eq. 4–19 and using Eq. 4–18 yields

$$\dot{r} = W_f^T \sigma_f' V_f^T \dot{x} - \dot{\hat{W}}_f^T \hat{\sigma}_f - \hat{W}_f^T \hat{\sigma}_f' \dot{\hat{V}}_f^T \hat{x} - \hat{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} + \dot{\varepsilon}_f(x) - kr - \gamma_f \tilde{x} - \beta_1 sgn(\tilde{x}) + \alpha\dot{\tilde{x}}. \tag{4–20}$$

The weight update laws for the DNN in Eq. 4–15 are developed based on the subsequent stability analysis as

$$\dot{\hat{W}}_f = proj(\Gamma_{wf} \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} \tilde{x}^T), \qquad \dot{\hat{V}}_f = proj(\Gamma_{vf} \dot{\hat{x}} \tilde{x}^T \hat{W}_f^T \hat{\sigma}_f'), \tag{4–21}$$

where $proj(\cdot)$ is a smooth projection operator [127], [128], and $\Gamma_{wf} \in \mathbb{R}^{L_f+1 \times L_f+1}$, $\Gamma_{vf} \in \mathbb{R}^{n \times n}$ are positive constant adaptation gain matrices. Adding and subtracting $\frac{1}{2} W_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} + \frac{1}{2} \hat{W}_f^T \hat{\sigma}_f' V_f^T \dot{\hat{x}}$, and grouping similar terms, the expression in Eq. 4–20 can be rewritten as

$$\dot{r} = \tilde{N} + N_{B1} + \hat{N}_{B2} - kr - \gamma_f \tilde{x} - \beta_1 sgn(\tilde{x}), \tag{4–22}$$

where the auxiliary signals, $\tilde{N}(x, \tilde{x}, r, \hat{W}_f, \hat{V}_f, t)$, $N_{B1}(x, \hat{x}, \hat{W}_f, \hat{V}_f, t)$, and $\hat{N}_{B2}(\hat{x}, \dot{\hat{x}}, \hat{W}_f, \hat{V}_f, t) \in \mathbb{R}^n$ in Eq. 4–22 are defined as

$$\tilde{N} \triangleq \alpha \dot{\tilde{x}} - \dot{\hat{W}}_f^T \hat{\sigma}_f - \hat{W}_f^T \hat{\sigma}_f' \dot{\hat{V}}_f^T \hat{x} + \frac{1}{2} W_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\tilde{x}} + \frac{1}{2} \hat{W}_f^T \hat{\sigma}_f' V_f^T \dot{\tilde{x}}, \tag{4–23}$$

$$N_{B1} \triangleq W_f^T \sigma_f' V_f^T \dot{x} - \frac{1}{2} W_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} - \frac{1}{2} \hat{W}_f^T \hat{\sigma}_f' V_f^T \dot{x} + \dot{\varepsilon}_f(x), \tag{4–24}$$

$$\hat{N}_{B2} \triangleq \frac{1}{2} \tilde{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} + \frac{1}{2} \hat{W}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{\hat{x}}. \tag{4–25}$$

To facilitate the subsequent stability analysis, an auxiliary term $N_{B2}(\hat{x}, \dot{x}, \hat{W}_f, \hat{V}_f, t) \in \mathbb{R}^n$ is defined by replacing $\dot{\hat{x}}(t)$ in $\hat{N}_{B2}(\cdot)$ by $\dot{x}(t)$, and $\tilde{N}_{B2}(\hat{x}, \dot{\tilde{x}}, \hat{W}_f, \hat{V}_f, t) \triangleq \hat{N}_{B2}(\cdot) - N_{B2}(\cdot)$. The terms $N_{B1}(\cdot)$ and $N_{B2}(\cdot)$ are grouped as $N_B \triangleq N_{B1} + N_{B2}$. Using Assumptions 4-2, 4-3, 4-4, and 4-6, Eqs. 4–19 and 4–21, 4–24 and 4–25 the following bounds can be obtained

$$\left\| \tilde{N} \right\| \leq \rho_1(\|z\|) \|z\|, \tag{4–26}$$

$$\|N_{B1}\| \leq \zeta_1, \quad \|N_{B2}\| \leq \zeta_2, \quad \left\| \dot{N}_B \right\| \leq \zeta_3 + \zeta_4 \rho_2(\|z\|) \|z\|, \tag{4–27}$$

$$\left\| \dot{\tilde{x}}^T \tilde{N}_{B2} \right\| \leq \zeta_5 \|\tilde{x}\|^2 + \zeta_6 \|r\|^2, \tag{4–28}$$

where $z \triangleq \left[\tilde{x}^T \; r^T\right]^T \in \mathbb{R}^{2n}$, $\rho_1(\cdot), \rho_2(\cdot) \in \mathbb{R}$ are positive, globally invertible, non-decreasing functions, and $\zeta_i \in \mathbb{R}$, $i = 1, ..., 6$ are computable positive constants. To facilitate the subsequent stability analysis, let $\mathcal{D} \subset \mathbb{R}^{2n+2}$ be a domain containing $y(t) = 0$, where $y(t) \in \mathbb{R}^{2n+2}$ is defined as

$$y \triangleq \left[\tilde{x}^T \quad r^T \quad \sqrt{P} \quad \sqrt{Q_f}\right]^T, \tag{4–29}$$

where the auxiliary function $P(t) \in \mathbb{R}$ is the generalized solution to the differential equation [129]

$$\dot{P} = -L, \qquad P(0) = \beta_1 \sum_{i=1}^{n} |\tilde{x}_i(0)| - \tilde{x}^T(0) N_B(0), \tag{4–30}$$

where the auxiliary function $L(t) \in \mathbb{R}$ is defined as

$$L \triangleq r^T(N_{B1} - \beta_1 sgn(\tilde{x})) + \dot{\tilde{x}}^T N_{B2} - \beta_2 \rho_2(\|z\|)\|z\|\|\tilde{x}\|, \tag{4–31}$$

where $\beta_1, \beta_2 \in \mathbb{R}$ are chosen according to the following sufficient conditions, such that $P(t) \geq 0$

$$\beta_1 > \max(\zeta_1 + \zeta_2,\ \zeta_1 + \frac{\zeta_3}{\alpha}), \qquad \beta_2 > \zeta_4. \tag{4–32}$$

The auxiliary function $Q_f(\tilde{W}_f, \tilde{V}_f) \in \mathbb{R}$ in Eq. 4–29 is defined as

$$Q_f \triangleq \frac{1}{4}\alpha \left[tr(\tilde{W}_f^T \Gamma_{wf}^{-1} \tilde{W}_f) + tr(\tilde{V}_f^T \Gamma_{vf}^{-1} \tilde{V}_f)\right],$$

where $tr(\cdot)$ denotes the trace of a matrix.

Theorem 4.1. *For the system in Eq. 4–1, the identifier developed in Eq. 4–15 along with its weight update laws in Eq. 4–21 ensures asymptotic identification of the state and its derivative, in the sense that*

$$\lim_{t\to\infty} \|\tilde{x}(t)\| = 0 \qquad \textit{and} \qquad \lim_{t\to\infty} \left\|\dot{\tilde{x}}(t)\right\| = 0,$$

provided Assumptions 4-4 through 4-6 hold, and the control gains k and γ_f are chosen sufficiently large based on the initial conditions of the states[1] *, and satisfy the following sufficient conditions*

$$\alpha\gamma_f > \zeta_5, \qquad k > \zeta_6, \tag{4–33}$$

where ζ_5 and ζ_6 are introduced in Eq. 4–28, and β_1, β_2 introduced in Eq. 4–31, are chosen according to the sufficient conditions in Eq. 4–32.

[1] See subsequent stability analysis.

Proof. Let $V_I(y) : \mathcal{D} \to \mathbb{R}$ be a Lipschitz continuous regular positive definite function defined as

$$V_I \triangleq \frac{1}{2} r^T r + \frac{1}{2} \gamma_f \tilde{x}^T \tilde{x} + P + Q_f, \tag{4–34}$$

which satisfies the following inequalities:

$$U_1(y) \leq V_I(y) \leq U_2(y), \tag{4–35}$$

where $U_1(y)$, $U_2(y) \in \mathbb{R}$ are continuous positive definite functions defined as

$$U_1 \triangleq \frac{1}{2} min(1, \gamma_f) \|y\|^2 \qquad U_2 \triangleq max(1, \gamma_f) \|y\|^2 .$$

From Eqs. 4–18, 4–21, 4–22, and 4–30, the differential equations of the closed-loop system are continuous except in the set $\{y | \tilde{x} = 0\}$. Using Filippov's differential inclusion [116, 118], the existence of solutions can be established for $\dot{y} = f(y)$, where $f(y) \in \mathbb{R}^{2n+2}$ denotes the right-hand side of the the closed-loop error signals. Under Filippov's framework, a generalized Lyapunov stability theory can be used ([119–121] for further details) to establish strong stability of the closed-loop system. The generalized time derivative of Eq. 4–34 exists almost everywhere (a.e.), and $\dot{V}_I(y) \in^{a.e.} \dot{\tilde{V}}_I(y)$ where

$$\dot{\tilde{V}}_I = \bigcap_{\xi \in \partial V_I(y)} \xi^T K \left[\dot{r}^T \;\; \dot{\tilde{x}}^T \;\; \frac{1}{2} P^{-\frac{1}{2}} \dot{P} \;\; \frac{1}{2} Q^{-\frac{1}{2}} \dot{Q} \right]^T ,$$

where ∂V_I is the generalized gradient of V_I [119], and $K[\cdot]$ is defined as [120, 121]

$$K[f](y) \triangleq \bigcap_{\delta > 0} \bigcap_{\mu M = 0} \overline{co} f(B(y, \delta) - M),$$

where $\bigcap_{\mu M = 0}$ denotes the intersection of all sets M of Lebesgue measure zero, $\overline{co}$ denotes convex closure, and $B(y, \delta) = \{x \in \mathbb{R}^{2n+2} | \; \|y - x\| < \delta\}$. Since $V_I(y)$ is a Lipschitz

continuous regular function,

$$\dot{\tilde{V}}_I = \nabla V_I^T K \left[\dot{r}^T \;\; \dot{\tilde{x}}^T \;\; \frac{1}{2} P^{-\frac{1}{2}} \dot{P} \;\; \frac{1}{2} Q^{-\frac{1}{2}} \dot{Q} \right]^T$$
$$= \left[r^T \;\; \gamma_f \tilde{x}^T \;\; 2P^{\frac{1}{2}} \;\; 2Q^{\frac{1}{2}} \right] K \left[\dot{r}^T \;\; \dot{\tilde{x}}^T \;\; \frac{1}{2} P^{-\frac{1}{2}} \dot{P} \;\; \frac{1}{2} Q^{-\frac{1}{2}} \dot{Q} \right]^T.$$

Using the calculus for $K[\cdot]$ from [121], and substituting the dynamics from Eqs. 4–22 and 4–30, yields

$$\begin{aligned}
\dot{\tilde{V}}_I \subset\; & r^T(\tilde{N} + N_{B1} + \hat{N}_{B2} - kr - \beta_1 K[sgn(\tilde{x})] - \gamma_f \tilde{x}) + \gamma_f \tilde{x}^T (r - \alpha \tilde{x}) - r^T (N_{B1} - \beta_1 K[sgn(\tilde{x})]) \\
& - \dot{\tilde{x}}^T N_{B2} + \beta_2 \rho_2(\|z\|) \|z\| \|\tilde{x}\| - \frac{1}{2}\alpha \left[tr(\tilde{W}_f^T \Gamma_{wf}^{-1} \dot{\hat{W}}_f) + tr(\tilde{V}_f^T \Gamma_{vf}^{-1} \dot{\hat{V}}_f) \right]. \\
= & -\alpha \gamma_f \tilde{x}^T \tilde{x} - k r^T r + r^T \tilde{N} + \frac{1}{2}\alpha \tilde{x}^T \tilde{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} + \frac{1}{2}\alpha \tilde{x}^T \hat{W}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{\hat{x}} + \dot{\tilde{x}}^T (\hat{N}_{B2} - N_{B2}) \\
& + \beta_2 \rho_2(\|z\|) \|z\| \|\tilde{x}\| - \frac{1}{2}\alpha tr(\tilde{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\hat{x}} \tilde{x}^T) - \frac{1}{2}\alpha tr(\tilde{V}_f^T \dot{\hat{x}} \tilde{x}^T \hat{W}_f^T \hat{\sigma}_f'), \qquad (4\text{–}36)
\end{aligned}$$

where Eq. 4–21 and the fact that $(r^T - r^T)_i SGN(\tilde{x}_i) = 0$ is used (the subscript i denotes the i^{th} element), where $K[sgn(\tilde{x})] = SGN(\tilde{x})$ [121], such that $SGN(\tilde{x}_i) = 1$ if $\tilde{x}_i > 0$, $[-1, 1]$ if $\tilde{x}_i = 0$, and -1 if $\tilde{x}_i < 0$. Canceling common terms, substituting for $k \triangleq k_1 + k_2$ and $\gamma_f \triangleq \gamma_1 + \gamma_2$, using Eqs. 4–26, 4–28, and completing the squares, the expression in Eq. 4–36 can be upper bounded as

$$\dot{\tilde{V}}_I \leq -(\alpha\gamma_1 - \zeta_5) \|\tilde{x}\|^2 - (k_1 - \zeta_6) \|r\|^2 + \frac{\rho_1(\|z\|)^2}{4k_2} \|z\|^2 + \frac{\beta_2^2 \rho_2(\|z\|)^2}{4\alpha\gamma_2} \|z\|^2. \qquad (4\text{–}37)$$

Provided the sufficient conditions in Eq. 4–33 are satisfied, the expression in Eq. 4–37 can be rewritten as

$$\begin{aligned}
\dot{\tilde{V}}_I &\leq -\lambda \|z\|^2 + \frac{\rho(\|z\|)^2}{4\eta} \|z\|^2 \\
&\leq -U(y) \qquad \forall y \in \mathcal{D}, \qquad (4\text{–}38)
\end{aligned}$$

where $\lambda \triangleq min\{\alpha\gamma_1 - \zeta_5,\ k_1 - \zeta_6\}$, $\rho(\|z\|)^2 \triangleq \rho_1(\|z\|)^2 + \rho_2(\|z\|)^2$, $\eta \triangleq min\{k_2,\ \frac{\alpha\gamma_2}{\beta_2^2}\}$, and $U(y) = c\|z\|^2$, for some positive constant c, is a continuous, positive semi-definite function

defined on the domain

$$\mathcal{D} \triangleq \left\{ y(t) \in \mathbb{R}^{2n+2} \mid \|y\| \leq \rho^{-1}\left(2\sqrt{\lambda\eta}\right) \right\}.$$

The size of the domain $\mathcal{D}$ can be increased by increasing the gains k and γ. The result in Eq. 4–38 indicates that $\dot{V}_I(y) \leq -U(y)$ $\forall$ $\dot{V}_I(y) \overset{a.e.}{\in} \dot{\tilde{V}}_I(y)$ $\forall y \in \mathcal{D}$. The inequalities in Eqs. 4–35 and 4–38 can be used to show that $V_I(y) \in \mathcal{L}_\infty$ in $\mathcal{D}$; hence, $\tilde{x}(t), r(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$. Using Eq. 4–19, standard linear analysis can be used to show that $\dot{\tilde{x}}(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$, and since $\dot{x}(t) \in \mathcal{L}_\infty$, $\dot{\hat{x}}(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$. Since $\hat{W}_f(t) \in \mathcal{L}_\infty$ from the use of projection in Eq. 4–21, $\hat{\sigma}_f(t) \in \mathcal{L}_\infty$ from Assumption 4-6, and $u(t) \in \mathcal{L}_\infty$ from Assumption 4-2, $\mu(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$ from Eq. 4–15. Using the above bounds and the fact that $\hat{\sigma}'_f(t), \dot{\varepsilon}_f(t) \in \mathcal{L}_\infty$, it can be shown from Eq. 4–20 that $\dot{r}(t) \in \mathcal{L}_\infty$ in $\mathcal{D}$. Since $\tilde{x}(t)$,$r(t) \in \mathcal{L}_\infty$, the definition of $U(y)$ can be used to show that it is uniformly continuous in $\mathcal{D}$. Let $\mathcal{S} \subset \mathcal{D}$ denote a set defined as

$$\mathcal{S} \triangleq \left\{ y(t) \subset \mathcal{D} \mid U_2(y(t)) < \frac{1}{2}\left(\rho^{-1}\left(2\sqrt{\lambda\eta}\right)\right)^2 \right\}. \tag{4–39}$$

The region of attraction in Eq. 4–39 can be made arbitrarily large to include any initial conditions by increasing the control gain η (i.e. a semi-global type of stability result), and hence

$$c\|z\|^2 \to 0 \qquad \text{as} \qquad t \to \infty \qquad \forall y(0) \in \mathcal{S},$$

and using the definition of $z(t)$ the following result can be shown

$$\|\tilde{x}(t)\|, \left\|\dot{\tilde{x}}(t)\right\|, \|r\| \to 0 \qquad \text{as} \qquad t \to \infty \qquad \forall y(0) \in \mathcal{S}.$$

□

4.4 Actor-Critic Design

Using Assumption 4-1 and Eq. 4–6, the optimal value function and the optimal controls can be represented by NNs as

$$V^*(x) = W^T\phi(x) + \varepsilon(x), \qquad u_1^*(x) = -\frac{1}{2}R^{-1}g_1^T(x)\left(\phi'(x)^TW + \varepsilon'(x)\right), \qquad (4\text{–}40)$$
$$u_2^*(x) = \frac{1}{2\gamma^2}g_2^T(x)\left(\phi'(x)^TW + \varepsilon'(x)\right),$$

where $W \in \mathbb{R}^N$ is the unknown ideal NN weight, N is the number of neurons, $\phi(x) = [\phi_1(x)\ \ \phi_2(x)\ldots\phi_N(x)]^T \in \mathbb{R}^N$ are smooth NN activation functions, such that $\phi_i(0) = 0$ and $\phi_i'(0) = 0 \quad i = 1...N$, where $\phi'(\cdot)$ denotes the first time derivative of the activation functions, and $\varepsilon(\cdot) \in \mathbb{R}$ is the function reconstruction errors.

Assumption 4.7. *The NN activation function $\{\phi_j(x) : j = 1...N\}$ are chosen such that as $N \to \infty$, $\phi(x)$ provides a complete independent basis for $V^*(x)$.*

Using Assumption 4-7 and Weierstrass higher-order approximation theorem, both $V^*(x)$ and ∇V^* can be uniformly approximated by NNs in Eq. 4–40, i.e. as $N \to \infty$, the approximation errors $\varepsilon(x), \varepsilon'(x) \to 0$, respectively. The critic $\hat{V}(x)$ and the actor $\hat{u}(x)$ approximate the optimal value function and the optimal controls in Eq. 4–40, and are given as

$$\hat{V}(x) = \hat{W}_c^T\phi(x), \qquad \hat{u}_1(x) = -\frac{1}{2}R^{-1}g_1^T(x)\phi'(x)^T\hat{W}_{1a} \qquad (4\text{–}41)$$
$$\hat{u}_2(x) = \frac{1}{2\gamma^2}g_2^T(x)\phi'(x)^T\hat{W}_{2a},$$

where $\hat{W}_c(t) \in \mathbb{R}^N$ and $\hat{W}_{1a}(t), \hat{W}_{2a}(t) \in \mathbb{R}^N$ are estimates of the ideal weights of the critic and actor NNs, respectively. The weight estimation errors for the critic and actor are defined as $\tilde{W}_c(t) \triangleq W - \hat{W}_c(t)$ and $\tilde{W}_{ia}(t) \triangleq W - \hat{W}_{ia}(t)$, for $i = 1, 2$ respectively. The actor and critic NN weights are both updated based on the minimization of the Bellman error $\delta_{hjb}(\cdot)$ in Eq. 4–12, which can be rewritten by substituting $\hat{V}$ from Eq. 4–41 as

$$\delta_{hjb} = \hat{W}_c^T\phi'\hat{F}_{\hat{u}} + r(x, \hat{u}_1, \hat{u}_2) = \hat{W}_c^T\omega + r(x, \hat{u}_1, \hat{u}_2), \qquad (4\text{–}42)$$

where $\omega(x, \hat{u}_1, \hat{u}_2, t) \triangleq \phi' \hat{F}_{\hat{u}} \in \mathbb{R}^N$ is the critic NN regressor vector.

