Modeling, Approximation, and Control for a Class of Nonlinear Systems

Parag Suhas Bobade

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Engineering Mechanics

Andrew J. Kurdila, Chair

John A. Burns

Jeffrey T. Borggaard

Craig A. Woolsey

Shane D. Ross

September 29, 2017

Blacksburg, Virginia

Keywords: Adaptive Estimation, Approximation Theory, Functional Differential Equations

Modeling, Approximation, and Control for a Class of Nonlinear Systems

Parag S. Bobade

(ABSTRACT)

This work investigates modeling, approximation, estimation, and control for classes of nonlinear systems whose state evolves in space $\mathbb{R}^n \times H$, where $\mathbb{R}^n$ is a n-dimensional Euclidean space and H is a infinite dimensional Hilbert space. Specifically, two classes of nonlinear systems are studied in this dissertation. The first topic develops a novel framework for adaptive estimation of nonlinear systems using reproducing kernel Hilbert spaces. A nonlinear adaptive estimation problem is cast as a time-varying estimation problem in $\mathbb{R}^d \times H$. In contrast to most conventional strategies for ODEs, the approach here embeds the estimate of the unknown nonlinear function appearing in the plant in a reproducing kernel Hilbert space (RKHS), H. Furthermore, the well-posedness of the framework in the new formulation is established. We derive the sufficient conditions for existence, uniqueness, and stability of an infinite dimensional adaptive estimation problem. A condition for persistence of excitation in a RKHS in terms of an evaluation functional is introduced to establish the convergence of finite dimensional approximations of the unknown function in RKHS. Lastly, a numerical validation of this framework is presented, which could have potential applications in terrain mapping algorithms. The second topic delves into estimation and control of history dependent differential equations. This study is motivated by the increasing interest in estimation and control techniques for robotic systems whose governing equations include

history dependent nonlinearities. The governing dynamics are modeled using a specific form of functional differential equations. The class of history dependent differential equations in this work is constructed using integral operators that depend on distributed parameters. Consequently, the resulting estimation and control equations define a distributed parameter system whose state, and distributed parameters evolve in finite and infinite dimensional spaces, respectively. The well-posedness of the governing equations is established by deriving sufficient conditions for existence, uniqueness and stability for the class of functional differential equations. The error estimates for multiwavelet approximation of such history dependent operators are derived. These estimates help determine the rate of convergence of finite dimensional approximations of the online estimation equations to the infinite dimensional solution of distributed parameter system. At last, we present the adaptive sliding mode control strategy developed for the history dependent functional differential equations and numerically validate the results on a simplified pitch-plunge wing model.

Modeling, Approximation, and Control for a Class of Nonlinear Systems

Parag S. Bobade

(GENERAL AUDIENCE ABSTRACT)

This dissertation aims to contribute towards our understanding of certain classes of estimation and control problems that arise in applications where the governing dynamics are modeled using nonlinear ordinary differential equations and certain functional differential equations. A common theme throughout this dissertation is to leverage ideas from approximation theory to extend the conventional adaptive estimation and control frameworks. The first topic develops a novel framework for adaptive estimation of nonlinear systems using reproducing kernel Hilbert spaces. The numerical validation of the framework presented has potential applications in terrain mapping algorithms. The second topic delves into estimation and control of history dependent differential equations. This study is motivated by the increasing interest in estimation and control techniques for robotic systems whose governing equations include history dependent nonlinearities.

...... to Aai and Baba

Acknowledgments

Let me begin by expressing my deepest gratitude to my advisor, Andrew Kurdila. During the course of my PhD, Dr. K as we fondly address him, introduced me to several new avenues in research, and some of these areas were not my forte when I began. He instilled in me the qualities of analytical rigor and attention to details. He always showed keen interest in listening to my ideas and encouraged independent thinking. I shall always remember and cherish the memories of our weekend research meetings at The Mill Mountains cafe. Dr. K wholeheartedly supported and constantly encouraged me on all my endeavors. I could not have ever thought of pursuing a Masters in Mathematics simultaneously along with the doctoral studies, had it not been for my advisor's tutelage and meticulous training.

A big thanks goes to my committee members John A. Burns, Jeffrey T. Borggaard, Shane Ross, and Craig Woolsey for all the advice and suggestions that helped improve the quality of this dissertation. I am deeply thankful to my mathematics advisor, Dr. Burns, for taking me under his wings and guiding me on my little excursion in Mathematics. His take on several topics pertaining to the mathematical aspects in this dissertation has helped me immensely to acquire wisdom and understand the deeper challenges better. My first têtê-á-têtê with mathematics (at VT) was in a two-semester course on Calculus of Variation and Optimal Control taught by Dr. Borggaard. I am indebted for his teachings that helped me rekindle the interest in mathematics. Dr. Ross over the years has helped me expand my understanding on wide variety of topics in dynamical systems and nonlinear dynamics. I couldn't be more glad to have these interactions with him on topics that I thoroughly enjoy learning more about. I have also always enjoyed the discussions with Dr. Woolsey on topics pertaining to Geometric Mechanics and Control. I am grateful to Dr. Woolsey for sharing his experience on various aspects of research and academia.