Least Squares Update for the Critic. Consider the integral squared Bellman error $E_c(\hat{W}_c(t), t)$

$$E_c = \int_0^t \delta_{hjb}^2(\tau) d\tau. \tag{4–43}$$

The LS update law for the critic $\hat{W}_c(t)$ is generated by minimizing the total prediction error in Eq. 4–43

$$\begin{aligned}
\frac{\partial E_c}{\partial \hat{W}_c} &= 2 \int_0^t \delta_{hjb}(\tau) \frac{\partial \delta_{hjb}(\tau)}{\partial \hat{W}_c(\tau)} \, d\tau = 0 \\
&= \hat{W}_c^T(t) \int_0^t \omega(\tau)\omega(\tau)^T \, d\tau + \int_0^t \omega(\tau)^T r(\tau) \, d\tau = 0 \\
\hat{W}_c(t) &= -\left(\int_0^t \omega(\tau)\omega(\tau)^T \, d\tau \right)^{-1} \int_0^t \omega(\tau) r(\tau) \, d\tau,
\end{aligned}$$

which gives the LS estimate of the critic weights, provided the inverse $\left(\int_0^t \omega(\tau)\omega(\tau)^T \, d\tau \right)^{-1}$ exists. The recursive formulation of the normalized LS algorithm [130] gives the update laws for the critic weight as

$$\dot{\hat{W}}_c = -\eta_c \Gamma_c \frac{\omega}{1 + \nu \omega^T \Gamma_c \omega} \delta_{hjb}, \tag{4–44}$$

where $\nu, \eta_c \in \mathbb{R}$ are constant positive gains and $\Gamma_c(t) \triangleq \left(\int_0^t \omega(\tau)\omega(\tau)^T \, d\tau \right)^{-1} \in \mathbb{R}^{N \times N}$ is a symmetric estimation gain matrix generated by

$$\dot{\Gamma}_c = -\eta_c \Gamma_c \frac{\omega \omega^T}{1 + \nu \omega^T \Gamma_c \omega} \Gamma_c; \qquad \Gamma_c(t_r^+) = \Gamma_c(0) = \varphi_0 I, \tag{4–45}$$

where t_r^+ is the resetting time at which $\lambda_{min}\{\Gamma_c(t)\} \leq \varphi_1$, and $\varphi_0 > \varphi_1 > 0$. The covariance resetting ensures that $\Gamma_c(t)$ is positive-definite for all time and prevents arbitrarily small values in some directions, making adaptation in those directions very slow (also called the covariance wind-up problem) [130]. From Eq. 4–45 it is clear that $\dot{\Gamma}_c \leq 0$,

which means that the covariance matrix $\Gamma_c(t)$ can be bounded as follows

$$\varphi_1 I \leq \Gamma_c \leq \varphi_0 I \tag{4–46}$$

Gradient Update for the Actor. The actor update, like the critic update in Section 4.4, is based on the minimization of the Bellman error $\delta_{hjb}(\cdot)$. However, unlike the critic weights, the actor weights appear nonlinearly in $\delta_{hjb}(\cdot)$, making it problematic to develop a LS update law. Hence, a gradient update law is developed for the actor which minimizes the squared Bellman error $E_a(t) \triangleq \delta_{hjb}^2$, whose gradients are given as

$$\frac{\partial E_a}{\partial \hat{W}_{1a}} = (\hat{W}_{1a} - \hat{W}_c)^T \phi' G_1 (\phi')^T \delta_{hjb}, \tag{4–47}$$

$$\frac{\partial E_a}{\partial \hat{W}_{2a}} = -(\hat{W}_{2a} + \hat{W}_c)^T \phi' G_2 (\phi')^T \delta_{hjb}, \tag{4–48}$$

where $G_1 \triangleq g_1 R^{-1} g_1^T \in \mathbb{R}^{n\times n}$ and $G_2 \triangleq \gamma^{-2} g_2 g_2^T \in \mathbb{R}^{n\times n}$ are symmetric matrices. Using Eq. 4–47, the actors NNs are updated as

$$\dot{\hat{W}}_{1a} = proj\left\{\frac{-\Gamma_{11a}}{\sqrt{1+\omega^T\omega}} \phi' G_1 (\phi')^T (\hat{W}_{1a} - \hat{W}_c)\delta_{hjb} - \Gamma_{12a}(\hat{W}_{1a} - \hat{W}_c)\right\}, \tag{4–49}$$

$$\dot{\hat{W}}_{2a} = proj\left\{\frac{-\Gamma_{21a}}{\sqrt{1+\omega^T\omega}} \phi' G_2 (\phi')^T (\hat{W}_{2a} + \hat{W}_c)\delta_{hjb} - \Gamma_{22a}(\hat{W}_{2a} - \hat{W}_c)\right\}, \tag{4–50}$$

where $\Gamma_{i1a}, \Gamma_{i2a} \in \mathbb{R}$ for $i = 1, 2$ are positive adaptation gains, and $proj\{\cdot\}$ is a projection operator used to bound the weight estimates [127], [128]. Using the Assumption 4-3 and the projection algorithm in Eq. 4–49, the actor NN weight estimation error can be bounded as

$$\left\|\tilde{W}_{1a}\right\| \leq \kappa_1, \qquad \left\|\tilde{W}_{2a}\right\| \leq \kappa_2, \tag{4–51}$$

where $\kappa_1, \kappa_2 \in \mathbb{R}$ is some positive constant. The first term in Eq. 4–49 is normalized and the last term is added as feedback for stability (based on the subsequent stability analysis).

4.5 Stability Analysis

The dynamics of the critic weight estimation error $\tilde{W}_c(t)$ can be developed using Eqs. 4–9-4–12, 4–42 and 4–44, as

$$\begin{aligned}\dot{\tilde{W}}_c &= \eta_c\Gamma_c\frac{\omega}{1+\nu\omega^T\Gamma_c\omega}\left[-\tilde{W}_c^T\omega - W^T\phi'\tilde{F}_{\hat{u}} + -\varepsilon' F_{u^*} + \hat{u}_1^T R\hat{u}_1 - u_1^{*^T} R u_1^*\right. \\ &\left. + W^T\phi'\left(g_1(\hat{u}_1-u_1^*)+g_2(\hat{u}_2-u_2^*)\right) - \gamma^2\hat{u}_2^T\hat{u}_2 + \gamma^2 u_2^{*^T}u_2^*\right].\end{aligned} \tag{4–52}$$

Substituting for $(u_1^*(x), u_2^*(x))$ and $(\hat{u}_1(x), \hat{u}_2(x))$ from Eqs. 4–40 and 4–41, respectively, in Eq. 4–52 yields

$$\begin{aligned}\dot{\tilde{W}}_c &= -\eta_c\Gamma_c\psi\psi^T\tilde{W}_c + \eta_c\Gamma_c\frac{\omega}{1+\nu\omega^T\Gamma_c\omega}\left[-W^T\phi'\tilde{F}_{\hat{u}} + \frac{1}{4}\tilde{W}_{1a}^T\phi' G_1\phi'^T\tilde{W}_{1a}\right. \\ &\left. + \frac{1}{4}\tilde{W}_{2a}^T\phi' G_2\phi'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon'\left(G_1+G_2\right)\varepsilon'^T - \varepsilon' F_{u^*}\right],\end{aligned} \tag{4–53}$$

where $\psi(t) \triangleq \frac{\omega(t)}{\sqrt{1+\nu\omega(t)^T\Gamma_c(t)\omega(t)}} \in \mathbb{R}^N$ is the normalized critic regressor vector, bounded as

$$\|\psi\| \leq \frac{1}{\sqrt{\nu\varphi_1}}, \tag{4–54}$$

where φ_1 is introduced in Eq. 4–46. The error systems in Eq. 4–53 can be represented as the following perturbed systems

$$\dot{\tilde{W}}_c = \Omega + \Delta, \tag{4–55}$$

where $\Omega(\tilde{W}_c, t) \triangleq -\eta_c\Gamma_c\psi\psi^T\tilde{W}_c \in \mathbb{R}^N$, denotes the nominal system, and

$$\begin{aligned}\Delta(t) \triangleq \frac{\eta_c\Gamma_c\omega}{1+\nu\omega^T\Gamma_c\omega}&\left[-W^T\phi'\tilde{F}_{\hat{u}} + \frac{1}{4}\tilde{W}_{1a}^T\phi' G_1\phi'^T\tilde{W}_{1a}\right. \\ &\left. + \frac{1}{4}\tilde{W}_{2a}^T\phi' G_2\phi'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon'\left(G_1+G_2\right)\varepsilon'^T - \varepsilon' F_{u^*}\right],\end{aligned}$$

denotes the perturbations. Using Theorem 2.5.1 in [130], it can be shown that the nominal systems

$$\dot{\tilde{W}}_c = -\eta_c\Gamma_c\psi\psi^T\tilde{W}_c, \tag{4–56}$$

are exponentially stable, if the bounded signals ψ is PE, i.e.

$$\mu_2 I \geq \int_{t_0}^{t_0+\delta} \psi(\tau)\psi(\tau)^T d\tau \geq \mu_1 I \qquad \forall t_0 \geq 0,$$

where $\mu_1, \mu_2, \delta \in \mathbb{R}$ are some positive constants. Since $\Omega(\tilde{W}_c, t)$ is continuously differentiable and the Jacobian $\frac{\partial \Omega}{\partial \tilde{W}_c} = -\eta_c \Gamma_c \psi \psi^T$ is bounded for the exponentially stable system Eq. 4–56 the converse Lyapunov Theorem 4.14 in [131] can be used to show that there exists a function $V_c : \mathbb{R}^N \times [0, \infty) \to \mathbb{R}$, which satisfies the following inequalities

$$\begin{aligned} c_1 \left\|\tilde{W}_c\right\|^2 \leq V_c(\tilde{W}_c, t) \leq c_2 \left\|\tilde{W}_c\right\|^2 \\ \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c}\Omega(\tilde{W}_c, t) \leq -c_3 \left\|\tilde{W}_c\right\|^2 \\ \left\|\frac{\partial V_c}{\partial \tilde{W}_c}\right\| \leq c_4 \left\|\tilde{W}_c\right\|, \end{aligned} \tag{4–57}$$

for some positive constants $c_1,\ c_2,\ c_3,\ c_4 \in \mathbb{R}$ for $i = 1, 2$. Using Assumptions 4-1 through 4-3, and 4-5 through 4-7, the projection bounds in Eq. 4–49, the fact that $F_{u^*} \in \mathcal{L}_\infty$ (since $(u_1^*(x), u_2^*(x))$ is stabilizing), and provided the conditions of Theorem 4-1 hold (required to prove that $\tilde{F}_{\hat{u}} \in \mathcal{L}_\infty$), the following bounds are developed to facilitate the subsequent stability proof

$$\begin{aligned} \left\|\tilde{W}_{1a}\right\| \leq \kappa_1 \qquad \left\|\tilde{W}_{2a}\right\| \leq \kappa_2, \\ \left\|\tilde{W}_{1a}^T \phi' G_1 \phi'^T \tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T \phi' G_2 \phi'^T \tilde{W}_{2a}\right\| \leq \kappa_{3a}, \\ \left\|-W^T \phi' \tilde{F}_{\hat{u}} - \frac{1}{4}\varepsilon' \left(G_1 - G_2\right) \varepsilon'^T - \varepsilon' F_{u^*}\right\| \leq \kappa_{3b}, \\ \left\|\frac{1}{2}\left(W^T \phi' + \varepsilon'\right)\left(\left(G_1 + G_2\right)\varepsilon'^T + G_1 \phi'^T \tilde{W}_{1a} + G_2 \phi'^T \tilde{W}_{2a}\right)\right\| \leq \kappa_4, \\ \left\|\phi' G_1 \phi'^T\right\| \leq \kappa_5 \qquad \left\|\phi' G_2 \phi'^T\right\| \leq \kappa_6, \end{aligned} \tag{4–58}$$

where $\kappa_3 \triangleq \kappa_{3a} + \kappa_{3b}$ and $\kappa_j \in \mathbb{R}$ for $j = 1, \ldots, 6$ are computable positive constants.

Theorem 4.2. *If Assumptions 4-1 through 4-7 hold, the regressors* $\psi(t) \triangleq \frac{\omega}{\sqrt{1+\omega^T \Gamma_c \omega}}$ *is PE, and provided Eq. 4–32, Eq. 4–33 and the following sufficient gain conditions are*

satisfied

$$c_3 > \Gamma_{11a}\kappa_1\kappa_5 + \Gamma_{21a}\kappa_2\kappa_6, \qquad \Gamma_{22a} > \frac{\kappa_6}{4},$$

where Γ_{11a}, Γ_{21a}, Γ_{22a}, c_3, κ_1, κ_2, κ_5 *and* κ_6 *are introduced in Eqs. 4–49, 4–57, and 4–58, then the controller in Eq. 4–41, the actor-critic weight update laws in Eqs. 4–44-4–45 and 4–49, and the identifier in Eqs. 4–15 and 4–21, guarantee that the state of the system* $x(t)$, *and the actor-critic weight estimation errors* $\tilde{W}_{1a}(t)$, $\tilde{W}_{2a}(t)$ *and* $\tilde{W}_c(t)$ *are UUB.*

Proof. To investigate the stability of the the system Eq. 4–1 with control $(\hat{u}_1, \hat{u}_2)$, and the perturbed system Eq. 4–55, consider $V_L : \mathcal{X} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times [0, \infty) \to \mathbb{R}$ as the continuously differentiable, positive-definite Lyapunov function candidate, given as

$$V_L(x, \tilde{W}_c, \tilde{W}_{1a}, \tilde{W}_{2a}, t) \triangleq V^*(x) + V_c(\tilde{W}_c, t) + \frac{1}{2}\tilde{W}_{1a}^T\tilde{W}_{1a} + \frac{1}{2}\tilde{W}_{2a}^T\tilde{W}_{2a},$$

where $V^*(x)$ (the optimal value function for Eq. 4–1), is the Lyapunov function for Eq. 4–1, and $V_c(\tilde{W}_c, t)$ is the Lyapunov function for the exponentially stable system in Eq. 4–56. Since $V^*(x)$ are continuously differentiable and positive-definite from Eq. 4–3, from Lemma 4.3 in [131], there exist class $\mathcal{K}$ functions α_1 and α_2 defined on $[0,\ r]$, where $B_r \subset \mathcal{X}$, such that

$$\alpha_1(\|x\|) \leq V^*(x) \leq \alpha_2(\|x\|) \qquad \forall x \in B_r. \tag{4–59}$$

Using Eqs. 4–57 and 4–59, $V_L(x, \tilde{W}_{1a}, \tilde{W}_{2a}, t)$ can be bounded as

$$\begin{aligned} \alpha_1(\|x\|) + c_1\left\|\tilde{W}_c\right\|^2 + \frac{1}{2}\left(\left\|\tilde{W}_{1a}\right\|^2 + \left\|\tilde{W}_{2a}\right\|^2\right) &\leq V_L, \\ \alpha_2(\|x\|) + c_2\left\|\tilde{W}_c\right\|^2 + \frac{1}{2}\left(\left\|\tilde{W}_{1a}\right\|^2 + \left\|\tilde{W}_{2a}\right\|^2\right) &\geq V_L \end{aligned}$$

which can be written as

$$\alpha_3(\|w\|) \leq V_L(x, \tilde{W}_{1a}, \tilde{W}_{2a}, t) \leq \alpha_4(\|w\|) \qquad \forall w \in B_s,$$

where $w(t) \triangleq [x(t)^T\ \tilde{W}_c(t)^T\ \tilde{W}_{1a}(t)^T\ \tilde{W}_{2a}(t)^T]^T \in \mathbb{R}^{n+3N}$, α_3 and α_4 are class $\mathcal{K}$ functions defined on $[0,\ s]$, where $B_s \subset \mathcal{X} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N$. Taking the time derivative of $V_L(\cdot)$

yields

$$\dot{V}_L = \nabla V^* \left(f + g_1\hat{u}_1 + g_2\hat{u}_2\right) + \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c}\Omega + \frac{\partial V_c}{\partial \tilde{W}_c}\Delta \qquad (4\text{–}60)$$
$$-\tilde{W}_{1a}^T\dot{\hat{W}}_{1a} - \tilde{W}_{2a}^T\dot{\hat{W}}_{2a},$$

where the time derivative of $V^*(\cdot)$ is taken along the the trajectories of the system Eq. 5–16 with control inputs $(\hat{u}_1(\cdot), \hat{u}_2(\cdot))$ and the time derivative of $V_c(\cdot)$ is taken along the along the trajectories of the perturbed system Eq. 4–55. Using the HJI equation Eq. 4–10, $\nabla V^* f = -\nabla V^* \left(g_1 u_1^* + g_2 u_2^*\right) - Q(x) - u_1^{*^T} R u_1^* + \gamma^2 u_2^{*^T} u_2^*$. Substituting for the $\nabla V^* f$ terms in Eq. 4–60, using the fact that $\nabla V^* g_1 = -2u_1^{*^T} R$ and $\nabla V^* g_2 = 2\gamma^2 u_2^{*^T}$ from Eqs. 4–6 and 4–7, and using Eqs. 4–49 and 4–57, 4–60 can be upper bounded as

$$\dot{V}_L \leq -Q - u_1^{*^T} R u_1^* + \gamma u_2^{*^T} u_2^* - c_3 \left\|\tilde{W}_c\right\|^2 \qquad (4\text{–}61)$$
$$+ c_4 \left\|\tilde{W}_c\right\| \|\Delta\| + 2u_1^{*^T} R(u_1^* - \hat{u}_1) - 2\gamma^2 u_2^{*^T}(u_2^* - \hat{u}_2)$$
$$+ \tilde{W}_{1a}^T \left[\frac{\Gamma_{11a}}{\sqrt{1+\omega^T\omega}}\phi' G_1 \left(\phi'\right)^T (\hat{W}_{1a} - \hat{W}_c)\delta_{hjb} + \Gamma_{12a}(\hat{W}_{1a} - \hat{W}_c)\right]$$
$$+ \tilde{W}_{2a}^T \left[\frac{\Gamma_{21a}}{\sqrt{1+\omega^T\omega}}\phi' G_2 \left(\phi'\right)^T (\hat{W}_{2a} + \hat{W}_c)\delta_{hjb} + \Gamma_{22a}(\hat{W}_{2a} - \hat{W}_c)\right].$$

Substituting for u^*, $\hat{u}$, δ_{hjb}, and Δ using Eqs. 4–6, 4–7, 4–41, 4–52, and 4–55, respectively, and using Eqs. 4–46 and 4–54 in Eq. 4–61, yields

$$\begin{aligned}
\dot{V}_L \leq & -Q - c_3 \left\|\tilde{W}_c\right\|^2 - \left(\Gamma_{22a} - \frac{1}{4}\left\|\phi' G_2 \phi'^T\right\|\right)\left\|\tilde{W}_{2a}\right\|^2 - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 \\
& + \frac{1}{2}\left(W^T\phi' + \varepsilon'\right)\left(\left(G_1 + G_2\right)\varepsilon'^T + G_1\phi'^T\tilde{W}_{1a} + G_2\phi'^T\tilde{W}_{2a}\right) \\
& + c_4\frac{\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}}\left\| -W^T\phi'\tilde{F}_{\hat{u}} + \frac{1}{4}\tilde{W}_{1a}^T\phi' G_1\phi'^T\tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T\phi' G_2\phi'^T\tilde{W}_{2a}\right. \\
& \left. -\frac{1}{4}\varepsilon'\left(G_1 - G_2\right)\varepsilon'^T - \varepsilon' F_{u^*}\right\|\left\|\tilde{W}_c\right\| \\
& + \frac{\Gamma_{11a}}{\sqrt{1+\omega^T\omega}}\tilde{W}_{1a}^T\phi' G_1\phi'^T\left(\tilde{W}_{1c} - \tilde{W}_{1a}\right)\left(-\tilde{W}_c^T\omega - W^T\phi'\tilde{F}_{\hat{u}}\right. \\
& \left. + \frac{1}{4}\tilde{W}_{1a}^T\phi' G_1\phi'^T\tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T\phi' G_2\phi'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon'\left(G_1 - G_2\right)\varepsilon'^T - \varepsilon' F_{u^*}\right) \\
& + \frac{\Gamma_{21a}}{\sqrt{1+\omega^T\omega}}\tilde{W}_{2a}^T\phi' G_2\phi'^T\left(-\tilde{W}_c - \tilde{W}_{2a} + 2W\right)\left(-\tilde{W}_c^T\omega - W^T\phi'\tilde{F}_{\hat{u}}\right. \\
& \left. + \frac{1}{4}\tilde{W}_{1a}^T\phi' G_1\phi'^T\tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T\phi' G_2\phi'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon'\left(G_1 - G_2\right)\varepsilon'^T - \varepsilon' F_{u^*}\right) \\
& + \left(\Gamma_{12a}\left\|\tilde{W}_{1a}\right\| + \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|\right)\left\|\tilde{W}_c\right\|.
\end{aligned} \tag{4–62}$$

Using the bounds developed in Eq. 4–58 and Assumption 4-3, Eq. 4–62 can be further upper bounded as

$$\begin{aligned}
\dot{V}_L \leq & -Q - \left(c_3 - \Gamma_{11a}\kappa_1\kappa_5 - \Gamma_{21a}\kappa_2\kappa_6\right)\left\|\tilde{W}_c\right\|^2 - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 \\
& - \left(\Gamma_{22a} - \frac{1}{4}\kappa_6\right)\left\|\tilde{W}_{2a}\right\|^2 + \Psi\left\|\tilde{W}_c\right\| \\
& + \Gamma_{11a}\kappa_1^2\kappa_5\kappa_3 + \Gamma_{21a}\left(\kappa_1^2 + 2\bar{W}\right)\kappa_6\kappa_3 + \kappa_4,
\end{aligned}$$

where

$$\begin{aligned}
\Psi \triangleq & \frac{c_4\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}}\kappa_3 + \left(\Gamma_{11a}\kappa_1\kappa_5 + \Gamma_{21a}\kappa_2\kappa_6\right)\kappa_3 + \Gamma_{11a}\kappa_1^2\kappa_5 \\
& + \Gamma_{21a}\left(\kappa_1 + 2\bar{W}\right)\kappa_1\kappa_5 + \Gamma_{12a}\kappa_1 + \Gamma_{22a}\kappa_2.
\end{aligned}$$

Provided $c_3 > \Gamma_{11a}\kappa_1\kappa_5 + \Gamma_{21a}\kappa_2\kappa_6$ and $4\Gamma_{22a} > \kappa_6$, and completing the square yields

$$\begin{aligned}\dot{V}_L \leq & -Q - (1-\theta_1)(c_3 - \Gamma_{11a}\kappa_1\kappa_5 - \Gamma_{21a}\kappa_2\kappa_6)\left\|\tilde{W}_c\right\|^2 \\ & - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 - \left(\Gamma_{22a} - \frac{1}{4}\kappa_6\right)\left\|\tilde{W}_{2a}\right\|^2 \\ & + \frac{\Psi^2}{4\theta_1(c_3 - \Gamma_{11a}\kappa_1\kappa_5 - \Gamma_{21a}\kappa_2\kappa_6)} \\ & + \Gamma_{11a}\kappa_1^2\kappa_5\kappa_3 + \Gamma_{21a}\left(\kappa_1^2 + 2\bar{W}\right)\kappa_6\kappa_3 + \kappa_4,\end{aligned} \tag{4–63}$$

where $\theta_1 \in (0,1)$. Since $Q(x)$ is positive definite, according to Lemma 4.3 in [131], there exist class $\mathcal{K}$ functions α_5 and α_6 such that

$$\alpha_5(\|w\|) \leq F(\|w\|) \leq \alpha_6(\|w\|) \qquad \forall w \in B_s, \tag{4–64}$$

where

$$\begin{aligned}F(\|w\|) = & \; Q + (1-\theta_1)(c_3 - \Gamma_{11a}\kappa_1\kappa_5 - \Gamma_{21a}\kappa_2\kappa_6)\left\|\tilde{W}_c\right\|^2 \\ & + \left(\Gamma_{22a} - \frac{1}{4}\kappa_6\right)\left\|\tilde{W}_{2a}\right\|^2 + \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2.\end{aligned}$$

Using Eq. 4–64, the expression in Eq. 4–63 can be further upper bounded as

$$\dot{V}_L \leq -\alpha_5(\|w\|) + \Lambda$$

where

$$\Lambda = \frac{\Psi^2}{4\theta_1(c_3 - \Gamma_{11a}\kappa_1\kappa_5 - \Gamma_{21a}\kappa_2\kappa_6)} + \Gamma_{11a}\kappa_1^2\kappa_5\kappa_3 + \Gamma_{21a}\left(\kappa_1^2 + 2\bar{W}\right)\kappa_6\kappa_3 + \kappa_4,$$

which proves that $\dot{V}_L(\cdot)$ is negative whenever $w(t)$ lies outside the compact set $\Omega_w \triangleq \left\{w : \|w\| \leq \alpha_5^{-1}(\Lambda)\right\}$, and hence, $\|w(t)\|$ is UUB, according to Theorem 4.18 in [131]. □

Remark 4.1. Since the actor, critic and identifier are continuously updated, the developed RL algorithm can be compared to fully optimistic policy iteration (PI) in machine learning literature [87], where policy evaluation and policy improvement are done after

every state transition. This differs from traditional PI, where policy improvement is done only after the convergence of each policy evaluation step. Convergence behavior of optimistic PI is not fully understood, and by considering an adaptive control framework, this result investigates the convergence and stability behavior of fully optimistic PI in continuous-time. The requirement of PE condition in Theorem 4-2 is equivalent to the exploration paradigm in RL which ensures sufficient sampling of the state space and convergence to the optimal policy [2]. The theorem shows that the PE condition is needed for proper identification of the value function. The theorem makes no mention of finding the minimum non-negative definite solution to the HJI. However it does guarantee convergence to a solution (u_1, u_2) such that the dynamics in Eq. 4–1 are stable. This is only accomplished by the minimal non-negative definite HJI solution.

4.6 Convergence to Nash Solution

In addition to establishing convergence of the actor and critic weights, it is prudent to also consider the convergence of the control strategies to the saddle point Nash equilibrium. The subsequent analysis demonstrates that the actor and critic NN approximations converge to the approximate HJI equation in Eq. 3–15. It can also be shown that the approximate controllers in Eq. 4–41 approximate the optimal solutions to the two player Nash game for the dynamic system given in Eq. 4–1. To facilitate the subsequent analysis the following assumption is made.

Assumption 4.8. *For each set of admissible control policies, the HJI equation Eq. 4–8 has a locally smooth solution $V(x) \geq 0, \forall x \in \Omega_x$, where Ω_x is the set described in Section 4.1. On $\Omega_x \subseteq \mathbb{R}^n$, $f(\cdot)$ is Lipschitz and bounded by $\|f\| \leq c_f \|x\|$, where $c_f \in \mathbb{R}$ is a positive constant.*

Theorem 4.3. *Given that the Assumptions and sufficient gain constraints in Theorem 4-2 hold, then the actor and critic NNs converge to the approximate HJI solution, in the sense that the HJIs in Eq. 4–11 are UUB.*

Proof. Consider the approximate HJI in Eq. 5–13 and after substituting the approximate control laws in Eq. 5–19 yields

$$\begin{aligned} H\left(x, \nabla\hat{V}, \hat{u}_1, \hat{u}_2\right) &= r_{\hat{u}} + \nabla\hat{V} F_{\hat{u}} \\ &= Q(x) + \frac{1}{4}\hat{W}_{1a}^T \phi' G_1 \phi'^T \hat{W}_{1a} - \frac{1}{4}\hat{W}_{2a}^T \phi' G_2 \phi'^T \hat{W}_{2a} \\ &\quad + \hat{W}_c^T \phi' f(x) - \frac{1}{2}\hat{W}_c^T \phi' \left(G_1 \phi'^T \hat{W}_{1a} - G_2 \phi'^T \hat{W}_{2a}\right). \end{aligned}$$

After adding and subtracting $\left(W^T\phi' + \varepsilon'\right) f = -\left(W^T\phi' + \varepsilon'\right)\left(g_1 u_1^* + g_2 u_2^*\right) - Q(x) - u_1^{*T} R u_1^* + \gamma^2 u_2^{*T} u_2^*$ and substituting for the optimal control law in Eq. 5–18 as

$$\begin{aligned} H &= -\tilde{W}_c^T \phi_1' f - \varepsilon' f + \frac{1}{4}\hat{W}_{1a}^T \phi' G_1 \phi'^T \hat{W}_{1a} - \frac{1}{4} W^T \phi' G_1 \phi'^T W \qquad (4\text{–}65) \\ &\quad - \frac{1}{4}\hat{W}_{2a}^T \phi' G_2 \phi'^T \hat{W}_{2a} + \frac{1}{4} W^T \phi' G_2 \phi'^T W \\ &\quad - \frac{1}{2}\hat{W}_c^T \phi' \left(G_1 \phi'^T \hat{W}_{1a} - G_2 \phi'^T \hat{W}_{2a}\right) \\ &\quad + \frac{1}{2}\left(W^T\phi' + \varepsilon'\right)\left(G_1 - G_2\right)\phi'^T W + \frac{1}{4}\varepsilon'\left(G_1 - G_2\right)\varepsilon'^T. \qquad (4\text{–}66) \end{aligned}$$

Substituting the NN mismatch errors $\tilde{W}_c(t) \triangleq W - \hat{W}_c(t)$ and $\tilde{W}_{ia}(t) \triangleq W - \hat{W}_{ia}(t)$, for $i = 1, 2$ respectively into 4–65,

$$\begin{aligned} H &= -\tilde{W}_c^T \phi' f(x) - \varepsilon' f + \frac{1}{4}\tilde{W}_{1a}^T \phi' G_1 \phi'^T \tilde{W}_{1a} + \frac{1}{4}\varepsilon'\left(G_1 - G_2\right)\varepsilon'^T \qquad (4\text{–}67) \\ &\quad - \frac{1}{4}\tilde{W}_{2a}^T \phi' G_2 \phi'^T \tilde{W}_{2a} + \frac{1}{2}\tilde{W}_c^T \phi'\left(G_1 - G_2\right)\phi'^T W \\ &\quad - \frac{1}{2}\tilde{W}_c^T \phi' \left(G_1 \phi'^T \tilde{W}_{1a} - G_2 \phi'^T \tilde{W}_{2a}\right) + \frac{1}{2}\varepsilon'\left(G_1 - G_2\right)\phi'^T W. \end{aligned}$$

Using the bounds developed in Eq. 4–58, Assumption 4-2 through 4-4 and Assumption 4-8, Eq. 4–67 can be upper bounded as

$$\begin{aligned} \|H\| &\leq \left(c_f \bar{\phi}' \|x\| + \kappa_5\left(\bar{W} + \kappa_1\right) + \kappa_6\left(\bar{W} + \kappa_2\right)\right)\left\|\tilde{W}_c\right\| + \bar{\varepsilon}' c_f \|x\| \qquad (4\text{–}68) \\ &\quad + \kappa_5 \left\|\tilde{W}_{1a}\right\|^2 - \kappa_6 \left\|\tilde{W}_{2a}\right\|^2 + \bar{\varepsilon}' \left\|\left(G_1 - G_2\right)\right\|\left(\bar{\phi}'\bar{W} + \bar{\varepsilon}'\right). \end{aligned}$$

Using the Assumptions and Theorem 4-2, it is easy to see that all terms to the right of the inequality are UUB, therefore the approximate HJI is also UUB. □

Theorem 4.4. *Given that the Assumptions and sufficient gain constraints in Theorem 4-2 hold, the approximate control laws in Eq. 4–41 converge to the approximate Nash equilibrium solution of the zero-sum game.*

Proof. Consider the control errors $(\tilde{u}_1, \tilde{u}_2)$ between the optimal control laws in Eqs. 4–6 and 4–7, and the approximate control laws in Eq. 4–41 given as

$$\tilde{u}_1 \triangleq u_1^* - \hat{u}_1, \qquad \tilde{u}_2 \triangleq u_2^* - \hat{u}_2.$$

Substituting for the optimal control laws in Eqs. 4–6 and 4–7, and the approximate control laws in Eq. 4–41 and using $\tilde{W}_{ia}(t) \triangleq W_i - \hat{W}_{ia}(t)$ for $i = 1, 2$, yields

$$\begin{aligned} \tilde{u}_1 &= -\frac{1}{2} R_{11}^{-1} g_1^T \phi' \left(\tilde{W}_{1a} + \varepsilon' \right) \\ \tilde{u}_2 &= \frac{1}{2\gamma^2} g_2^T \phi' \left(\tilde{W}_{2a} + \varepsilon' \right). \end{aligned} \tag{4–69}$$

Using Assumptions 4-1 through 4-4, Eq. 4–69 can be upper bounded as

$$\begin{aligned} \|\tilde{u}_1\| &\leq \frac{1}{2} \lambda_{min} \left(R_{11}^{-1} \right) \bar{g}_1 \bar{\phi}' \left(\left\| \tilde{W}_{1a} \right\| + \bar{\varepsilon}'_1 \right) \\ \|\tilde{u}_2\| &\leq \frac{1}{2\gamma^2} \bar{g}_2 \bar{\phi}' \left(\left\| \tilde{W}_{2a} \right\| + \bar{\varepsilon}'_2 \right). \end{aligned}$$

Given that the Assumptions and sufficient gain constraints in Theorem 4-2 hold, then all terms to the right of the inequality are UUB, therefore the control errors $(\tilde{u}_1, \tilde{u}_2)$ are UUB and the approximate control laws $(\hat{u}_1, \hat{u}_2)$ give the approximate Nash equilibrium solution. □

4.7 Simulation

The following nonlinear dynamics are considered in [107, 108, 132]

$$\dot{x} = f(x) + g_1(x) u_1(x) + g_2(x) u_2(x),$$

where

$$\begin{aligned} f(x) &= \begin{bmatrix} -x_1 + x_2 \\ -x_1^3 - x_2^3 + \frac{1}{4}x_2 \left(cos\left(2x_2\right) + 2\right)^2 - \frac{1}{4\gamma^2}x_2 \left(sin\left(4x_1\right) + 2\right)^2 \end{bmatrix} \\ g_1(x) &= \begin{bmatrix} 0 & cos\left(2x_2\right) + 2 \end{bmatrix}^T \qquad g_2(x) = \begin{bmatrix} 0 & sin\left(4x_1\right) + 2 \end{bmatrix}^T. \end{aligned}$$

The initial state is given as $x(0) = [3, -1]^T$ and the local cost function is defined as

$$r = x^T Q x + u_1^T R u_1 - \gamma^2 u_2^T u_2$$

where

$$R = 1, \qquad \gamma^2 = 8, \qquad Q = \mathbb{I}_{2\times 2}.$$

The optimal value function is

$$V^*(x) = \frac{1}{4}x_1^4 + \frac{1}{2}x_2^2,$$

and the optimal control inputs are given as

$$u_1^* = -\left(cos\left(2x_1\right) + 2\right)x_2, \qquad u_2^* = \frac{1}{\gamma^2}\left(sin\left(4x_1\right) + 2\right)x_2.$$

The activation function for the critic NN is chosen as

$$\phi = \begin{bmatrix} x_1^2 & x_2^2 & x_1^4 & x_2^4 \end{bmatrix},$$

while the activation function for the identifier DNN is chosen as a symmetric sigmoid with 5 neurons in the hidden layer. The identifier gains are chosen as

$$k = 100, \qquad \alpha = 30, \qquad \gamma_f = 5, \qquad \beta_1 = 0.2, \qquad \Gamma_{wf} = 0.2\mathbb{I}_{6\times 6}, \qquad \Gamma_{vf} = 0.2\mathbb{I}_{2\times 2},$$

and the gains of the actor-critic learning laws are chosen as

$$\Gamma_{11a} = \Gamma_{21a} = 1, \qquad \Gamma_{12a} = \Gamma_{22a} = 0.5, \qquad \eta_c = 1, \qquad \nu = 0.005.$$

The covariance matrix is initialized to $\Gamma(0) = .001$, all the NN weights are randomly initialized with values between $[-1, 1]$, and the states are initialized to $x(0) = [3, -1]$. A small amplitude exploratory signal (noise) is added to the control to excite the states for the first 10 seconds of the simulation, as seen from the evolution of states in Figure 4-1. The identifier approximates the system dynamics, and the state derivative estimation error is shown in Figure 4-2. The time histories of the critic NN weights and the actors NN weights are given in Figure 4-3 and 4-4. Persistence of excitation ensures that the weights converge. Figure 4-5 shows the difference between the optimal value function and the approximate one. Figure 4-6 demonstrates the approximation error between the optimal controller and the approximated controller for player 1 and 2, respectively. Figures 4-7, 4-8, 4-9, and 4-10 demonstrate that for a PE signal that is not removed the weights converge, however the PE signal degrades the performance of the states.

Remark 4.2. An implementation issue in using the developed algorithm is to ensure PE of the critic regressor vector. Unlike linear systems, where PE of the regressor translates to the sufficient richness of the external input, no verifiable method exists to ensure PE in nonlinear systems. In this simulation, a small exploratory signal consisting of sinusoids of varying frequencies was added to the control to ensure PE qualitatively, and convergence of critic weights to their optimal values is achieved. The exploratory signal $n(t)$ is present in the first 3 seconds of the simulation and is given by

$$n(t) = (1.2 - \exp(.01t)) \left(cos^2(0.2t) + sin^2(2.0t)\, cos(0.1t) + sin^2(-1.2t)\, cos(.5t) + sin^5(t) \right).$$

4.8 Summary

A generalized solution for a two player zero-sum differential game is sought utilizing the ACI architecture for nonlinear a HJI equation. The ACI architecture implements the actor and critic approximation simultaneously and in real-time. The use of a robust DNN-based identifier circumvents the need for complete model knowledge, yielding an identifier which is proven to be asymptotically convergent. Least squares approximation

and adaptive control theory techniques are utilized to update the weights for the critic and actor NNs to approximate the value function and approximate control policies. Using the identifier and the critic, an approximation to the optimal control law (actor) is developed which stabilizes the closed loop system and approaches the optimal solutions to the two player zero-sum game.

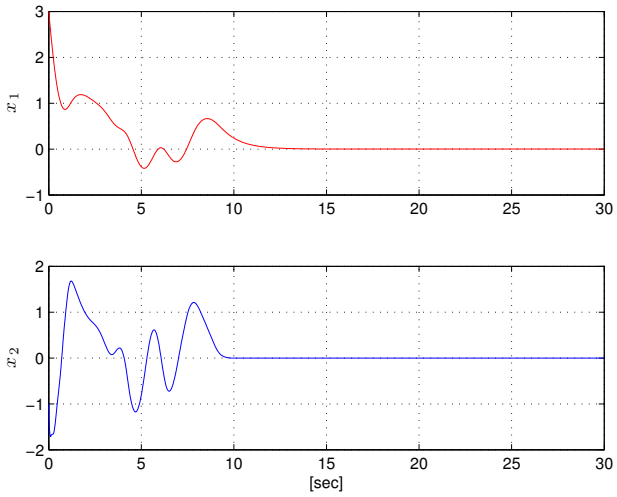

Figure 4-1. The evolution of the system states for the zero-sum game, with persistently excited input for the first 10 seconds.

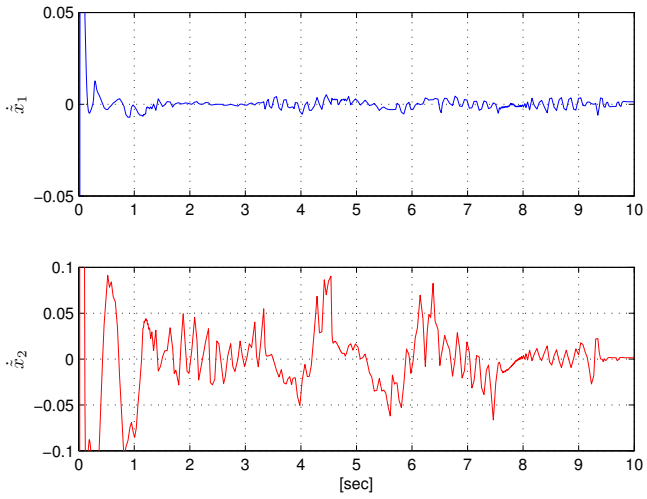

Figure 4-2. Error in estimating the state derivatives, with the identifier for the zero-sum game.

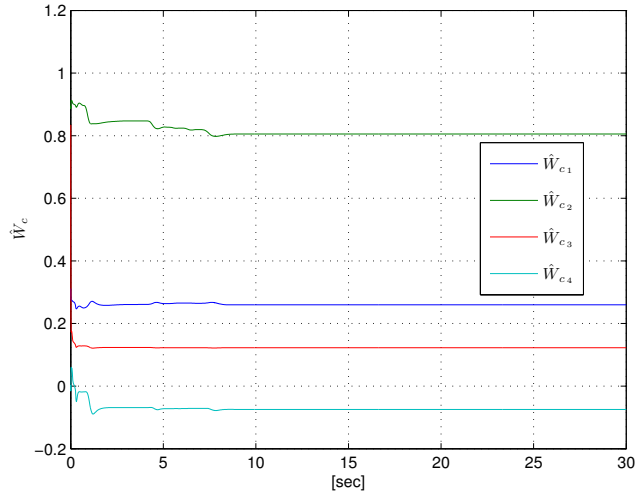

Figure 4-3. Convergence of critic weights for the zero-sum game.

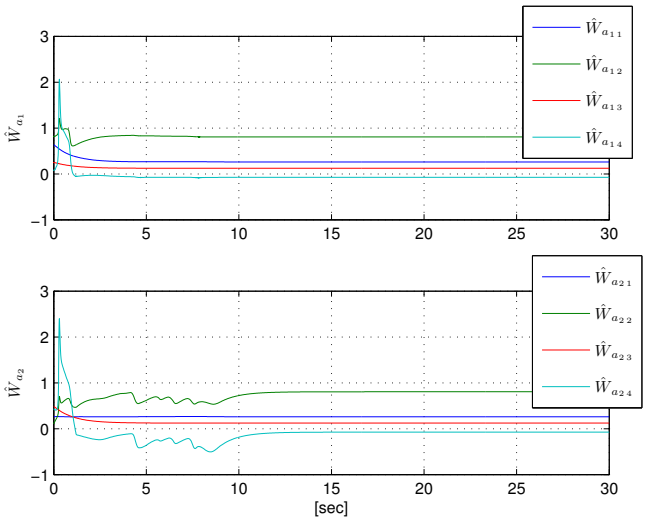

Figure 4-4. Convergence of actor weights for player 1 and player 2 in a zero-sum game.

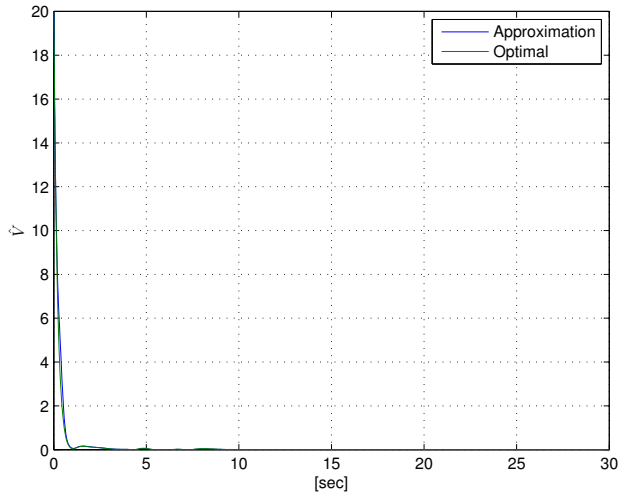

Figure 4-5. Optimal value function approximation $\hat{V}(x)$, for a zero-sum game.

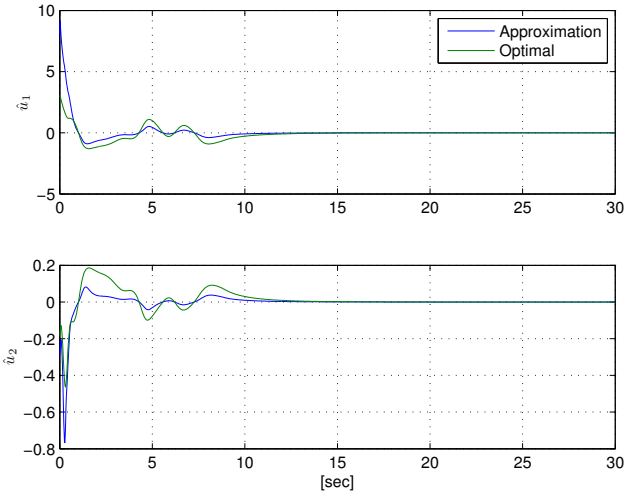

Figure 4-6. Optimal control approximations $\hat{u}_1$ and $\hat{u}_2$, in a zero-sum game.

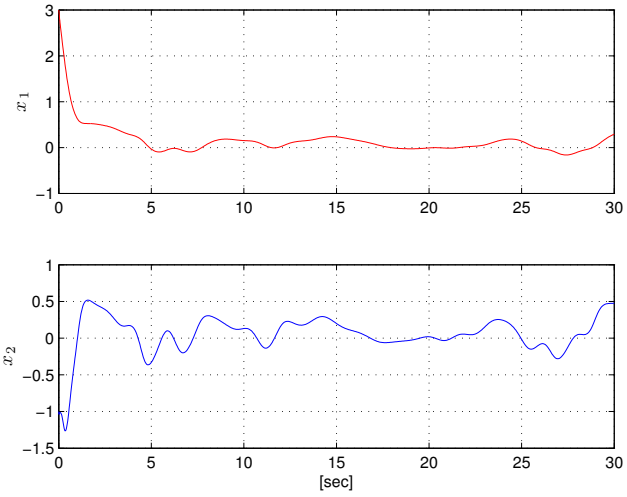

Figure 4-7. The evolution of the system states for the zero-sum game, with a continuous persistently excited input.

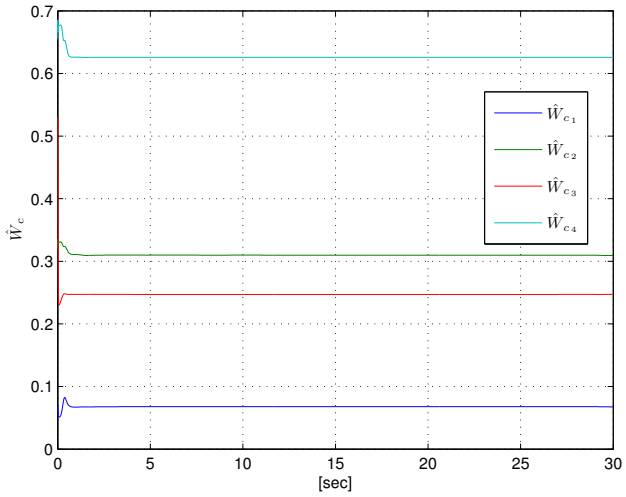

Figure 4-8. Convergence of critic weights for the zero-sum game, with a continuous persistently excited input.

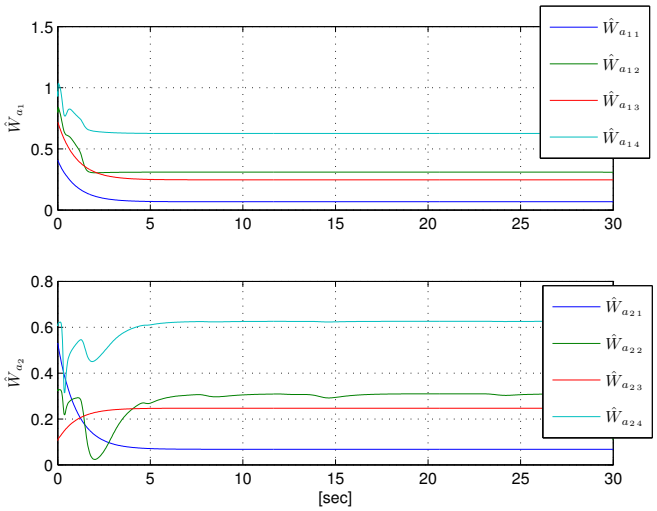

Figure 4-9. Convergence of actor weights for player 1 and player 2 in a zero-sum game, with a continuous persistently excited input.

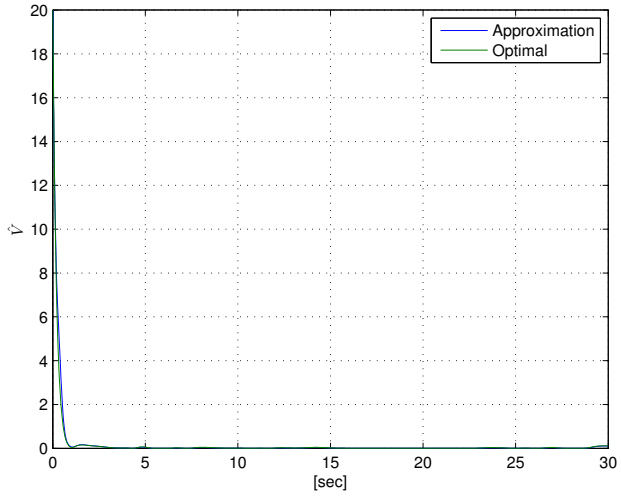

Figure 4-10. Optimal value function approximation $\hat{V}(x)$ for a zero-sum game, with a continuous persistently excited input..

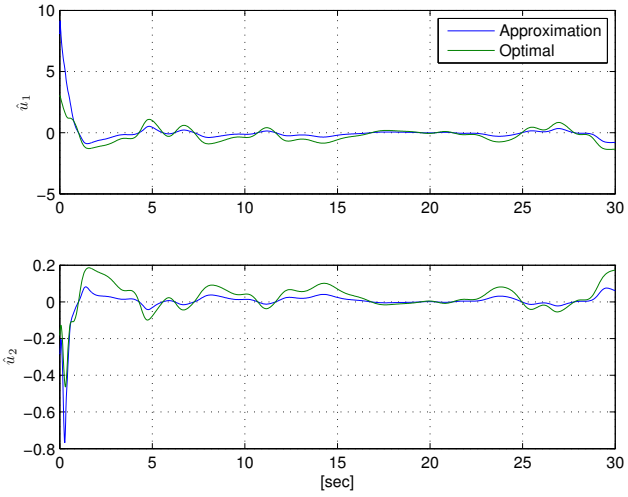

Figure 4-11. Optimal control approximations $\hat{u}_1$ and $\hat{u}_2$ in a zero-sum game, with a continuous persistently excited input.

CHAPTER 5
APPROXIMATE N-PLAYER NONZERO-SUM GAME SOLUTION FOR AN UNCERTAIN CONTINUOUS NONLINEAR SYSTEM

Chapter 4 focused on solving a two player infinite horizon zero-sum game subject to nonlinear time-invariant affine in the input dynamics. This chapter expands the technique from Chapter 4 to a more general class of problems. The focus of this chapter is the derivation of a solution to an N-player infinite horizon nonzero-sum game subject to nonlinear time-invariant affine in the input dynamics. This problem has inherent complexity as compared to the zero-sum game in the fact that a set of optimal strategies are trying to minimize a set of coupled cost functions, which lead to a set of coupled nonlinear HJB partial differential equations. Nonzero-sum differential games resemble classical optimal control problems in some respects, but due to multiple cost criteria the game formulation must be further specified as to what is demanded of an optimal solution. To approach a feasible solution for this problem an online solution method based on an approximation of the set of HJBs is presented. This technique utilizes an approximate optimal adaptive controller that has two sets of adaptive structures, a critic set to approximate for the value (cost) functions and an actor set to approximate for the control policies. In addition, a DNN is used to robustly identify the system parameters. The two adaptive structures are tuned simultaneously online to learn the solution to the set of coupled HJB equations and the set of optimal policies. This chapter presents an adaptive control method that converges online to an approximate solution set of the N-player differential game. Parameter update laws are given to tune the weights of the online critic and actor neural networks simultaneously to converge to the solution set of the coupled HJB equations and the set of Nash equilibrium policies, while also guaranteeing closed-loop stability. The set of policies guarantee uniformly ultimately bounded (UUB) tracking error for the closed-loop system.

5.1 N-player Nonzero-Sum Differential Game

Consider the N-player nonlinear time-invariant affine in the input dynamic system give by

$$\dot{x} = f(x) + \sum_{j=1}^{N} g_j(x) u_j \tag{5–1}$$

where $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state vector, $u_j(x) \in \mathcal{U} \subseteq \mathbb{R}^{m_j}$ are the control inputs, and $f(x) \in \mathbb{R}^n$, and $g(x) \in \mathbb{R}^{n \times m}$ are the drift, input matrices. Assume that $g_1(x), \ldots, g_N(x)$, and $f(x)$ are second order differentiable and Lipschitz continuous, and that $f(0) = 0$ such that $x = 0$ is an equilibrium point for Eq. 5–1. The infinite-horizon scalar cost functional $J_i(x(t), u_1, u_2, \ldots u_N)$ associated with each player can be defined as

$$J_i = \int_t^{\infty} r_i(x(s), u_1, u_2, \ldots u_N)\, ds \quad i \in N, \tag{5–2}$$

where t is the initial time, and $r_i(x, u_1, u_2, \ldots u_N) \in \mathbb{R}$ is the local cost for the state, and control, defined as

$$r_i = Q_i(x) + \sum_{j=1}^{N} u_j^T R_{ij} u_j \quad i \in N, \tag{5–3}$$

where $Q_i(x) \in \mathbb{R}$, $R_{ij} = R_{ij}^T \in \mathbb{R}^{m_j \times m_j}$ are continuously differentiable and positive definite, and $R_{ii} \in \mathbb{R}^{m_i \times m_i}$are positive definite symmetric matrices. The cost functional may also be written as [109]

$$J_i = \frac{1}{N}\sum_{j=1}^{N} J_j + \frac{1}{N}\sum_{j=1}^{N} (J_i - J_j) \equiv \bar{J} + \tilde{J}_i \quad i \in N, \tag{5–4}$$

where $\bar{J}$ is an overall cooperative *team* cost and $\tilde{J}_i$ a *conflict* cost for player i. The cost function in Eq. 5–4 can be cast as a zero-sum game by setting $\bar{J} = 0$, and can be further reduced to a two player zero-sum game when $J_1 = -J_2$. Such games have been extensively studied in control systems and result in the saddle-point Nash equilibrium solution. However, general *team* games may have both cooperative objectives and selfish objectives, which is captured in nonzero-sum games, as detailed in Eq. 5–4. The objective of the N-player game is to find a set of admissible feedback policies $(u_1^*, u_2^*, \ldots, u_N^*)$ such that the

value function $V_i\left(x\left(t\right), u_1, u_2, \ldots u_N\right)$ given in Eq. 5–2

$$V_i = \min_{u_i} \int_t^{\infty} \left(Q_i\left(x\right) + \sum_{j=1}^{N} u_j^T R_{ij} u_j, \right) ds \quad i \in N, \tag{5–5}$$

is minimized. This chapter focuses on the Nash equilibrium solution for the N-player game, in which the following N inequalities are satisfied for all $u_i^* \in \Omega_i, i \in N$:

$$\left.\begin{aligned} V_1^* \triangleq V_1\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq V_1\left(x\left(t\right), u_1, u_2^*, \ldots u_N^*\right) \\ V_2^* \triangleq V_2\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq V_2\left(x\left(t\right), u_1^*, u_2, \ldots u_N^*\right) \\ \cdots \\ \cdots \\ V_N^* \triangleq V_N\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N^*\right) \leq V_N\left(x\left(t\right), u_1^*, u_2^*, \ldots u_N\right). \end{aligned}\right\} \tag{5–6}$$

The Nash equilibrium outcome of the N-player game is given by the N-tuple of quantities $\{V_1^*, V_2^*, \ldots, V_N^*\}$. The value functions can be alternately presented by a differential equivalent given by the following nonlinear Lyapunov equation [109]

$$0 = r\left(x, u_1, \ldots, u_N\right) + \nabla V_i^* \left(f\left(x\right) + \sum_{j=1}^{N} g_j\left(x\right) u_j \right), \; V_i^*\left(0\right) = 0, \; i \in N, \tag{5–7}$$

where $\nabla V_i^* \triangleq \frac{\partial V_i^*(x)}{\partial x} \in \mathbb{R}^{n \times 1}$. Assuming the value functional is continuously differentiable, Bellman's principle of optimality can be used to derive the following optimality condition

$$\begin{aligned} 0 \;\; = \;\; & \min_{u_i} \left[\nabla V_i^* \left(f\left(x\right) + \sum_{j=1}^{N} g_j\left(x\right) u_j \right) + r\left(x, u_1, \ldots, u_N\right) \right], \\ & V_i^*\left(0\right) = 0, \; i \in N, \end{aligned} \tag{5–8}$$

which is a N-coupled set of nonlinear PDEs, also called the HJB equation. Suitable nonnegative definite solutions to Eq. 5–7 can be used to evaluate the infinite integral Eq. 5–5 along the systems trajectories. A closed-form expression of the optimal feedback control policies are given by

$$u_i^*\left(x\right) = -\frac{1}{2} R_{ii}^{-1} g_i^T\left(x\right) \nabla V_i^* \quad i \in N. \tag{5–9}$$

The closed form expression for the optimal control policies in Eq. 5–9, obviates the need to search for a set of feedback policies that minimize the value function; however, the solutions $V_i^*(x)$ to the HJB equations given in Eq. 5–8 are required. The HJB equations in Eq. 5–8, can be rewritten by substituting for the local cost in Eq. 5–3 and the optimal control policy in Eq. 5–9, respectively, as

$$\begin{aligned} 0 &= Q_i(x) + \nabla V_i^* f(x) - \frac{1}{2}\nabla V_i^* \sum_{j=1}^{N} g_j(x) R_{jj}^{-1} g_j^T(x) \nabla V_j^* \\ &\quad + \frac{1}{4}\sum_{j=1}^{N} \nabla V_j^* g_j(x) R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x) \nabla V_j^*, \quad V_i^*(0) = 0. \end{aligned} \tag{5–10}$$

Although nonzero-sum games contain non-cooperative components, the solution to each player's coupled HJB equation in Eq. 5–10 requires knowledge of all the other player's strategies in Eq. 5–9. The underlying assumption of rational opponents [41] is characteristic of differential game theory problems and it implies that the players share information, yet they agree to adhere to the equilibrium policy determined from the Nash game.

5.2 HJB Approximation via ACI

This chapter generalizes the ACI approximation architecture to solve the N-player nonzero-sum game for Eq. 5–10. The ACI architecture eliminates the need for exact model knowledge and utilizes a DNN to robustly identify the system, a critic NN to approximate the value function and an actor NN to find a set of control policies which minimizes the value functions. This section introduces the ACI architecture for the N-player game, and subsequent sections give details of the design for the N-player nonzero-sum game solution.

The Hamiltonian $H_i(x, \nabla V_{x_i}, u_1, \ldots, u_N)$ of the system in Eq. 5–1 can be defined as

$$H_i = r_{u_i} + \nabla V_i F_u, \;\; i \in N, \tag{5–11}$$

where ∇V_i denotes the Jacobian of the value functions V_i, $F_u(x, u_1, \ldots, u_N) \triangleq f(x) + \sum_{j=1}^{N} g_j(x) u_j \in \mathbb{R}^n$ denotes the system dynamics, and $r_{u_i}(x, u_1, \ldots, u_N) \triangleq Q_i(x) + \sum_{j=1}^{N} u_j^T R_{ij} u_j$ denotes the local cost. The optimal policies in Eq. 5–9 and the associated value functions $V_i^*(x)$ satisfy the HJB equation

$$H_i(x, \nabla V_i^*, u_1^*, \ldots, u_N^*) = r_{u_i^*} + \nabla V_i^* F_{u^*} = 0 \;\; i \in N. \tag{5–12}$$

Replacing the optimal Jacobian ∇V_i^* and optimal control policies u_i^* by estimates $\nabla \hat{V}_i$ and $\hat{u}_i$, respectively, yields the approximate HJB equation

$$H_i\left(x, \nabla \hat{V}_i, \hat{u}_1, \ldots, \hat{u}_N\right) = r_{\hat{u}_i} + \nabla \hat{V}_i F_{\hat{u}}, \;\; i \in N. \tag{5–13}$$

It is evident that the approximate HJB in Eq. 5–13 is dependent on the complete knowledge of the system. To overcome this limitation, an online system identifier replaces the system dynamics which modifies the approximate HJB in Eq. 5–13, and is defined as

$$H_i\left(x, \hat{x}, \nabla \hat{V}_i, \hat{u}_1, \ldots, \hat{u}_N\right) = r_{\hat{u}_i} + \nabla \hat{V}_i \hat{F}_{\hat{u}}, \;\; i \in N, \tag{5–14}$$

where $\hat{F}_{\hat{u}}$ is an approximation of the system dynamics $F_{\hat{u}}$. Taking the error between the optimal and approximate HJB equations in Eqs. 5–12 and 5–14, respectively, yields the Bellman residual errors $\delta_{hjb_i}\left(x, \hat{x}, \hat{u}_i, \nabla \hat{V}_i\right)$ defined as

$$\delta_{hjb_i} \;\triangleq\; H_i\left(x, \hat{x}, \nabla \hat{V}_i, \hat{u}_1, \ldots, \hat{u}_N\right) - H_i(x, \nabla V_i^*, u_1^*, \ldots, u_N^*) \;\; i \in N. \tag{5–15}$$

However since $H_i(x, \nabla V_i^*, u_1^*, \ldots, u_N^*) = 0 \quad \forall i \in N$ then the Bellman residual error can be defined in a measurable form as

$$\delta_{hjb_i} = H_i\left(x, \hat{x}, \nabla \hat{V}_i, \hat{u}_1, \ldots, \hat{u}_N\right) \;\; i \in N.$$

The objective is to update both $\hat{u}_i$ (actors) and $\hat{V}_i$ (critics) simultaneously, based on the minimization of the Bellman residual errors δ_{hjb_i}. All together, the actors $\hat{u}_i$, the critics $\hat{V}_i$,

and the identifier $\hat{F}_{\hat{u}}$ constitute the ACI architecture. Assumptions 4-1 through 4-6 from Chapter 4 are used in this derivation to facilitate the subsequent analysis.

5.3 System Identifier

Consider the two player case for the dynamics given in Eq. 5–1 as

$$\dot{x} = f(x) + g_1(x) u_1(x) + g_2(x) u_2(x), \quad x(0) = x_0, \tag{5–16}$$

where $u_1(x), u_2(x) \in \mathbb{R}^n$ are the control inputs, and the state $x(t) \in \mathbb{R}^n$ is assumed to be measurable. The system identifier is identical to the identifier presented in Chapter 4. For brevity this chapter presents the main theorem and refers to Chapter 4 for further details.

Theorem 5.1. *For the system in Eq. 5–16, the identifier developed in Eq. 4–15 along with its weight update laws in Eq. 4–21 ensures asymptotic identification of the state and its derivative, in the sense that*

$$\lim_{t\to\infty} \|\tilde{x}(t)\| = 0 \qquad \textit{and} \qquad \lim_{t\to\infty} \left\|\dot{\tilde{x}}(t)\right\| = 0,$$

provided Assumptions 4-4 through 4-6 hold, and the control gains k and γ_f are chosen sufficiently large based on the initial conditions of the states, and satisfy the following sufficient conditions

$$\alpha\gamma_f > \zeta_5, \qquad k > \zeta_6, \tag{5–17}$$

where ζ_5 and ζ_6 are introduced in Eq. 4–28, and β_1, β_2 introduced in Eq. 4–31, are chosen according to the sufficient conditions in Eq. 4–32.

Proof. Refer to Theorem 4-1 in Chapter 4. □

5.4 Actor-Critic Design

Using Assumption 5-1 and Eq. 5–9, the optimal value function and the optimal controls can be represented by NNs as

$$V_1^*(x) = W_1^T\phi_1(x) + \varepsilon_1(x), \qquad u_1^*(x) = -\frac{1}{2}R_{11}^{-1}g_1^T(x)\underbrace{\left(\phi_1'(x)^T W_1 + \varepsilon_1'(x)^T\right)}_{\nabla V_1^*}, \tag{5–18}$$
$$V_2^*(x) = W_2^T\phi_2(x) + \varepsilon_2(x), \qquad u_2^*(x) = -\frac{1}{2}R_{22}^{-1}g_2^T(x)\underbrace{\left(\phi_2'(x)^T W_2 + \varepsilon_2'(x)^T\right)}_{\nabla V_2^*},$$

where $W_1, W_2 \in \mathbb{R}^N$ are unknown ideal NN weights, N is the number of neurons, $\phi_i(x) = [\phi_{i1}(x)\ \ \phi_{i2}(x). . . \phi_{iN}(x)]^T \in \mathbb{R}^N$ are smooth NN activation functions, such that $\phi_{ij}(0) = 0$ and $\phi_{ij}'(0) = 0 \quad j = 1...N$ and $i = 1, 2$, and $\varepsilon_1(\cdot), \varepsilon_2(\cdot) \in \mathbb{R}$ are the function reconstruction errors.

Assumption 5.1. *The NN activation function* $\{\phi_{ij}(x) : j = 1...N,\ i = 1, 2\}$ *are chosen such that as* $N \to \infty$, $\phi_i(x)$ *provides a complete independent basis for* $V_i^*(x)$.

Using Assumption 5-1 and Weierstrass higher-order approximation theorem, both $V_i^*(x)$ and ∇V_i^* can be uniformly approximated by NNs in Eq. 5–18, i.e. as $N \to \infty$, the approximation errors $\varepsilon_i(x), \varepsilon_i'(x) \to 0$ for $i = 1, 2$, respectively. The critic $\hat{V}_i(x)$ and the actor $\hat{u}_i(x)$ approximate the optimal value function $V_i^*(x)$ and the optimal controls $u_i^*(x)$ in Eq. 5–18, and are given as

$$\hat{V}_1(x) = \hat{W}_{1c}^T\phi_1(x), \qquad \hat{u}_1(x) = -\frac{1}{2}R_{11}^{-1}g_1^T(x)\phi_1'(x)\hat{W}_{1a}, \tag{5–19}$$
$$\hat{V}_2(x) = \hat{W}_{2c}^T\phi_2(x), \qquad \hat{u}_2(x) = -\frac{1}{2}R_{22}^{-1}g_2^T(x)\phi_2'(x)\hat{W}_{2a},$$

where $\hat{W}_{1c}(t), \hat{W}_{2c}(t) \in \mathbb{R}^N$ and $\hat{W}_{1a}(t), \hat{W}_{2a}(t) \in \mathbb{R}^N$ are estimates of the ideal weights of the critic and actor NNs, respectively. The weight estimation errors for the critic and actor are defined as $\tilde{W}_{ic}(t) \triangleq W_i - \hat{W}_{ic}(t)$ and $\tilde{W}_{ia}(t) \triangleq W_i - \hat{W}_{ia}(t)$ for $i = 1, 2$, respectively. The actor and critic NN weights are both updated based on the minimization of the Bellman error $\delta_{hjb}(\cdot)$ in Eq. 5–14, which can be rewritten by substituting $\hat{V}_1$ and $\hat{V}_2$

from Eq. 5–19 as

$$\delta_{hjb_1} = \hat{W}_{1c}^T \phi_1' \hat{F}_{\hat{u}} + r_1(x, \hat{u}_1, \hat{u}_2) = \hat{W}_{1c}^T \omega_1 + r_1(x, \hat{u}_1, \hat{u}_2), \tag{5–20}$$
$$\delta_{hjb_2} = \hat{W}_{2c}^T \phi_2' \hat{F}_{\hat{u}} + r_2(x, \hat{u}_1, \hat{u}_2) = \hat{W}_{2c}^T \omega_2 + r_2(x, \hat{u}_1, \hat{u}_2),$$

where $\omega_i(x, \hat{u}, t) \triangleq \phi_i' \hat{F}_{\hat{u}} \in \mathbb{R}^N$ for $i = 1, 2$, is the critic NN regressor vector.

Least Squares Update for the Critic. Consider the integral of the sum of the squared Bellman errors $E_c(\hat{W}_{1c}(t), \hat{W}_{2c}(t), t)$

$$E_c = \int_0^t \left(\delta_{hjb_1}^2(\tau) + \delta_{hjb_2}^2(\tau)\right) d\tau. \tag{5–21}$$

The LS update law for the critic $\hat{W}_{1c}(t)$ is generated by minimizing the total prediction error in Eq. 5–21

$$\frac{\partial E_c}{\partial \hat{W}_{1c}} = 2 \int_0^t \delta_{hjb_1}(\tau) \frac{\partial \delta_{hjb_1}(\tau)}{\partial \hat{W}_{1c}(\tau)} d\tau = 0$$
$$= \hat{W}_{1c}^T(t) \int_0^t \omega_1(\tau)\omega_1(\tau)^T d\tau + \int_0^t \omega_1(\tau)^T r_1(\tau) d\tau = 0$$
$$\hat{W}_{1c}(t) = -\left(\int_0^t \omega_1(\tau)\omega_1(\tau)^T d\tau\right)^{-1} \int_0^t \omega_1(\tau) r_1(\tau) d\tau,$$

which gives the LS estimate of the critic weights, provided the inverse $\left(\int_0^t \omega_1(\tau)\omega_1(\tau)^T d\tau\right)^{-1}$ exists. Likewise, the LS update law for the critic $\hat{W}_{2c}(t)$ is generated by

$$\hat{W}_{2c}(t) = -\left(\int_0^t \omega_2(\tau)\omega_2(\tau)^T d\tau\right)^{-1} \int_0^t \omega_2(\tau) r_2(\tau) d\tau.$$

The recursive formulation of the normalized LS algorithm [130] gives the update laws for the two critic weights as

$$\begin{aligned}\dot{\hat{W}}_{1c} &= -\eta_{1c}\Gamma_{1c}\frac{\omega_1}{1+\nu_1\omega_1^T\Gamma_{1c}\omega_1}\delta_{hjb_1}, \\ \dot{\hat{W}}_{2c} &= -\eta_{2c}\Gamma_{2c}\frac{\omega_2}{1+\nu_2\omega_2^T\Gamma_{2c}\omega_2}\delta_{hjb_2},\end{aligned} \tag{5–22}$$

where $\nu_1, \nu_2, \eta_{1c}, \eta_{2c} \in \mathbb{R}$ are constant positive gains and $\Gamma_{ic}(t) \triangleq \left(\int_0^t \omega(\tau)\omega(\tau)^T\, d\tau\right)^{-1} \in \mathbb{R}^{N\times N}$ for $i = 1, 2$, are symmetric estimation gain matrices generated by

$$\begin{aligned}\dot{\Gamma}_{1c} &= -\eta_{1c}\Gamma_{1c}\frac{\omega_1{\omega_1}^T}{1+\nu_1\omega_1^T\Gamma_{1c}\omega_1}\Gamma_{1c}, \qquad \Gamma_{1c}(t_{r1}^+) = \Gamma_{1c}(0) = \varphi_{01}I, \\ \dot{\Gamma}_{c2} &= -\eta_{2c}\Gamma_{2c}\frac{\omega_2{\omega_2}^T}{1+\nu_2\omega_2^T\Gamma_{2c}\omega_2}\Gamma_{2c}, \qquad \Gamma_{2c}(t_{r2}^+) = \Gamma_{2c}(0) = \varphi_{02}I,\end{aligned} \tag{5–23}$$

where t_{r1}^+ and t_{r2}^+ are the resetting times at which $\lambda_{min}\{\Gamma_{1c}(t)\} \leq \varphi_1$ and $\lambda_{min}\{\Gamma_{2c}(t)\} \leq \varphi_2$, $\varphi_{01} > \varphi_1 > 0$, and $\varphi_{02} > \varphi_2 > 0$. The covariance resetting ensures that $\Gamma_{1c}(t)$ and $\Gamma_{2c}(t)$ are positive-definite for all time and prevent arbitrarily small values in some directions, which would make adaptation in those directions very slow (also called the covariance wind-up problem) [130]. From Eq. 5–23 it is clear that $\dot{\Gamma}_{1c} \leq 0$ and $\dot{\Gamma}_{2c} \leq 0$, which means that the covariance matrices $(\Gamma_{1c}(t), \Gamma_{2c}(t))$ can be bounded as follows

$$\varphi_1 I \leq \Gamma_{1c} \leq \varphi_{01} I, \qquad \varphi_2 I \leq \Gamma_{2c} \leq \varphi_{02} I. \tag{5–24}$$

Gradient Update for the Actor. The actor update is also based on the minimization of the Bellman errors $\delta_{hjb_i}(\cdot)$. However, unlike the critic weights, the actor weights appear nonlinearly in $\delta_{hjb_i}(\cdot)$, making it problematic to develop a LS update law. Hence, a gradient update law is developed for the actor which minimizes the squared Bellman error $E_a(t) \triangleq \delta_{hjb_1}^2 + \delta_{hjb_2}^2$, whose gradients are given as

$$\begin{aligned}\frac{\partial E_a}{\partial \hat{W}_{1a}} &= (\hat{W}_{1a} - \hat{W}_{1c})^T\phi_1' G_1 \phi_1'^T \delta_{hjb_1} + (\hat{W}_{1a}\phi_1' G_{21} - \hat{W}_{2c}\phi_2' G_2)^T \phi_1'^T \delta_{hjb_2}, \\ \frac{\partial E_a}{\partial \hat{W}_{2a}} &= (\hat{W}_{2a}\phi_2' G_{12} - \hat{W}_{1c}\phi_1' G_1)^T \phi_2'^T \delta_{hjb_1} + (\hat{W}_{2a} - \hat{W}_{2c})^T \phi_2' G_2 \phi_2'^T \delta_{hjb_2},\end{aligned} \tag{5–25}$$

where $G_i \triangleq g_i R_{ii}^{-1} g_i \in \mathbb{R}^{n\times n}$ and $G_{ji} \triangleq g_i R_{ii}^{-1} R_{ji} R_{ii}^{-1} g_i \in \mathbb{R}^{n\times n}$, for $i = 1, 2$ and $j = 1, 2$, are symmetric matrices. Using Eq. 5–25, the actor NNs are updated as

$$\dot{\hat{W}}_{1a} = proj\left\{-\frac{\Gamma_{11a}}{\sqrt{1+\omega_1^T\omega_1}}\frac{\partial E_a}{\partial \hat{W}_{1a}} - \Gamma_{12a}(\hat{W}_{1a} - \hat{W}_{1c})\right\}, \tag{5–26}$$
$$\dot{\hat{W}}_{2a} = proj\left\{-\frac{\Gamma_{21a}}{\sqrt{1+\omega_2^T\omega_2}}\frac{\partial E_a}{\partial \hat{W}_{2a}} - \Gamma_{22a}(\hat{W}_{2a} - \hat{W}_{2c})\right\},$$

where $\Gamma_{11a}, \Gamma_{12a}, \Gamma_{21a}, \Gamma_{22a} \in \mathbb{R}$ are positive adaptation gains, and $proj\{\cdot\}$ is a projection operator used to bound the weight estimates [127], [128]. Using Assumption 4-2 and the projection algorithm in Eq. 5–26, the actor NN weight estimation error can be bounded as

$$\left\|\tilde{W}_{1a}\right\| \leq \kappa_1; \qquad \left\|\tilde{W}_{2a}\right\| \leq \kappa_2, \tag{5–27}$$

where $\kappa_1, \kappa_2 \in \mathbb{R}$ is some positive constants. The first term in Eq. 5–26 is normalized and the last term is added as feedback for stability (based on the subsequent stability analysis).

5.5 Stability Analysis

The dynamics of the critic weight estimation errors $\tilde{W}_{1c}(t)$ and $\tilde{W}_{2c}(t)$ can be developed using Eqs. 5–11-5–14, 5–20 and 5–22, as

$$\begin{aligned}
\dot{\tilde{W}}_{1c} &= \eta_{1c}\Gamma_{1c}\frac{\omega_1}{1+\nu_1\omega_1^T\Gamma_{1c}\omega_1}\left[-\tilde{W}_{1c}^T\omega_1 - W_1^T\phi_1'\tilde{F}_{\hat{u}} - \varepsilon_1' F_{u^*} + \hat{u}_1^T R_{11}\hat{u}_1\right. \\
&\quad \left. -u_1^{*T} R_{11} u_1^* - u_2^{*T} R_{12} u_2^* + W_1^T\phi_1'\left(g_1(\hat{u}_1 - u_1^*) + g_2(\hat{u}_2 - u_2^*)\right) + \hat{u}_2^T R_{12}\hat{u}_2\right], \\
\dot{\tilde{W}}_{2c} &= \eta_{2c}\Gamma_{2c}\frac{\omega_2}{1+\nu_2\omega_2^T\Gamma_{2c}\omega_2}\left[-\tilde{W}_{2c}^T\omega_2 - W_2^T\phi_2'\tilde{F}_{\hat{u}} - \varepsilon_2' F_{u^*} + \hat{u}_2^T R_{22}\hat{u}_2\right. \\
&\quad \left. -u_1^{*T} R_{21} u_1^* - u_2^{*T} R_{22} u_2^* + W_2^T\phi_2'\left(g_1(\hat{u}_1 - u_1^*) + g_2(\hat{u}_2 - u_2^*)\right) + \hat{u}_1^T R_{21}\hat{u}_1\right].
\end{aligned} \tag{5–28}$$

Substituting for $(u_1^*(x), u_2^*(x))$ and $(\hat{u}_1(x), \hat{u}_2(x))$ from Eqs. 5–18 and 5–19, respectively, in Eq. 5–28 yields

$$\begin{aligned}
\dot{\tilde{W}}_{1c} &= -\eta_{1c}\Gamma_{1c}\psi_1\psi_1^T\tilde{W}_{1c} + \frac{\eta_{1c}\Gamma_{1c}\omega_1}{1+\nu_1\omega_1^T\Gamma_{1c}\omega_1}\left[-W_1^T\phi_1'\tilde{F}_{\hat{u}} - \varepsilon_1' F_{u^*}\right. \\
&\quad + \frac{1}{4}\tilde{W}_{1a}^T\phi_1' G_1 \phi_1'^T \tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T\phi_2' G_{12}\phi_2'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon_1' G_1 \varepsilon_1'^T - \frac{1}{4}\varepsilon_2' G_{12}\varepsilon_2'^T \\
&\quad \left. + \frac{1}{2}\left(\tilde{W}_{2a}\phi_2' + \varepsilon_2'^T\right)\left(G_2\phi_1'^T W_1 - G_{12}\phi_2'^T W_2\right)\right], \\
\dot{\tilde{W}}_{2c} &= -\eta_{2c}\Gamma_{2c}\psi_2\psi_2^T\tilde{W}_{2c} + \frac{\eta_{2c}\Gamma_{2c}\omega_2}{1+\nu_2\omega_2^T\Gamma_{2c}\omega_2}\left[-W_2^T\phi_2'\tilde{F}_{\hat{u}} - \varepsilon_2' F_{u^*}\right. \\
&\quad + \frac{1}{4}\tilde{W}_{1a}^T\phi_1' G_{21}\phi_1'^T\tilde{W}_{1a} + \frac{1}{4}\tilde{W}_{2a}^T\phi_2' G_2\phi_2'^T\tilde{W}_{2a} - \frac{1}{4}\varepsilon_2' G_2\varepsilon_2'^T - \frac{1}{4}\varepsilon_1' G_{21}\varepsilon_1'^T \\
&\quad \left. + \frac{1}{2}\left(\tilde{W}_{1a}\phi_1' + \varepsilon_1'^T\right)\left(G_1\phi_2'^T W_2 - G_{21}\phi_1'^T W_1\right)\right],
\end{aligned} \tag{5–29}$$

where $\psi_i(t) \triangleq \frac{\omega_i(t)}{\sqrt{1+\nu_i\omega_i(t)^T\Gamma_{ic}(t)\omega_i(t)}} \in \mathbb{R}^N$ are the normalized critic regressor vectors for $i = 1, 2$, respectively, bounded as

$$\|\psi_1\| \leq \frac{1}{\sqrt{\nu_1\varphi_{11}}}, \qquad \|\psi_2\| \leq \frac{1}{\sqrt{\nu_2\varphi_{12}}}, \tag{5–30}$$

where φ_{11} and φ_{12} are introduced in Eq. 5–24. The error systems in Eq. 5–29 can be represented as the following perturbed systems

$$\dot{\tilde{W}}_{1c} = \Omega_1 + \Lambda_{01}\Delta_1, \qquad \dot{\tilde{W}}_{2c} = \Omega_2 + \Lambda_{02}\Delta_2, \tag{5–31}$$

where $\Omega_i(\tilde{W}_{ic}, t) \triangleq -\eta_{ic}\Gamma_{ic}\psi_i\psi_i^T\tilde{W}_{ic} \in \mathbb{R}^N \quad i = 1, 2$, denotes the nominal system, $\Lambda_{0i} \triangleq \frac{\eta_{ic}\Gamma_{ic}\omega_i}{1+\nu_i\omega_i^T\Gamma_{ic}\omega_i}$ denotes the perturbation gain, and

$$\begin{aligned}
\Delta_i(t) \triangleq &\left[-W_i^T\phi_i'\tilde{F}_{\hat{u}} + \frac{1}{4}\tilde{W}_{ia}^T\phi_i' G_i\phi_i'^T\tilde{W}_{ia} - \varepsilon_i' F_{u^*}\right. \\
&+ \frac{1}{4}\tilde{W}_{ka}^T\phi_k' G_{ik}\phi_k'^T\tilde{W}_{ka} - \frac{1}{4}\varepsilon_k' G_{ik}\varepsilon_k'^T - \frac{1}{4}\varepsilon_i' G_i\varepsilon_i'^T \\
&\left. + \frac{1}{2}\left(\tilde{W}_{ka}\phi_k' + \varepsilon_k'^T\right)\left(G_k\phi_i'^T W_i - G_{ik}\phi_k'^T W_k\right)\right] \in \mathbb{R}^N,
\end{aligned}$$

where $i = 1, 2$ and $k = 3 - i$, denotes the perturbations. Using Theorem 2.5.1 in [130], it can be shown that the nominal systems

$$\dot{\tilde{W}}_{1c} = -\eta_{1c}\Gamma_{1c}\psi_1\psi_1^T\tilde{W}_{1c}, \qquad \dot{\tilde{W}}_{2c} = -\eta_{2c}\Gamma_{2c}\psi_2\psi_2^T\tilde{W}_{2c}, \tag{5–32}$$

are exponentially stable, if the bounded signals $(\psi_1(t), \psi_2(t))$ are PE, i.e.

$$\mu_{i2}I \geq \int_{t_0}^{t_0+\delta_i} \psi_i(\tau)\psi_i(\tau)^T d\tau \geq \mu_{i1}I \qquad \forall t_0 \geq 0,\ i = 1, 2,$$

where $\mu_{i1}, \mu_{i2}, \delta_i \in \mathbb{R}$ are some positive constants. Since $\Omega_i(\tilde{W}_c, t)$ is continuously differentiable and the Jacobian $\frac{\partial \Omega_i}{\partial \tilde{W}_{ic}} = -\eta_{ic}\Gamma_{ic}\psi_i\psi_i^T$ is bounded for the exponentially stable system Eq. 5–32 for $i = 1, 2$, the converse Lyapunov Theorem 4.14 in [131] can be used to show that there exists a function $V_c : \mathbb{R}^N \times \mathbb{R}^N \times [0,\ \infty) \to \mathbb{R}$, which satisfies the following inequalities

$$\begin{gathered}
c_{11}\left\|\tilde{W}_{1c}\right\|^2 + c_{12}\left\|\tilde{W}_{2c}\right\|^2 \leq V_c(\tilde{W}_{1c}, \tilde{W}_{2c}, t) \leq c_{21}\left\|\tilde{W}_{1c}\right\|^2 + c_{22}\left\|\tilde{W}_{2c}\right\|^2 \\
\frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_{1c}}\Omega_1(\tilde{W}_{1c}, t) + \frac{\partial V_c}{\partial \tilde{W}_{2c}}\Omega_2(\tilde{W}_{2c}, t) \leq -c_{31}\left\|\tilde{W}_{1c}\right\|^2 - c_{32}\left\|\tilde{W}_{2c}\right\|^2 \\
\left\|\frac{\partial V_c}{\partial \tilde{W}_{1c}}\right\| \leq c_{41}\left\|\tilde{W}_{1c}\right\| \qquad \left\|\frac{\partial V_c}{\partial \tilde{W}_{2c}}\right\| \leq c_{42}\left\|\tilde{W}_{2c}\right\|,
\end{gathered} \tag{5–33}$$

for some positive constants $c_{1i},\ c_{2i},\ c_{3i}, c_{4i} \in \mathbb{R}$ for $i = 1, 2$. Using Assumptions 4-1 through 4-6 and 5-1, the projection bounds in Eq. 5–26, the fact that $F_{u^*} \in \mathcal{L}_\infty$ (since $(u_1^*(x), u_2^*(x))$ is stabilizing), and provided the conditions of Theorem 5-1 hold (required to prove that $\tilde{F}_{\hat{u}} \in \mathcal{L}_\infty$), the following bounds are developed to facilitate the subsequent

stability proof

$$\begin{aligned}
\left\|\tilde{W}_{1a}\right\| \leq \kappa_1; \qquad \left\|\tilde{W}_{2a}\right\| \leq \kappa_2, \\
\left\|\phi_1' G_1 \phi_1'^T\right\| \leq \kappa_3; \qquad \left\|\phi_2' G_2 \phi_2'^T\right\| \leq \kappa_4, \\
\|\Delta_1\| \leq \kappa_5; \qquad \|\Delta_2\| \leq \kappa_6, \\
\left\|\frac{1}{2}\left(W_1^T\phi_1' + W_2^T\phi_2' + \varepsilon_1' + \varepsilon_2'\right)\left(G_1\varepsilon_1'^T + G_2\varepsilon_2'^T\right)\right\| \leq \kappa_7, \\
\left\|\frac{1}{2}\left(W_1^T\phi_1' + W_2^T\phi_2' + \varepsilon_1' + \varepsilon_2'\right)\left(G_1\phi_1'^T\tilde{W}_{1a} + G_2\phi_2'^T\tilde{W}_{2a}\right)\right\| \leq \kappa_8, \\
\left\|\phi_1' G_{21} \phi_1'^T\right\| \leq \kappa_9; \qquad \left\|\phi_2' G_2 \phi_1'^T\right\| \leq \kappa_{10}, \\
\left\|\phi_1' G_1 \phi_2'^T\right\| \leq \kappa_{11}; \qquad \left\|\phi_2' G_{12} \phi_2'^T\right\| \leq \kappa_{12},
\end{aligned} \tag{5–34}$$

where $\kappa_j \in \mathbb{R}$ for $j = 1, \ldots, 12$ are computable positive constants.

Theorem 5.2. *If Assumptions 4-1-4-6 and 5-1 hold, the regressors* $\psi_i(t) \triangleq \frac{\omega_i}{\sqrt{1+\omega_i^T\Gamma_{ic}\omega_i}}$ *for* $i = 1, 2$ *are PE, and provided Eq. 4–32, Eq. 5–17 and the following sufficient gain conditions are satisfied*

$$\frac{c_{31}}{\Gamma_{11a}} > \kappa_1\kappa_3; \qquad \frac{c_{32}}{\Gamma_{21a}} > \kappa_2\kappa_4,$$

where Γ_{11a}, Γ_{21a}, c_{31}, c_{32}, κ_1, κ_2, κ_3, *and* κ_4 *are introduced in Eqs. 5–26, 5–33, and 5–34, then the controller in Eq. 5–19, the actor-critic weight update laws in Eqs. 5–22-5–23 and 5–26, and the identifier in Eqs. 4–15 and 4–21, guarantee that the state of the system* $x(t)$, *and the actor-critic weight estimation errors* $\left(\tilde{W}_{1a}(t), \tilde{W}_{2a}(t)\right)$ *and* $\left(\tilde{W}_{1c}(t), \tilde{W}_{2c}(t)\right)$ *are UUB.*

Proof. To investigate the stability of the the system Eq. 5–16 with control $(\hat{u}_1, \hat{u}_2)$, and the perturbed system Eq. 5–31, consider $V_L : \mathcal{X} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times [0, \infty) \to \mathbb{R}$ as the continuously differentiable, positive-definite Lyapunov function candidate, given as

$$V_L(x, \tilde{W}_{1c}, \tilde{W}_{2c}, \tilde{W}_{1a}, \tilde{W}_{2a}, t) \triangleq V_1^*(x) + V_2^*(x) + V_c(\tilde{W}_{1c}, \tilde{W}_{2c}, t) + \frac{1}{2}\tilde{W}_{1a}^T\tilde{W}_{1a} + \frac{1}{2}\tilde{W}_{2a}^T\tilde{W}_{2a},$$

where $V_i^*(x)$ for $i = 1, 2$ (the optimal value function for Eq. 5–16, are the Lyapunov function for Eq. 5–16, and $V_c(\tilde{W}_c, t)$ is the Lyapunov function for the exponentially stable system in Eq. 5–32. Since $(V_1^*(x), V_2^*(x))$ are continuously differentiable and positive-definite from Eq. 5–5, from Lemma 4.3 in [131], there exist class $\mathcal{K}$ functions α_1 and α_2 defined on $[0,\ r]$, where $B_r \subset \mathcal{X}$, such that

$$\alpha_1(\|x\|) \leq V_1^*(x) + V_2^*(x) \leq \alpha_2(\|x\|) \qquad \forall x \in B_r. \tag{5–35}$$

Using Eqs. 5–33 and 5–35, $V_L(x, \tilde{W}_{1c}, \tilde{W}_{2c}, \tilde{W}_{1a}, \tilde{W}_{2a}, t)$ can be bounded as

$$\begin{aligned} \alpha_1(\|x\|) + c_{11}\left\|\tilde{W}_{1c}\right\|^2 + c_{12}\left\|\tilde{W}_{2c}\right\|^2 + \frac{1}{2}\left(\left\|\tilde{W}_{1a}\right\|^2 + \left\|\tilde{W}_{2a}\right\|^2\right) &\leq V_L, \\ \alpha_2(\|x\|) + c_{21}\left\|\tilde{W}_{1c}\right\|^2 + c_{22}\left\|\tilde{W}_{2c}\right\|^2 + \frac{1}{2}\left(\left\|\tilde{W}_{1a}\right\|^2 + \left\|\tilde{W}_{2a}\right\|^2\right) &\geq V_L, \end{aligned}$$

which can be written as

$$\alpha_3(\|w\|) \leq V_L(x, \tilde{W}_{1c}, \tilde{W}_{2c}, \tilde{W}_{1a}, \tilde{W}_{2a}, t) \leq \alpha_4(\|w\|) \qquad \forall w \in B_s,$$

where $w(t) \triangleq [x(t)^T\ \tilde{W}_{1c}(t)^T\ \tilde{W}_{2c}(t)^T\ \tilde{W}_{1a}(t)^T\ \tilde{W}_{2a}(t)^T]^T \in \mathbb{R}^{n+4N}$, α_3 and α_4 are class $\mathcal{K}$ functions defined on $[0,\ s]$, where $B_s \subset \mathcal{X} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N$. Taking the time derivative of $V_L(\cdot)$ yields

$$\begin{aligned} \dot{V}_L &= \left(\nabla V_1^* + \nabla V_2^*\right)\left(f + g_1\hat{u}_1 + g_2\hat{u}_2\right) + \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_{1c}}\Omega_1 \\ &\quad + \frac{\partial V_c}{\partial \tilde{W}_{1c}}\Lambda_{01}\Delta_1 + \frac{\partial V_c}{\partial \tilde{W}_{2c}}\Omega_2 + \frac{\partial V_c}{\partial \tilde{W}_{2c}}\Lambda_{02}\Delta_2 - \tilde{W}_{1a}^T\dot{\hat{W}}_{1a} - \tilde{W}_{2a}^T\dot{\hat{W}}_{2a}, \end{aligned} \tag{5–36}$$

where the time derivatives of $V_i^*(\cdot)$ for $i = 1, 2$, are taken along the the trajectories of the system Eq. 5–16 with control inputs $(\hat{u}_1(\cdot), \hat{u}_2(\cdot))$ and the time derivative of $V_c(\cdot)$ is taken along the along the trajectories of the perturbed system Eq. 5–31. Using the HJB equation Eq. 5–12, $\nabla V_i^* f = -\nabla V_i^*\left(g_1 u_1^* + g_2 u_2^*\right) - Q_i(x) - \sum_{j=1}^{2} u_j^T R_{ij} u_j$ for $i = 1, 2$. Substituting for the $\nabla V_i^* f$ terms in Eq. 5–36, using the fact that $\nabla V_i^* g_i = -2{u_i^*}^T R_{ii}$ from

Eq. 5–9, and using Eqs. 5–26 and 5–33, Eq. 5–36 can be upper bounded as

$$\begin{aligned}\dot{V}_L \leq & -Q - u_1^{*^T}\left(R_{11}+R_{21}\right)u_1^* - u_2^{*^T}\left(R_{22}+R_{12}\right)u_2^* - c_{31}\left\|\tilde{W}_{1c}\right\|^2 - c_{32}\left\|\tilde{W}_{2c}\right\|^2 \\ & + c_{41}\Lambda_{01}\left\|\tilde{W}_{1c}\right\|\|\Delta_1\| + c_{42}\Lambda_{02}\left\|\tilde{W}_{2c}\right\|\|\Delta_2\| + 2u_1^{*^T}R_{11}(u_1^*-\hat{u}_1) + 2u_2^{*^T}R_{22}(u_2^*-\hat{u}_2) \\ & + \tilde{W}_{1a}^T\left[\frac{\Gamma_{11a}}{\sqrt{1+\omega_1^T\omega_1}}\frac{\partial E_a}{\partial \hat{W}_{1a}} + \Gamma_{12a}(\hat{W}_{1a}-\hat{W}_{1c})\right] + \nabla V_1^* g_2\left(\hat{u}_2-u_2^*\right) \\ & + \tilde{W}_{2a}^T\left[\frac{\Gamma_{21a}}{\sqrt{1+\omega_2^T\omega_2}}\frac{\partial E_a}{\partial \hat{W}_{2a}} + \Gamma_{22a}(\hat{W}_{2a}-\hat{W}_{2c})\right] + \nabla V_2^* g_1\left(\hat{u}_1-u_1^*\right),\end{aligned} \tag{5–37}$$

where $Q\left(x\right) \triangleq Q_1\left(x\right) + Q_2\left(x\right)$. Substituting for u_i^*, $\hat{u}_i$, δ_{hjb_i}, and Δ_i for $i = 1, 2$ using Eqs. 5–9, 5–19, 5–28, and 5–31, respectively, and using Eqs. 5–24 and 5–30 in Eq. 5–37, yields

$$\begin{aligned}\dot{V}_L \leq & -Q - c_{31}\left\|\tilde{W}_{1c}\right\|^2 - c_{32}\left\|\tilde{W}_{2c}\right\|^2 - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 - \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|^2 \\ & + \frac{1}{2}\left(W_1^T\phi_1' + W_2^T\phi_2' + \varepsilon_1' + \varepsilon_2'\right)\left(G_1\varepsilon_1'^T + G_2\varepsilon_2'^T\right) \\ & + \frac{1}{2}\left(W_1^T\phi_1' + W_2^T\phi_2' + \varepsilon_1' + \varepsilon_2'\right)\left(G_1\phi_1'^T\tilde{W}_{1a} + G_2\phi_2'^T\tilde{W}_{2a}\right) \\ & + c_{41}\frac{\eta_{1c}\varphi_{01}}{2\sqrt{\nu_1\varphi_{11}}}\|\Delta_1\|\left\|\tilde{W}_{1c}\right\| + c_{42}\frac{\eta_{2c}\varphi_{02}}{2\sqrt{\nu_2\varphi_{12}}}\|\Delta_2\|\left\|\tilde{W}_{2c}\right\| \\ & + \frac{\Gamma_{11a}}{\sqrt{1+\omega_1^T\omega_1}}\tilde{W}_{1a}^T\left((\tilde{W}_{1c}-\tilde{W}_{1a})^T\phi_1'G_1\phi_1'^T\left(-\tilde{W}_{1c}^T\omega_1+\Delta_1\right)\right. \\ & \left.+(\tilde{W}_{1c}\phi_1'G_{21}-\tilde{W}_{2a}\phi_2'G_2)^T\phi_1'^T\left(-\tilde{W}_{2c}^T\omega_2+\Delta_2\right)\right) + \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|\left\|\tilde{W}_{1c}\right\| \\ & + \frac{\Gamma_{21a}}{\sqrt{1+\omega_2^T\omega_2}}\tilde{W}_{2a}^T\left((\tilde{W}_{2c}\phi_2'G_{12}-\tilde{W}_{1a}\phi_1'G_1)^T\phi_2'^T\left(-\tilde{W}_{1c}^T\omega_1+\Delta_1\right)\right. \\ & \left.+(\tilde{W}_{2c}-\tilde{W}_{2a})^T\phi_2'G_2\phi_2'^T\left(-\tilde{W}_{2c}^T\omega_2+\Delta_2\right)\right) + \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|\left\|\tilde{W}_{2c}\right\| \\ & + \frac{\Gamma_{11a}}{\sqrt{1+\omega_1^T\omega_1}}\tilde{W}_{1a}^T\left((W_1\phi_1'G_{21}-W_2\phi_2'G_2)^T\phi_1'^T\left(-\tilde{W}_{2c}^T\omega_2+\Delta_2\right)\right) \\ & + \frac{\Gamma_{21a}}{\sqrt{1+\omega_2^T\omega_2}}\tilde{W}_{2a}^T\left((W_2\phi_2'G_{12}-W_1\phi_1'G_1)^T\phi_2'^T\left(-\tilde{W}_{1c}^T\omega_1+\Delta_1\right)\right).\end{aligned} \tag{5–38}$$

Using the bounds developed in Eq. 5–34, Eq. 5–38 can be further upper bounded as

$$\begin{aligned}\dot{V}_L \leq & -Q - (c_{31} - \Gamma_{11a}\kappa_1\kappa_3)\left\|\tilde{W}_{1c}\right\|^2 - (c_{32} - \Gamma_{21a}\kappa_2\kappa_4)\left\|\tilde{W}_{2c}\right\|^2 \\ & - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 - \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|^2 + \Phi_1\left\|\tilde{W}_{1c}\right\| + \Phi_2\left\|\tilde{W}_{2c}\right\| \\ & + \Gamma_{11a}\left(\kappa_1^2\kappa_3\kappa_5 + \kappa_1\left(\kappa_2\kappa_{10} + \bar{W}_1\kappa_9 + \bar{W}_2\kappa_{10}\right)\kappa_6\right) \\ & + \Gamma_{21a}\left(\kappa_2^2\kappa_4\kappa_6 + \kappa_2\left(\kappa_1\kappa_{11} + \bar{W}_1\kappa_{11} + \bar{W}_2\kappa_{12}\right)\kappa_5\right) \\ & + \Phi_3\left\|\tilde{W}_{1c}\right\|\left\|\tilde{W}_{2c}\right\| + \kappa_7 + \kappa_8,\end{aligned}$$

where

$$\begin{aligned}\Phi_1 &\triangleq \left(\frac{c_{41}\eta_{1c}\varphi_{01}}{2\sqrt{\nu_1\varphi_{11}}}\kappa_5 + \Gamma_{11a}\left(\kappa_1\kappa_3\kappa_5 + \kappa_1^2\kappa_3 + \kappa_1\kappa_9\kappa_6\right)\right. \\ &\quad \left. + \Gamma_{12a}\kappa_1 + \Gamma_{21a}\kappa_2\left(\kappa_1\kappa_{11} + \bar{W}_1\kappa_{11} + \bar{W}_2\kappa_{12}\right)\right), \\ \Phi_2 &\triangleq \left(\frac{c_{42}\eta_{2c}\varphi_{02}}{2\sqrt{\nu_2\varphi_{12}}}\kappa_6 + \Gamma_{21a}\left(\kappa_2\kappa_3\kappa_6 + \kappa_2\kappa_{12}\kappa_5 + \kappa_2^2\kappa_4\right)\right. \\ &\quad \left. + \Gamma_{22a}\kappa_2 + \Gamma_{11a}\kappa_1\left(\kappa_2\kappa_{10} + \bar{W}_1\kappa_9 + \bar{W}_2\kappa_{10}\right)\right), \\ \Phi_3 &\triangleq \left(\Gamma_{11a}\kappa_1\kappa_9 + \Gamma_{21a}\kappa_2\kappa_{12}\right).\end{aligned}$$

Provided $c_{31} > \Gamma_{11a}\kappa_1\kappa_3$ and $c_{32} > \Gamma_{21a}\kappa_2\kappa_4$, using Young's inequality $\left\|\tilde{W}_{1c}\right\|\left\|\tilde{W}_{2c}\right\| \leq \frac{1}{2}\left\|\tilde{W}_{1c}\right\|^2 + \frac{1}{2}\left\|\tilde{W}_{2c}\right\|^2$, and completing the square yields

$$\begin{aligned}\dot{V}_L \leq & -Q - (1-\theta_1)(c_{31} - \Gamma_{11a}\kappa_1\kappa_3 - \frac{1}{2}\Phi_3)\left\|\tilde{W}_{1c}\right\|^2 - \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 \\ & - (1-\theta_2)(c_{32} - \Gamma_{21a}\kappa_2\kappa_4 - \frac{1}{2}\Phi_3)\left\|\tilde{W}_{2c}\right\|^2 - \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|^2 \\ & + \Gamma_{11a}\left(\kappa_1^2\kappa_3\kappa_5 + \kappa_1\left(\kappa_2\kappa_{10} + \bar{W}_1\kappa_9 + \bar{W}_2\kappa_{10}\right)\kappa_6\right) \\ & + \Gamma_{21a}\left(\kappa_2^2\kappa_4\kappa_6 + \kappa_2\left(\kappa_1\kappa_{11} + \bar{W}_1\kappa_{11} + \bar{W}_2\kappa_{12}\right)\kappa_5\right) \\ & + \frac{\Phi_1^2}{4\theta_1(c_{31} - \Gamma_{11a}\kappa_1\kappa_3 - \frac{1}{2}\Phi_3)} + \frac{\Phi_2^2}{4\theta_2(c_{32} - \Gamma_{21a}\kappa_2\kappa_4 - \frac{1}{2}\Phi_3)},\end{aligned} \tag{5–39}$$

where $\theta_1, \theta_2 \in (0,1)$. Since $Q(x)$ is positive definite, according to Lemma 4.3 in [131], there exist class $\mathcal{K}$ functions α_5 and α_6 such that

$$\alpha_5(\|w\|) \leq F\left(\|w\|\right) \leq \alpha_6(\|w\|) \qquad \forall w \in B_s, \tag{5–40}$$

where

$$F\left(\|w\|\right) = Q + (1-\theta_1)(c_{31} - \Gamma_{11a}\kappa_1\kappa_3 - \frac{1}{2}\Phi_3)\left\|\tilde{W}_{1c}\right\|^2 + \Gamma_{12a}\left\|\tilde{W}_{1a}\right\|^2 + (1-\theta_2)(c_{32} - \Gamma_{21a}\kappa_2\kappa_4 - \frac{1}{2}\Phi_3)\left\|\tilde{W}_{2c}\right\|^2 + \Gamma_{22a}\left\|\tilde{W}_{2a}\right\|^2.$$

Using Eq. 5–40, the expression in Eq. 5–39 can be further upper bounded as

$$\dot{V}_L \leq -\alpha_5(\|w\|) + \Upsilon,$$

where

$$\Upsilon = \Gamma_{11a}\left(\kappa_1^2\kappa_3\kappa_5 + \kappa_1\left(\kappa_2\kappa_{10} + \bar{W}_1\kappa_9 + \bar{W}_2\kappa_{10}\right)\kappa_6\right) + \Gamma_{21a}\left(\kappa_2^2\kappa_4\kappa_6 + \kappa_2\left(\kappa_1\kappa_{11} + \bar{W}_1\kappa_{11} + \bar{W}_2\kappa_{12}\right)\kappa_5\right) + \frac{\Phi_1^2}{4\theta_1(c_{31} - \Gamma_{11a}\kappa_1\kappa_3 - \frac{1}{2}\Phi_3)} + \frac{\Phi_2^2}{4\theta_2(c_{32} - \Gamma_{21a}\kappa_2\kappa_4 - \frac{1}{2}\Phi_3)},$$

which proves that $\dot{V}_L(\cdot)$ is negative whenever $w(t)$ lies outside the compact set $\Omega_w \triangleq \left\{w : \|w\| \leq \alpha_5^{-1}\left(\Upsilon\right)\right\}$, and hence, $\|w(t)\|$ is UUB, according to Theorem 4.18 in [131]. □

5.6 Convergence to Nash Solution

The subsequent theorem demonstrates that the actor NN approximations converge to the approximate coupled HJB in Eq. 5–10. It can also be shown that the approximate controllers in Eq. 5–19 approximate the optimal solutions to the two player Nash game for the dynamic system given in Eq. 5–16.

Assumption 5.2. *For each admissible control policies the HJB equations Eq. 5–10 have a locally smooth solution $V_i(x) \geq 0$, for $i = 1, 2$, and $f(\cdot)$ is Lipschitz and bounded by $\|f\| \leq c_f\|x\|$, where $c_f \in \mathbb{R}$ is a positive constant.*

Theorem 5.3. *Given that the Assumptions and sufficient gain constraints in Theorem 5-2 hold, then the actor and critic NNs converge to the approximate coupled HJB solution, in the sense that the HJBs in Eq. 5–13 are UUB.*

Proof. Consider the approximate HJB in Eq. 5–13 and after substituting the approximate control laws in Eq. 5–19 yields

$$\begin{aligned} H_1\left(x, \nabla\hat{V}_1, \hat{u}_1, \hat{u}_2\right) &= r_{\hat{u}_1} + \nabla\hat{V}_1 F_{\hat{u}} \\ &= Q_1(x) + \frac{1}{4}\hat{W}_{1a}^T\phi_1' G_1 \phi_1'^T \hat{W}_{1a} + \hat{W}_{1c}^T\phi_1' f(x) \\ &\quad + \frac{1}{4}\hat{W}_{2a}^T\phi_2' G_{12}\phi_2'^T\hat{W}_{2a} \\ &\quad - \frac{1}{2}\hat{W}_{1c}^T\phi_1'\left(G_1\phi_1'^T\hat{W}_{1a} + G_2\phi_2'^T\hat{W}_{2a}\right), \end{aligned}$$

and

$$\begin{aligned} H_2\left(x, \nabla\hat{V}_2, \hat{u}_1, \hat{u}_2\right) &= r_{\hat{u}_2} + \nabla\hat{V}_2 F_{\hat{u}} \\ &= Q_2(x) + \frac{1}{4}\hat{W}_{2a}^T\phi_2' G_2 \phi_2'^T \hat{W}_{2a} + \hat{W}_{1c}^T\phi_1' f(x) \\ &\quad + \frac{1}{4}\hat{W}_{1a}^T\phi_1' G_{21}\phi_1'^T\hat{W}_{1a} \\ &\quad - \frac{1}{2}\hat{W}_{2c}^T\phi_2'\left(G_1\phi_1'^T\hat{W}_{1a} + G_2\phi_2'^T\hat{W}_{2a}\right). \end{aligned}$$

After adding and subtracting $\left(W_i^T\phi_i' + \varepsilon_i'\right) f = -\left(W_i^T\phi_i' + \varepsilon_i'\right)(g_1u_1^* + g_2u_2^*) - Q_i(x) - \sum_{j=1}^{2} u_j^T R_{ij} u_j$ for $i = 1, 2$ and substituting for the optimal control law in Eq. 5–18 as

$$\begin{aligned} H_1 &= -\tilde{W}_{1c}^T\phi_1' f(x) - \varepsilon_1' f + \frac{1}{4}\hat{W}_{1a}^T\phi_1' G_1\phi_1'^T\hat{W}_{1a} - \frac{1}{4}W_1^T\phi_1' G_1\phi_1'^T W_1 \\ &\quad + \frac{1}{4}\hat{W}_{2a}^T\phi_2' G_{12}\phi_2'^T\hat{W}_{2a} - \frac{1}{4}W_2^T\phi_2' G_{12}\phi_2'^T W_2 \\ &\quad - \frac{1}{2}\left(\varepsilon_2' G_{12} - \varepsilon_1' G_2\right)\phi_2'^T W_2 + \frac{1}{2}\varepsilon_1'\left(G_1\varepsilon_1'^T + G_2\varepsilon_2'^T\right) - \frac{1}{4}\varepsilon_2' G_{12}\varepsilon_2'^T \\ &\quad + \frac{1}{2}W_1^T\phi_1'\left(G_1\phi_1'^T W_1 + G_2\phi_2'^T W_2\right) + \frac{1}{2}W_1^T\phi_1'\left(G_1\varepsilon_1'^T + G_2\varepsilon_2'^T\right) \\ &\quad - \frac{1}{2}\hat{W}_{1c}^T\phi_1'\left(G_1\phi_1'^T\hat{W}_{1a} + G_2\phi_2'^T\hat{W}_{2a}\right), \end{aligned} \tag{5–41}$$

and

$$
\begin{aligned}
H_2 \;=\; & -\tilde{W}_{2c}^T\phi_2' f\left(x\right) - \varepsilon_2' f + \frac{1}{4}\hat{W}_{2a}^T\phi_2' G_2 \phi_2'^T \hat{W}_{2a} - \frac{1}{4} W_2^T \phi_2' G_2 \phi_2'^T W_2 \\
& + \frac{1}{4}\hat{W}_{1a}^T\phi_1' G_{21} \phi_1'^T \hat{W}_{1a} - \frac{1}{4} W_1^T \phi_1' G_{21} \phi_1'^T W_1 \\
& - \frac{1}{2}\left(\varepsilon_1' G_{21} - \varepsilon_2' G_1\right)\phi_1'^T W_1 + \frac{1}{2}\varepsilon_2'\left(G_2 \varepsilon_2'^T + G_1 \varepsilon_1'^T\right) - \frac{1}{4}\varepsilon_1' G_{21} \varepsilon_1'^T \\
& + \frac{1}{2} W_2^T \phi_2'\left(G_2 \phi_2'^T W_2 + G_1 \phi_1'^T W_1\right) + \frac{1}{2} W_2^T \phi_2'\left(G_2 \varepsilon_2'^T + G_1 \varepsilon_1'^T\right) \\
& - \frac{1}{2}\hat{W}_{2c}^T \phi_2'\left(G_2 \phi_2'^T \hat{W}_{2a} + G_1 \phi_1'^T \hat{W}_{1a}\right).
\end{aligned} \tag{5–42}
$$

Substituting the NN mismatch errors $\tilde{W}_{ic}(t) \triangleq W_i - \hat{W}_{ic}(t)$ and $\tilde{W}_{ia}(t) \triangleq W_i - \hat{W}_{ia}(t)$, for $i = 1, 2$ into 4–65 and 5–42, respectively, yields

$$
\begin{aligned}
H_1 \;=\; & -\tilde{W}_{1c}^T\phi_1' f\left(x\right) - \varepsilon_1' f - \frac{1}{4}\tilde{W}_1^T \phi_1' G_1 \phi_1'^T \tilde{W}_1 \\
& - \frac{1}{2}\tilde{W}_{2a}^T \phi_2' G_{12} \phi_2'^T W_{2a} + \frac{1}{2}\tilde{W}_2^T \phi_2' G_{12} \phi_2'^T \tilde{W}_2 \\
& - \frac{1}{2}\left(\varepsilon_2' G_{12} - \varepsilon_1' G_2\right)\phi_2'^T W_2 + \frac{1}{2}\varepsilon_1'\left(G_1 \varepsilon_1'^T + G_2 \varepsilon_2'^T\right) - \frac{1}{4}\varepsilon_2' G_{12} \varepsilon_2'^T \\
& + \frac{1}{2} W_1^T \phi_1'\left(G_1 \phi_1'^T\left(2\tilde{W}_{1a} - W_1\right) + G_2 \phi_2'^T \tilde{W}_2\right) + \frac{1}{2} W_1^T \phi_1'\left(G_1 \varepsilon_1'^T + G_2 \varepsilon_2'^T\right) \\
& + \frac{1}{2}\tilde{W}_1^T \phi_1'\left(G_1 \phi_1'^T\left(-\tilde{W}_{1a} + W_1\right) + G_2 \phi_2'^T\left(-\tilde{W}_{2a} + W_2\right)\right),
\end{aligned} \tag{5–43}
$$

and

$$
\begin{aligned}
H_2 \;=\; & -\tilde{W}_{2c}^T\phi_2' f\left(x\right) - \varepsilon_2' f - \frac{1}{4}\tilde{W}_2^T \phi_2' G_2 \phi_2'^T \tilde{W}_2 \\
& - \frac{1}{2}\tilde{W}_{2a}^T \phi_1' G_{21} \phi_1'^T W_{1a} + \frac{1}{2}\tilde{W}_1^T \phi_1' G_{21} \phi_1'^T \tilde{W}_1 \\
& - \frac{1}{2}\left(\varepsilon_1' G_{21} - \varepsilon_2' G_1\right)\phi_1'^T W_1 + \frac{1}{2}\varepsilon_2'\left(G_2 \varepsilon_2'^T + G_1 \varepsilon_1'^T\right) - \frac{1}{4}\varepsilon_1' G_{21} \varepsilon_1'^T \\
& + \frac{1}{2} W_2^T \phi_2'\left(G_2 \phi_2'^T\left(2\tilde{W}_{2a} - W_2\right) + G_1 \phi_1'^T \tilde{W}_1\right) + \frac{1}{2} W_2^T \phi_2'\left(G_2 \varepsilon_2'^T + G_1 \varepsilon_1'^T\right) \\
& + \frac{1}{2}\tilde{W}_2^T \phi_2'\left(G_2 \phi_2'^T\left(-\tilde{W}_{2a} + W_2\right) + G_1 \phi_1'^T\left(-\tilde{W}_{1a} + W_1\right)\right).
\end{aligned} \tag{5–44}
$$

It is easy to see that if the assumptions and sufficient gain constraints in Theorem 5-2 hold, then the right side of Eqs. 5–43 and 5–44 can be upper bounded by a function that

is UUB $\|H_i\| \leq \Theta_i \left(\tilde{W}_{ic}, \tilde{W}_{1a}, \tilde{W}_{2a}, t \right)$ for $i = 1, 2$, therefore the approximate HJBs are also UUB. □

Theorem 5.4. *Given that the assumptions and sufficient gain constraints in Theorem 2 hold, the approximate control laws in Eq. 4–41 converge to the approximate Nash solution of the game.*

Proof. Consider the control errors $(\tilde{u}_1, \tilde{u}_2)$ between the optimal control laws in Eq. 5–9 and the approximate control laws in Eq. 5–19 given as

$$\tilde{u}_1 \triangleq u_1^* - \hat{u}_1, \qquad \tilde{u}_2 \triangleq u_2^* - \hat{u}_2.$$

Substituting for the optimal control laws in Eq. 5–9 and the approximate control laws in Eq. 5–19 and using $\tilde{W}_{ia}(t) \triangleq W_i - \hat{W}_{ia}(t)$ for $i = 1, 2$, yields

$$\begin{aligned} \tilde{u}_1 &= -\frac{1}{2} R_{11}^{-1} g_1^T(x) \phi_1'(x) \left(\tilde{W}_{1a} + \varepsilon_1'(x)^T \right), \\ \tilde{u}_2 &= -\frac{1}{2} R_{22}^{-1} g_2^T(x) \phi_2'(x) \left(\tilde{W}_{2a} + \varepsilon_2'(x)^T \right). \end{aligned} \tag{5–45}$$

Using Assumptions 2-5, Eq. 5–45 can be upper bounded as

$$\begin{aligned} \|\tilde{u}_1\| &\leq \frac{1}{2} \lambda_{min} \left(R_{11}^{-1} \right) \bar{g}_1 \bar{\phi}_1' \left(\left\| \tilde{W}_{1a} \right\| + \bar{\varepsilon}'_1 \right), \\ \|\tilde{u}_2\| &\leq \frac{1}{2} \lambda_{min} \left(R_{22}^{-1} \right) \bar{g}_2 \bar{\phi}_2' \left(\left\| \tilde{W}_{2a} \right\| + \bar{\varepsilon}'_2 \right). \end{aligned}$$

Given that the assumptions and sufficient gain constraints in Theorem 5-2 hold, then all terms to the right of the inequality are UUB, therefore the control errors $(\tilde{u}_1, \tilde{u}_2)$ are UUB and the approximate control laws $(\hat{u}_1, \hat{u}_2)$ give the approximate Nash equilibrium solution. □

5.7 Simulation

The following nonlinear dynamics are considered in [101, 108, 109, 132]

$$\dot{x} = f(x) + g_1(x) u_1(x) + g_2(x) u_2(x),$$

where

$$f(x) = \begin{bmatrix} x_2 \\ -\frac{1}{2}x_1 - x_2 + \frac{1}{4}x_2 \left(cos\left(2x_1\right) + 2\right)^2 - \frac{1}{4}x_2 \left(sin\left(4x_1^2\right) + 2\right)^2 \end{bmatrix}$$

$$g_1(x) = \begin{bmatrix} 0 & cos\left(2x_1\right) + 2 \end{bmatrix}^T \qquad g_2(x) = \begin{bmatrix} 0 & sin\left(4x_1^2\right) + 2 \end{bmatrix}^T.$$

The initial state is given as $x(0) = [3, -1]^T$ and the local cost function is defined as

$$r_i = x^T Q_i x + u_i^T R_{ii} u_i + u_i^T R_{ji} u_i \qquad i = 1, 2,\ j = 3 - i,$$

where

$$R_{11} = 2R_{22} = 1, \qquad R_{12} = 2R_{21} = 2, \qquad Q_1 = 2Q_2 = \mathbb{I}_{2\times 2}.$$

The optimal value functions for the critics of player 1 and player 2 are given as

$$V_1^*(x) = \frac{1}{2}x_1^2 + x_2^2, \qquad V_2^*(x) = \frac{1}{4}x_1^2 + \frac{1}{2}x_2^2,$$

and the optimal control inputs are given as

$$u_1^* = -\left(cos\left(2x_1\right) + 2\right)x_2, \qquad u_2^* = -\left(sin\left(4x_1^2\right) + 2\right)x_2.$$

The activation function for the critic NN is chosen as

$$\phi = \begin{bmatrix} x_1^2 & x_1x_2 & x_2^2 \\ & & \end{bmatrix},$$

while the activation function for the identifier DNN is chosen as a symmetric sigmoid with 5 neurons in the hidden layer. The identifier gains are chosen as

$$k = 800, \qquad \alpha = 300, \qquad \gamma_f = 5, \qquad \beta_1 = 0.2, \qquad \Gamma_{wf} = 0.1\mathbb{I}_{6\times 6}, \qquad \Gamma_{vf} = 0.1\mathbb{I}_{2\times 2},$$

and the gains of the actor-critic learning laws are chosen as

$$\Gamma_{11a} = \Gamma_{22a} = 10, \qquad \Gamma_{12a} = \Gamma_{21a} = 50, \qquad \eta_{1c} = \eta_{1c} = 50, \qquad \nu_1 = \nu_2 = 0.001.$$

The covariance matrix is initialized to $\Gamma(0) = 5000$, all the NN weights are randomly initialized with values between $[-1, 1]$, and the states are initialized to $x(0) = [3, -1]$. A small amplitude exploratory signal (noise) is added to the control to excite the states for the first 3 seconds of the simulation, as seen from the evolution of states in Figure 5-1. The identifier approximates the system dynamics, and the state derivative estimation error is shown in Figure 5-2. The time histories of the critic NN weights and the actors NN weights are given in Figure 5-3 and 5-4. Persistence of excitation ensures that the weights converge. Figure 5-5 shows the optimal value functions and the approximate ones. Figure 5-6 shows the optimal controller and the approximated controller for player 1 and 2, respectively. Figures 5-7, 5-8, 5-9, and 5-10 demonstrate that for a PE signal that is not removed the weights converge, however the PE signal degrades the performance of the states.

Remark 5.1. An implementation issue in using the developed algorithm is to ensure PE of the critic regressor vector. Unlike linear systems, where PE of the regressor translates to the sufficient richness of the external input, no verifiable method exists to ensure PE in nonlinear systems. In this simulation, a small amplitude exploratory signal consisting of a sum of sines and cosines of varying frequencies is added to the control to ensure PE qualitatively, and convergence of critic weights to their optimal values is achieved. The exploratory signal $n(t)$ is present in the first 3 seconds of the simulation and is given by

$$n(t) = (1.2 - \exp(.01t)) \left(cos^2(0.2t) + sin^2(2.0t)\, cos(0.1t) + sin^2(-1.2t)\, cos(.5t) + sin^5(t)\right).$$

5.8 Summary

A generalized solution for a N-player nonzero-sum differential game is sought utilizing by a Hamilton-Jacobi-Bellman approximation by an actor-critic-identifier architecture. The ACI architecture implements the actor and critic approximation simultaneously and in real-time. The use of a robust DNN-based identifier circumvents the need for complete model knowledge, yielding an identifier which is proven to be asymptotically convergent.

A gradient-based weight update law is used for the critic NN to approximate the value function. Using the identifier and the critic, an approximation to the optimal control law (actor) is developed which stabilizes the closed loop system and approaches the optimal solutions to the N-player nonzero-sum game.

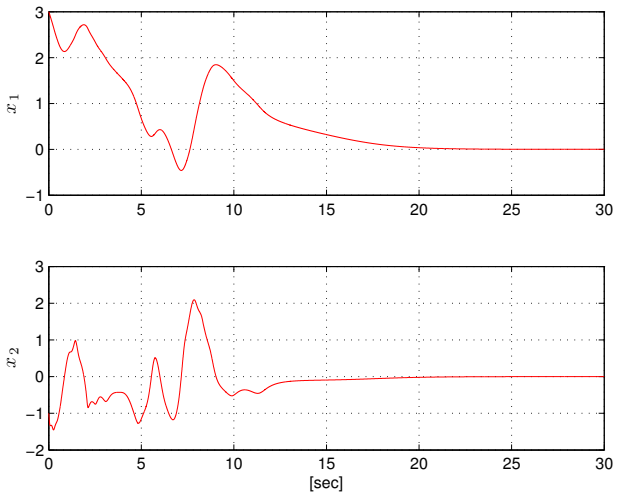

Figure 5-1. The evolution of the system states for the nonzero-sum game, with persistently excited input for the first 10 seconds.

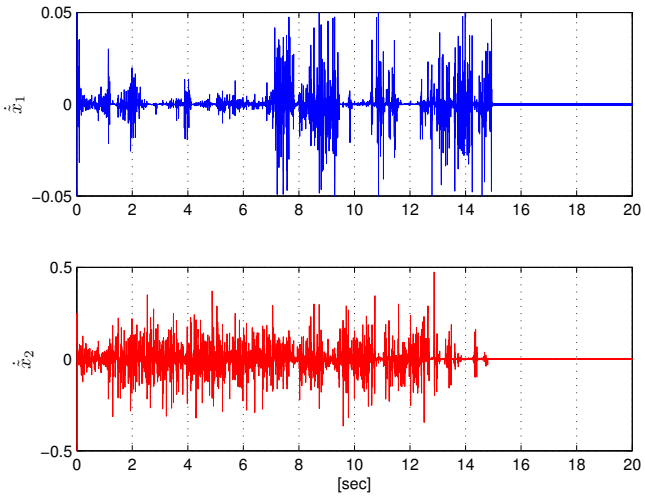

Figure 5-2. Error in estimating the state derivatives, with the identifier for the nonzero-sum game.

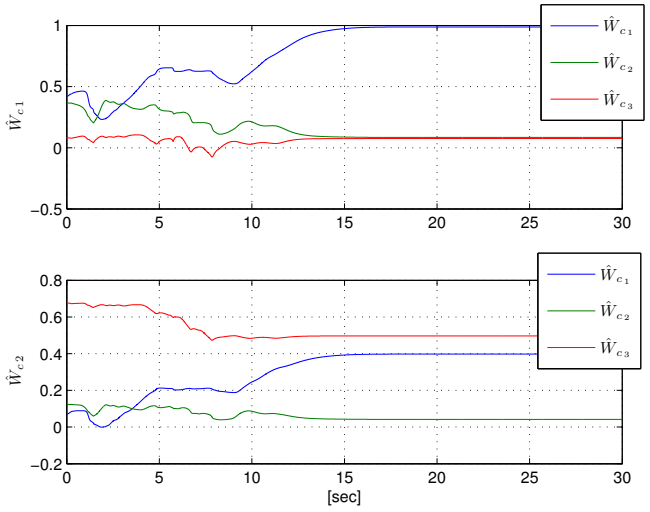

Figure 5-3. Convergence of critic weights for the nonzero-sum game.

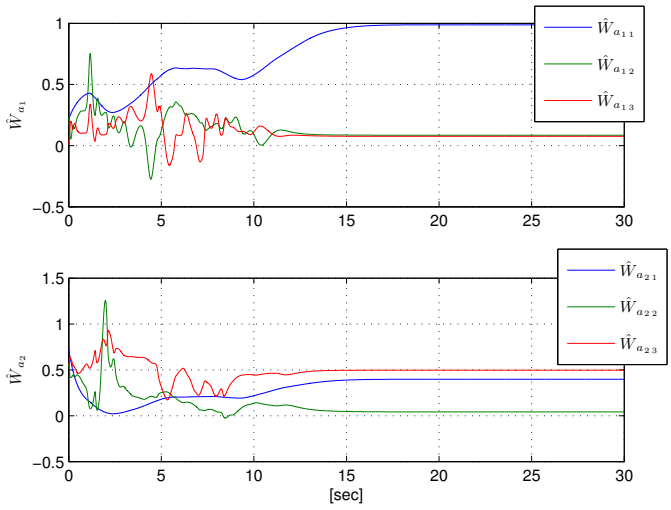

Figure 5-4. Convergence of actor weights for player 1 and player 2 in a nonzero-sum game.

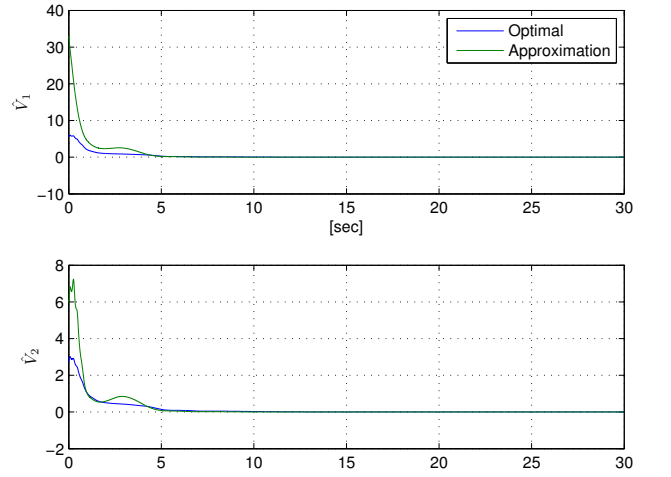

Figure 5-5. Value function approximation $\hat{V}(x)$, for a nonzero-sum game.

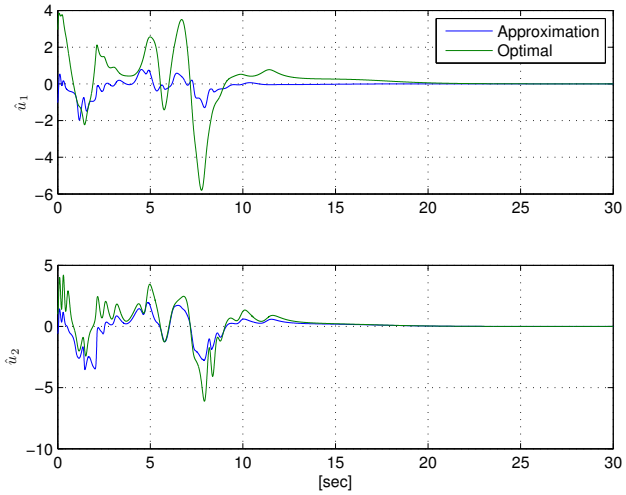

Figure 5-6. Optimal control approximation $\hat{u}$, for a nonzero-sum game.

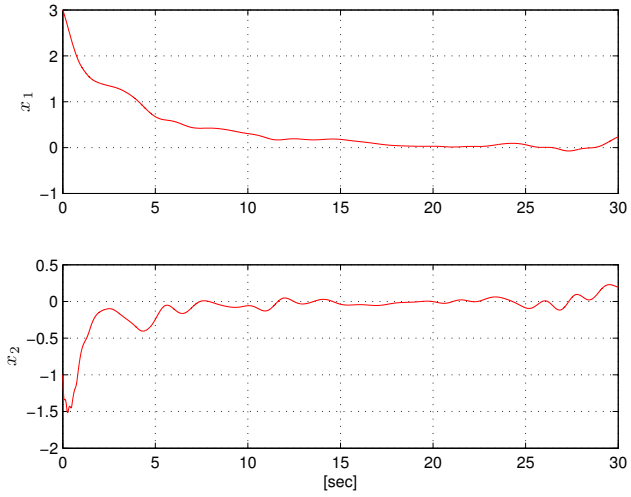

Figure 5-7. The evolution of the system states for the nonzero-sum game, with a continuous persistently excited input.

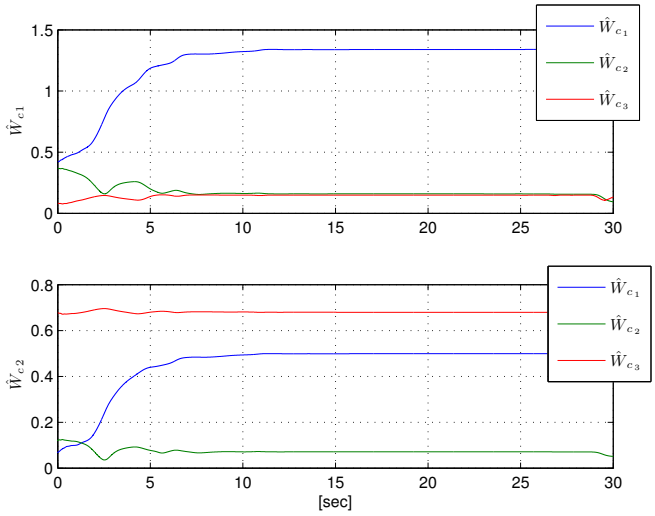

Figure 5-8. Convergence of critic weights for the nonzero-sum game, with a continuous persistently excited input.

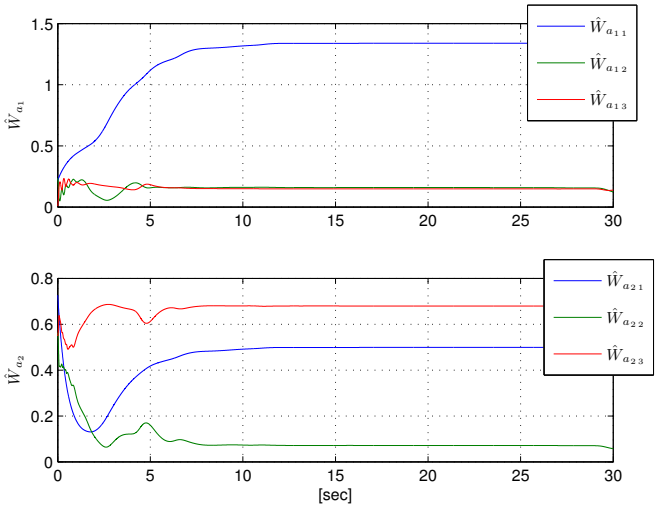

Figure 5-9. Convergence of actor weights for player 1 and player 2 in a nonzero-sum game, with a continuous persistently excited input.

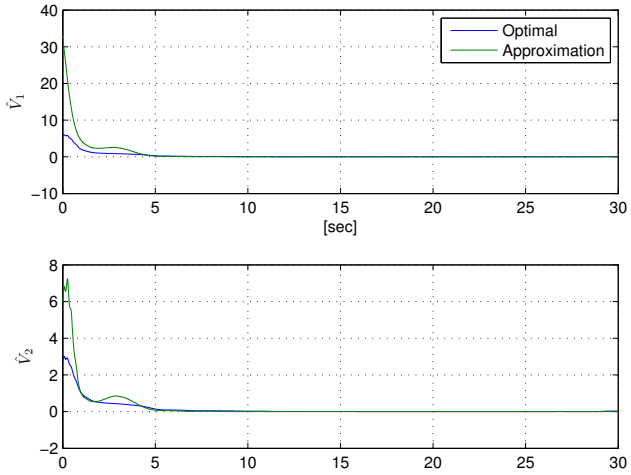

Figure 5-10. Value function approximation $\hat{V}(x)$ for a nonzero-sum game, with a continuous persistently excited input.

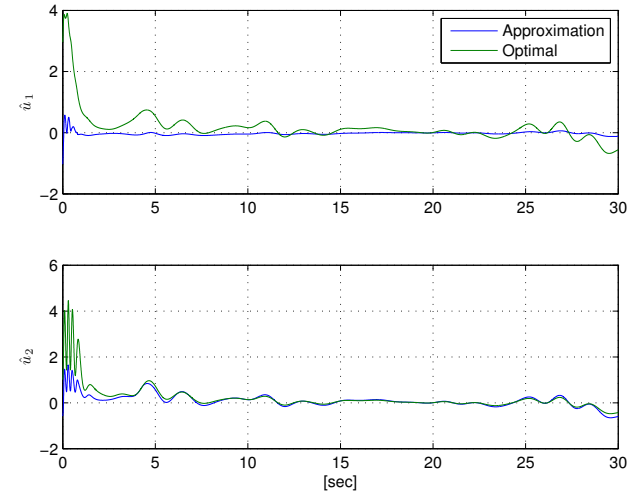

Figure 5-11. Optimal control approximation $\hat{u}$ for a nonzero-sum game, with a continuous persistently excited input.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Conclusion

The focus of this work is to develop techniques for approximating solutions to zero-sum and nonzero-sum noncooperative differential games and using these solutions to stabilize some classes of uncertain nonlinear systems. In the spirit of optimal control, two approaches were used based from Bellman's optimality principle and Pontryagin's maximum principle to approximate the solution to coupled nonlinear HJB equations. The first approach, using the maximum principle, involves partial feedback linearization of a particular class of nonlinear systems and synthesizing a differential game. The differential game yields a coupled set of DRE equations which are reduced to ARE and conditions are given for the solution to the AREs. The second approach uses the optimality principle, particularly the dynamic programming approach, to approximate the solution to the HJB. These approaches are shown to approximately solve a differential game and stabilize the dynamics. The specific contributions of each result are mentioned below.

Chapter 2 focuses on the development of robust (sub)optimal feedback Nash-based feedback control laws for an uncertain nonlinear system. This chapter incorporates the RISE controller with an optimal Nash strategy to stabilize an uncertain Euler-Lagrange system with additive disturbances. This chapter also illustrates the development of the RISE controller which is used to asymptotically identify the nonlinearities in the dynamics. By applying the RISE controller the nonlinear dynamics converge to a residual system, the solution to the feedback Nash game for the residual system is used to derive the stabilizing feedback control laws. The (sub)optimal feedback controllers are shown to minimize a cost functional in the presence of unknown bounded disturbances, and a Lyapunov-based stability analysis demonstrates asymptotic tracking for the combination of the RISE and Nash-based controllers.

The result from Chapter 2 is further refined in Chapter 3 for a class of systems in which additional information is provided to one of the players. The main contribution of this chapter is the development of robust (sub)optimal open-loop Stackelberg-based for the leader and follower, which both act as inputs to an uncertain nonlinear system. Similar to Chapter 2, this chapter utilizes the RISE controller and combines it with a differential game-based control strategy. The control formulation utilizes the solution to the hierarchical open-loop Stackelberg game to derive the feedback control laws. A Lyapunov-based asymptotic tracking derivation and a simulation is presented to validate the utility of the technique.

In contrast to the approaches in Chapters 2 and 3, which are largely based on Pontryagin's maximum principle, the techniques in Chapter 4 and 5 attempt to approximate the solution to the HJI. The main contribution of Chapter 4 is solving a two player zero-sum infinite horizon game subject to continuous-time unknown nonlinear dynamic that are affine in the input. In the developed method, two actor and one critic NNs using gradient and least squares-based update laws, respectively, are designed to minimize the Bellman error, which is the difference between the exact and the approximate HJI equation. The identifier DNN is a combination of a Hopfield-type [112] component, in parallel configuration with the system [113], and a RISE component. The Hopfield component of the DNN learns the system dynamics based on online gradient-based weight tuning laws, while the RISE term robustly accounts for the function reconstruction errors, guaranteeing asymptotic estimation of the state and the state derivative. The online estimation of the state derivative allows the ACI architecture to be implemented without knowledge of system drift dynamics; however, knowledge of the input gain matrix is required to implement the control policy. While the design of the actor and critic are coupled through a HJI equation, the design of the identifier is decoupled from actor-critic, and can be considered as a modular component in the actor-critic-identifier architecture. Convergence of the actor-critic-identifier-based algorithm and stability of the closed-loop system are analyzed

using Lyapunov-based adaptive control methods, and a PE condition is used to guarantee convergence to within a bounded region of the optimal control and UUB stability of the closed-loop system.

Nonzero-sum games pose different challenges as compared to zero-sum games. For nonlinear dynamics, the HJI for zero-sum games is equivalently a coupled set of nonlinear HJB equations for nonzero-sum games. Chapter 5 builds Chapter 4, by considering a N-player nonzero-sum infinite horizon game subject to continuous-time uncertain nonlinear dynamics. The main contribution of this work is deriving an approximate solution to a N-player nonzero-sum game with a technique that is continuous, online and based on adaptive control theory. Previous research in the area focused on simplistic scalar nonlinear systems or implemented iterative/hybrid techniques that required complete knowledge of the drift dynamics. The technique expands the ACI structure to solve a differential game problem, wherein two actor and two critic neural network structures are used to approximate the optimal control laws and the optimal value function set, respectively. The main traits of this online algorithm involve the use of ADP techniques and adaptive theory to determine the Nash equilibrium solution of the game in an online simultaneous procedure that does not require full knowledge of the system dynamics and the online version of a mathematical algorithm that solves the underlying set of coupled HJB equations of the game problem. For an equivalent nonlinear system, previous research makes use of offline procedures or requires full knowledge of the system dynamics to determine the Nash equilibrium. A Lyapunov proof shows that UUB tracking for the closed-loop system is guaranteed and a convergence analysis demonstrates that the approximate control policies converge to the optimal solutions in the sense of UUB.

6.2 Future Work

The work in this dissertation opens new doors for the research in the domain of nonlinear optimal control design. In this section, open problems related to the work in this

dissertation are discussed for a curious reader. The open problems are listed below. From Chapters 2 and 3:

1. The investigation of optimal output feedback solutions for nonzero-sum games subject to nonlinear dynamics, where full state feedback is not available. In game theory this scenario is considered an imperfect state game. Many engineering problems that can be defined by Euler-Lagrange dynamics do not always have full state feedback and thus output feedback designs are desirable for implementation. Nonlinear H_∞ control has examined an output feedback solution for a zero-sum game, however these controllers require the solution to the HJI equation. Furthermore the nonzero-sum output feedback design has been relatively unexplored.
2. The determination of an differential game-derived optimal control control law that is subject to saturated inputs and time delays. Hardware implementation of control designs are often plagued with time delays and small actuator bandwidth. Preliminary investigation with Heuristic Dynamic Programming have looked at ADP approaches to incorporating these phenomenon into the control design, however little research has gone into the development of existence and uniqueness conditions for these types of games and the implementation of a game-derived control design for a nonlinear system that incorporates these effects.
3. In regards to Euler-Lagrange dynamics, the derivation of a greedy optimal control that that has a moving horizon, thereby allowing for online realization. A greedy strategy only looks at the most optimal choice for the next iteration, whereas Chapters 2 and 3 focused on the infinite horizon optimal strategy. Defining a controller that looks at a smaller interval that changes with time allows for a computationally feasible online calculation of the game solution.

From Chapters 4 and 5:

1. Online adaptive optimal controllers for systems with periodic dynamics. Periodic dynamics can be seen as systems whom yield a periodic output (e.g. At various

level of modeling, automotive engine dynamics can be considered as a linear periodic system mechanically coordinated through the revolution of the crankshaft). Periodic systems are present in a wide variety of engineering applications, particularly robotic systems used in manufacturing, yet there have not been much investigation of controlling these systems using ADP techniques.

2. Online adaptive optimal controllers using output feedback. Nonlinear H_∞ control has examined an output feedback solution for a zero-sum game, however these controllers require the solution to the HJI equation. Furthermore the nonzero-sum output feedback design has been relatively unexplored. Using ADP techniques, the approximation of the HJI solution could yield more implementable solutions.
3. Existence and uniqueness proofs for multi-player nonzero-sum infinite horizon games with nonlinear dynamic constraints. Due to the complexity of N-coupled HJB equations, the investigation of uniqueness properties in the N-player nonzero-sum game are sparse. For nonlinear dynamics this is largely still seen as an open problem.
4. An online continuous ADP solution using a mixed strategy for a zero-sum infinite horizon game with nonlinear dynamic constraints. Zero-sum games are widely used in engineering problems, however, typically the solution of the HJI equation is assumed to exist or has local existence with conditions that are difficult to satisfy. For online continuous ADP games, the scenario in which the saddle point solution does not exist is an open problem. The use of the mixed strategy incorporated in an online continuous ADP technique could be a feasible solution.

REFERENCES

[1] A. Barto, R. Sutton, and C. Anderson, "Neuron-like adaptive elements that can solve difficult learning control problems," *IEEE Trans. Syst. Man Cybern.*, vol. 13, no. 5, pp. 834–846, 1983.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[3] R. Sutton, A. Barto, and R. Williams, "Reinforcement learning is direct adaptive optimal control," *IEEE Contr. Syst. Mag.*, vol. 12, no. 2, pp. 19–22, 1992.

[4] J. Campos and F. Lewis, "Adaptive critic neural network for feedforward compensation," in *Proc. Am. Control Conf.*, vol. 4, 1999.

[5] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Adaptive critic designs for discrete-time zero-sum games with application to h-[infinity] control," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 37, pp. 240–247, 2007.

[6] Y. Tessa and T. Erez, "Least squares solutions of the hjb equation with neural network value-function approximators," *Trans. on Neural Networks*, vol. 18, pp. 1031–1041, 2007.

[7] R. Beard, G. Saridis, and J. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, pp. 2159–2178, 1997.

[8] J. A. Primbs and V. Nevistic, "Optimality of nonlinear design techniques: A converse HJB approach," California Institute of Technology, Pasadena, CA 91125, Tech. Rep. CIT-CDS 96-022, 1996.

[9] T. Cheng, F. Lewis, and M. Abu-Khalaf, "Fixed-final-time-constrained optimal control of nonlinear systems using neural network HJB approach," *IEEE Trans. Neural Networks*, vol. 18, no. 6, pp. 1725–1737, 2007.

[10] M. Abu-Khalaf and F. Lewis, "Nearly optimal HJB solution for constrained input systems using a neural network least-squares approach," in *Proc. IEEE Conf. Decis. Control*, Las Vegas, NV, 2002, pp. 943–948.

[11] M. Krstic and Z.-H. Li, "Inverse optimal design of input-to-state stabilizing nonlinear controllers," in *Proc. IEEE Conf. Decis. Control*, vol. 4, Dec. 10–12, 1997, pp. 3479–3484.

[12] M. Krstic and P. Tsiotras, "Inverse optimality results for the attitude motion of a rigid spacecraft," in *Proc. Am. Control Conf.*, 4-6 June 1997, pp. 1884–1888.

[13] M. Krstic and H. Deng, *Stabilization of Nonlinear Uncertain Systems*. Springer, 1998.

[14] M. Krstic and Z.-H. Li, "Inverse optimal design of input-to-state stabilizing nonlinear controllers," *IEEE Trans. Autom. Control*, vol. 43, no. 3, pp. 336–350, March 1998.

[15] N. Kidane, Y. Yamashita, H. Nakamura, and H. Nishitani, "Inverse optimization for a nonlinear system with an input constraint," in *Proc. SICE Annu. Conf.*, vol. 2, 4-6 Aug. 2004, pp. 1210–1213.

[16] T. Fukao, "Inverse optimal tracking control of a nonholonomic mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, vol. 2, 28 Sept.-2 Oct. 2004, pp. 1475–1480.

[17] R. Freeman and J. Primbs, "Control Lyapunov functions: new ideas from an old source," in *Proc. IEEE Conf. Decis. Control*, 11-13 Dec. 1996, pp. 3926–3931.

[18] P. Gurfil, "Non-linear missile guidance synthesis using control Lyapunov functions," *Proc. IME G J. Aero. Eng.*, vol. 219, pp. 77–88, 2005.

[19] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1980.

[20] J. Nash, "Non-cooperative games," *Annals of Math.*, vol. 2, pp. 286–295, 1951.

[21] H. von Stackelberg, *The Theory of the Market Economy*. Oxford Univ. Press, 1952.

[22] R. Isaacs, *Differential Games*. John Wiley, 1967.

[23] K. Hipel, K. Radford, and L. Fang, "Multiple participant-multiple criteria decision making." *IEEE Trans. Syst. Man Cybern.*, vol. 23, pp. 1184–1189, 1993.

[24] S. Zoints, "Multiple criteria mathematical programming: An overview and several approaches." *Mathematics of Multi-Objective Optimization*, pp. 227–273, 1985.

[25] Y. Sawaragi, H. Nakayama, and T. Tanino, *Theory of Multi-objective Optimization*. Academic Press, 1985.

[26] K.-C. Chu, "Team decision theory and information structures in optimal control problems-part ii," *IEEE Trans. Autom. Contr.*, vol. 17, pp. 22–28, 1972.

[27] C.-Y. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems-part i," *IEEE Trans. Autom. Contr.*, vol. 17, pp. 15–22, 1972.

[28] K. Kim and F. W. Roush, *Team Theory*. Ellis Horwood Limited, 1987.

[29] W. Bialas, "Cooperative n-person Stackelberg games," *IEEE Conf. Decision and Control*, pp. 2439–2444, 1989.

[30] G. Owen, *Game Theory*. Academic Press, 1982.

[31] R. Isaacs, *Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization.* Dover Pubns, 1999.

[32] S. Tijs, *Introduction to Game Theory.* Hindustand Book Agency, 2003.

[33] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory.* SIAM, PA, 1999.

[34] M. Bloem, T. Alpcan, and T. Başar, "A Stackelberg game for power control and channel allocation in cognitive radio networks," in *Proc. Int. Conf. Perform. Eval. Methodol. Tools.* ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, pp. 1–9.

[35] T. Basar and H. Selbuz, "Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems," *IEEE Trans. Autom. Control*, vol. 24 no. 2, pp. 166–179, 1979.

[36] J. Medanic, "Closed-loop Stackelberg strategies in linear-quadratic problems," *IEEE Trans. Autom. Control*, vol. 23 no. 4, pp. 632–637, 1978.

[37] M. Simaan and J. Cruz, J., "A Stackelberg solution for games with many players," *IEEE Trans. Autom. Control*, vol. 18, pp. 322–324, 1973.

[38] G. Papavassilopoulos and J. Cruz, "Nonclassical control problems and Stackelberg games," *IEEE Trans. Autom. Control*, vol. 24 no. 2, pp. 155–166, 1979.

[39] A. Gambier, A. Wellenreuther, and E. Badreddin, "A new approach to design multi-loop control systems with multiple controllers," in *Proc. IEEE Conf. Decis. Control*, 13-15 2006, pp. 1828 –1833.

[40] J. Hongbin and C. Y. Huang, "Non-cooperative uplink power control in cellular radio systems," *Wireless Networks*, vol. 4 no. 3, pp. 233–240, 1998.

[41] T. Basar and P. Bernhard, *H-infinity Optimal Control and Related Minimax Design Problems.* Boston: Birkhäuser, 2008.

[42] A. Isidori and A. Astolfi, "Disturbance attenuation and H-infinity-control via measurement feedback in nonlinear systems," *IEEE Trans. Autom. Control*, vol. 37, no. 9, pp. 1283–1293, Sept. 1992.

[43] L. Pavel, "A noncooperative game approach to OSNR optimization in optical networks," *IEEE Trans. Autom. Control*, vol. 51 no. 5, pp. 848–852, 2006.

[44] C. J. Tomlin, J. Lygeros, and S. Shankar Sastry, "A game theoretic approach to controller design for hybrid systems," *Proc. of the IEEE*, vol. 88, no. 7, pp. 949–970, 2000.

[45] T. Basar and G. . J. Olsder, "Team-optimal closed loop Stackelberg strategies in hierarchical control problems," *Automatica*, vol. 16 no. 4, pp. 409–414, 1980.

[46] M. Jungers, E. Trelat, and H. Abou-Kandil, "Min-max and min-min Stackelberg strategy with closed-loop information," *HAL Hyper Articles en Ligne*, vol. 3, 2010.

[47] M. Abu-Khalaf and F. 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.

[48] A. Starr and C.-Y. Ho, "Nonzero-sum differential games," *J. Optimiz. Theory App.*, vol. 3, pp. 184–206, 1972.

[49] G. Leitmann, *Cooperative and Non-cooperative Many Playser Differential Games*. Springer, 1974.

[50] D. Yeung and L. Petrosyan, "Subgame consistent solutions of a cooperative stochastic differential game with nontransferable payoffs," *J. Optimiz. Theory App.*, vol. 124, pp. 701–724, 2005.

[51] A. Starr and Ho, "Further properties of nonzero-sum differential games," *J. Optimiz. Theory App.*, vol. 4, pp. 207–219, 1969.

[52] J. Engwerda and A. Weeren, "The open-loop nash equilibrium in lq-games revisited," *Center for Economic Research*, 1995.

[53] J. Case, "Toward a theory of many player differential games," *SIAM*, vol. 7, pp. 179–197, 1969.

[54] A. Friedman, *Differential games*. Wiley, 1971.

[55] T. Basar and P. Bernhard, *Hinfinity- optimal control and related minimax design problems: A dynamic game approach*. Birkhäuser, 1995.

[56] J. Cruz, "Leader-follower strategies for multilevel systems," *IEEE Trans. Autom. Control*, vol. 23 no. 2, pp. 244–255, 1978.

[57] C. I. Chen and J. B. Cruz, "Stackelberg solution for two-person games with biased information patterns," *IEEE Trans. Autom. Control*, vol. 17 no. 6, pp. 791–798, 1972.

[58] J. B. Cruz, "Survey of Nash and Stackelberg equilibrium strategies in dynamic games," *Ann. Econ. Soc. Meas.*, vol. 4 no. 2, pp. 339–344, 1975.

[59] M. Simaan and J. Cruz, "On the Stackelberg strategy in nonzero-sum games," *J. Optimiz. Theory App.*, vol. 11 no. 5, pp. 533–555, 1973.

[60] ——, "Additional aspects of the Stackelberg strategy in nonzero-sum games," *J. Optimiz. Theory App.*, vol. 11 no. 1, pp. 613–626, 1973.

[61] B. Gardner and J. Cruz, "Feedback Stackelberg strategy for a two-player game," *IEEE Trans. Autom. Control*, vol. 22 no. 2, pp. 244–255, 1977.

[62] A. Weeren, J. Schumacher, and J. Engwerda, "Asymptotic analysis of linear feedback nash equilibria in nonzero-sum linear-quadratic differential games," *J. Optimiz. Theory App.*, vol. 101, pp. 693–72, 1999.

[63] G. Freiling, G. Jank, and D. Kremer, "Solvability condition for a nonsymmetric riccati equation appearing in stackelberg games," *Proc. of the European Control Conference*, 2003.

[64] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory.* SIAM, 1999.

[65] A. Van der Schaft, "L2-gain analysis of nonlinear systems and nonlinear H-[infinity] control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, 1992.

[66] A. van der Schaft, "L2 gain analysis of nonlinear systems and nonlinear state-feedback H-infinity control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, June 1992.

[67] ——, "On a state space approach to h-infinity nonlinear control," *Syst. Contr. Lett.*, vol. 16, pp. 1–8, 1991.

[68] A. Isidori and W. Lin, "Global inverse L2-gain state feedback design for a class of nonlinear systems," *IEEE Conf. Decision and Control*, pp. 2831–2836, 1997.

[69] X. Cai, R. Lin, and S. Su, "Robust stabilization for a class of nonlinear systems," in *Proc. Chin. Control Decis. Conf.*, 2008, pp. 4840–4844.

[70] H. Kim, J. Back, H. Shim, and J. H. Seo, "Locally optimal and globally inverse optimal controller for multi-input nonlinear systems," in *Proc. Am. Control Conf.*, 2008, pp. 4486–4491.

[71] Y. Nakamura and H. Hanafusa, "Inverse kinematic solutions with singularity robustness for robot manipulator control," *J. Dyn. Syst. Meas. Contr.*, vol. 108, no. 3, pp. 163–171, 1986.

[72] J. Guojun, "Inverse optimal stabilization of a class of nonlinear systems," in *Proc. Chin. Control Conf.*, 2007, pp. 226–230.

[73] M. Krstic and P. Tsiotras, "Inverse optimal stabilization of a rigid spacecraft," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1042–1049, May 1999.

[74] M. Krstic, "Inverse optimal adaptive control-the interplay between update laws, control laws, and Lyapunov functions," in *Proc. Am. Control Conf.*, 2009, pp. 1250–1255.

[75] Z.-H. Li and M. Krstic, "Optimal design of adaptive tracking controllers for non-linear systems," in *Proc. Am. Control Conf.*, Albuquerque, New Mexico, 1997, pp. 1191–1197.

[76] J. Fausz, V.-S. Chellaboina, and W. Haddad, "Inverse optimal adaptive control for nonlinear uncertain systems with exogenous disturbances," in *Proc. IEEE Conf. Decis. Control*, Dec. 1997, pp. 2654–2659.

[77] W. Luo, Y.-C. Chu, and K.-V. Ling, "Inverse optimal adaptive control for attitude tracking of spacecraft," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1639–1654, Nov. 2005.

[78] L. Sonneveldt, E. Van Oort, Q. P. Chu, and J. A. Mulder, "Comparison of inverse optimal and tuning functions designs for adaptive missile control," *J. Guid. Contr. Dynam.*, vol. 31, no. 4, pp. 1176–1182, 2008.

[79] X.-S. Cai and Z.-Z. Han, "Inverse optimal control of nonlinear systems with structural uncertainty," *IEE Proc. Contr. Theor. Appl.*, vol. 152, no. 1, pp. 79–83, Jan. 2005.

[80] J. Cheng, H. Li, and Y. Zhang, "Robust low-cost sliding mode overload control for uncertain agile missile model," in *Proc. World Congr. Intell. Control Autom.*, Dalian, China, June 2006, pp. 2185–2188.

[81] T. Cheng and F. Lewis, "Neural network solution for finite-horizon H-infinity constrained optimal control of nonlinear systems," vol. 5, no. 1, 2007, pp. 1–11.

[82] T. Cheng, F. Lewis, and M. Abu-Khalaf, "A neural network solution for fixed-final time optimal control of nonlinear systems," *Automatica*, vol. 43, no. 3, pp. 482–490, 2007.

[83] Y. Kim, F. Lewis, and D. Dawson, "Intelligent optimal control of robotic manipulator using neural networks," *Automatica*, vol. 36, no. 9, pp. 1355–1364, 2000.

[84] Y. Kim and F. Lewis, "Optimal design of CMAC neural-network controller for robot manipulators," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 30, no. 1, pp. 22 –31, feb 2000.

[85] K. Dupree, P. Patre, Z. Wilcox, and W. Dixon, "Asymptotic optimal control of uncertain nonlinear euler-lagrange systems," *Automatica*, 2010.

[86] P. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York: Van Nostrand Reinhold, 1992.

[87] D. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming.* Athena Scientific, 1996.

[88] D. V. Prokhorov and I. Wunsch, D. C., "Adaptive critic designs," *IEEE Trans. Neural Networks*, vol. 8, pp. 997–1007, 1997.

[89] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 38, pp. 943–949, 2008.

[90] ——, "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.

[91] B. Widrow, N. Gupta, and S. Maitra, "Punish/reward: Learning with a critic in adaptive threshold systems," *IEEE Trans. Syst. Man Cybern.*, vol. 3, no. 5, pp. 455–465, 1973.

[92] S. Balakrishnan, "Adaptive-critic-based neural networks for aircraft optimal control," *J. Guid. Contr. Dynam.*, vol. 19, no. 4, pp. 893–898, 1996.

[93] G. Lendaris, L. Schultz, and T. Shannon, "Adaptive critic design for intelligent steering and speed control of a 2-axle vehicle," in *Int. Joint Conf. Neural Netw.*, 2000, pp. 73–78.

[94] S. Ferrari and R. Stengel, "An adaptive critic global controller," in *Proc. Am. Control Conf.*, vol. 4, 2002.

[95] D. Han and S. Balakrishnan, "State-constrained agile missile control with adaptive-critic-based neural networks," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 4, pp. 481–489, 2002.

[96] P. He and S. Jagannathan, "Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 37, no. 2, pp. 425–436, 2007.

[97] L. Baird, "Advantage updating," Wright Lab, Wright-Patterson Air Force Base, OH, Tech. Rep., 1993.

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

[99] J. Murray, C. Cox, G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 32, no. 2, pp. 140–153, 2002.

[100] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, vol. 22, no. 3, pp. 237 – 246, 2009.

[101] K. Vamvoudakis and F. Lewis, "Online synchronous policy iteration method for optimal control," in *Recent Advances in Intelligent Control Systems*. Springer, 2009, pp. 357–374.

[102] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. (submitted), 2011.

[103] Q. Wei and H. Zhang, "A new approach to solve a class of continuous-time nonlinear quadratic zero-sum game using adp," in *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on.* IEEE, 2008, pp. 507–512.

[104] H. Zhang, Q. Wei, and D. Liu, "An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games," *Automatica*, vol. 47, pp. 207–214 207–214 207–214, 2010.

[105] X. Zhang, H. Zhang, Y. Luo, and M. Dong, "Iteration algorithm for solving the optimal strategies of a class of nonaffine nonlinear quadratic zero-sum games," in *Control and Decision Conference*, 2010.

[106] A. Mellouk, Ed., *Advances in Reinforcement Learning.* InTech, 2011.

[107] K. Vamvoudakis and F. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, pp. 878–888, 2010.

[108] ——, "Online neural network solution of nonlinear two-player zero-sum games using synchronous policy iteration," in *Proc. IEEE Conf. Decis. Control*, 2010.

[109] ——, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled hamilton-jacobi equations," *Automatica*, 2011.

[110] M. Littman, "Value-function reinforcement learning in markov games," *Cognitive Systems Research*, vol. 2, no. 1, pp. 55–66, 2001.

[111] P. M. Patre, "Lyapunov-based robust and adaptive control of nonlinear systems using a novel feedback structure," Ph.D. dissertation, University of Florida, Gainesville, FL, 2009. [Online]. Available: http://ncr.mae.ufl.edu/dissertations/patre.pdf

[112] J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 81, no. 10, p. 3088, 1984.

[113] A. Poznyak, E. Sanchez, and W. Yu, *Differential neural networks for robust nonlinear control: identification, state estimation and trajectory tracking.* World Scientific Pub Co Inc, 2001.

[114] D. Kirk, *Optimal Control Theory: An Introduction.* Dover Pubns, 2004.

[115] T. Basar, "A counter example in linear-quadratic games: Existence of non-linear nash strategies," *J. Optimiz. Theory App.*, vol. 14, pp. 425–430, 1974.

[116] A. Filippov, "Differential equations with discontinuous right-hand side," *Am. Math. Soc. Transl.*, vol. 42 no. 2, pp. 199–231, 1964.

[117] ——, *Differential equations with discontinuous right-hand side*. Netherlands: Kluwer Academic Publishers, 1988.

[118] G. V. Smirnov, *Introduction to the theory of differential inclusions*. American Mathematical Society, 2002.

[119] F. H. Clarke, *Optimization and nonsmooth analysis*. SIAM, 1990.

[120] D. Shevitz and B. Paden, "Lyapunov stability theory of nonsmooth systems," *IEEE Trans. Autom. Control*, vol. 39 no. 9, pp. 1910–1914, 1994.

[121] B. Paden and S. Sastry, "A calculus for computing Filippov's differential inclusion with application to the variable structure control of robot manipulators," *IEEE Trans. Circuits Syst.*, vol. 34 no. 1, pp. 73–82, 1987.

[122] F. Kydland and E. Prescott, "Rules rather than discretion: The inconsistency of optimal plans," *Journal of Political Economy*, vol. 85 no. 3, pp. 473–492, 1977.

[123] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*. Birkhauser, 2003.

[124] F. L. Lewis, *Optimal Control*. John Wiley & Sons, 1986.

[125] M. Bardi and I. Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 1997.

[126] F. L. Lewis, R. Selmic, and J. Campos, *Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002.

[127] W. E. Dixon, A. Behal, D. M. Dawson, and S. Nagarkatti, *Nonlinear Control of Engineering Systems: A Lyapunov-Based Approach*. Birkhäuser Boston, 2003.

[128] M. Krstic, P. V. Kokotovic, and I. Kanellakopoulos, *Nonlinear and Adaptive Control Design*. John Wiley & Sons, 1995.

[129] P. M. Patre, W. MacKunis, K. Kaiser, and W. E. Dixon, "Asymptotic tracking for uncertain dynamic systems via a multilayer neural network feedforward and RISE feedback control structure," *IEEE Trans. Autom. Control*, vol. 53, no. 9, pp. 2180–2185, 2008.

[130] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[131] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[132] V. Nevistic and J. A. Primbs, "Constrained nonlinear optimal control: a converse HJB approach," California Institute of Technology, Pasadena, CA 91125, Tech. Rep. CIT-CDS 96-021, 1996.

BIOGRAPHICAL SKETCH

Marcus Johnson was born in Seattle, Washington. He received his Bachelor of Engineering degree in aerospace engineering from the University of Florida, USA. He then joined the Nonlinear Controls and Robotics (NCR) research group to pursue his doctoral research under the advisement of Warren E. Dixon and he completed his Doctorate of Philosophy from the University of Florda in August 2011. He has worked as a flight controls engineer for the Tybrin Corporation at NASA Dryden in Edwards CA, from May 2009 through November 2010. Since November 2010 he has currently been working as a guidance, navigation, and control engineer with the Boeing Co. in Huntington Beach CA.