Special thanks goes to my colleagues and collaborators, Shirin Dadashi and Suprotim Majumdar. I am grateful for their contributions in numerical validations of the work in this dissertation and without them this dissertation would have taken another year to see the end zone.

I also want say thank you to my amazing roommates and friends in Blacksburg: Balakrishnan, Mahesh, Priyal, Anirudh, Shyam, and Srivatsan for all the great times we had together. I would always be grateful to Saurabh Nath, a dear friend, for introducing me to the wonderful world of droplet dynamics and other interfacial fluid phenomena during our endless, intense conversations over coffee.

I am thankful to the all ESM and ME staff, especially Jessica Grimes and Elizabeth Joslin for their support on various administrative matters.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

All models are wrong, but some are useful.—George E. P. Box

The study of dynamical systems has evolved over centuries. It has now progressed to a multidisciplinary field that spans most areas in science. From fifteenth century celestial mechanics to the twenty first century systems biology, the underlying complexities have often been captured using mathematical models to gain deeper insight into understanding the behavior of a problem. Roughly speaking, a dynamical system models the time evolution of the state of the system. The law that describes this evolution is called the governing equation of the dynamical system. Common questions regarding a dynamical system include:

- How accurately does the governing equation capture the behavior of the system?
- Can we predict the behavior if any unmodeled disturbances affect the system ?

- Can we maintain the state or drive the state of the system to desired state ?

All the above questions often come up in engineering applications. The ideas developed from study of dynamical system theory provide perspicacity to tackle such questions. These questions and the corresponding answers are the focus of the field of control theory and control engineering.

Estimation and control problems are ubiquitous in a wide range of engineering applications. Fundamentally, estimation problems involve the ability to approximate or predict an unknown or uncertain quantity using the known input data or information. Consequently, having a reliable estimate of an unknown helps in designing a controller to deal with a possibly uncertain or worst-case scenario. Hence, often in the past decades, the progress in estimation theory and control theory has evolved hand-in-hand with each other, and each field has benefited from advancements in the other. This dissertation aims to contribute towards our understanding of certain classes of estimation and control problems that arise in applications where the governing dynamics are modeled using nonlinear ordinary differential equations and certain functional differential equations. Although these problems are motivated from real world practical applications, it is important to emphasize that rigorous mathematical treatment is necessary to understand the nuances of these formulations.

The outline of this dissertation is as follows. Chapter 1 presents the motivation for the problems studied in this dissertation. This chapter also reviews past literature on topics pertaining to the problems. The contributions of this dissertation are outlined at the end of this chapter. Chapter 2 covers the mathematical preliminaries and the necessary background for the subsequent chapters. The framework for adaptive estimation in reproducing kernel Hilbert spaces is developed in chapter 3. Subsequently, numerical validation of the adaptive estimation is presented. Chapter 4 delves into online estimation and control problem for history dependent functional differential equations. The class of history dependent operator is defined and the details regarding their construction are explained. The penultimate chapter 5 discusses persistence of excitation condition and its ramification for parameter convergence in adaptive estimation. Lastly we conclude by discussing the key lessons learned from this study and present future directions for extension of the work in this dissertation.

1.1 Motivation

This section is divided into two parts. The first part in Section 1.1 provides the motivation for the framework developed in the Chapter 3. The second part presents challenges in modeling the unsteady aerodynamics of a flapping wing robot that lays the groundwork for methods described in Chapter 4.

1.1.1 Road and Terrain Mapping

There has been a steep rise of interest in the last decade among researchers in academia and the commercial sector in autonomous vehicles and self driving cars. Although adaptive estimation has been studied for some time, applications such as terrain or road mapping continue to challenge researchers to further develop the underlying theory and algorithms in this field. These vehicles are required to sense the environment and navigate surrounding terrain without any human intervention. The environmental sensing capability of such vehicles must be able to navigate off-road conditions or to respond to other agents in urban settings. As a key ingredient to achieve these goals, it can be critical to have a good *a priori* knowledge of the surrounding environment as well as the position and orientation of the vehicle in the environment. To collect this data for the construction of terrain maps, mobile vehicles equipped with multiple high bandwidth, high resolution imaging sensors are deployed. The mapping sensors retrieve the terrain data relative to the vehicle and navigation sensors provide georeferencing relative to a fixed coordinate system. The geospatial data, which can include the digital terrain maps acquired from these mobile mapping systems, find applications in emergency response planning and road surface monitoring. Further, to improve the ride and handling characteristic of an autonomous vehicle, it might be necessary that these digital terrain maps have accuracy on a sub-centimeter scale.

One of the main areas of improvement in the current state of the art for robotic or autonomous vehicles is in their localization. Localization is the process of tracking the position of the vehicle relative to the surrounding environment. Since localization heavily relies on the quality of GPS/GNSS, IMU data, it is important to come up with novel approaches that fuse the data from multiple sensors to generate the best possible estimate of the environment. Contemporary data acquisition systems used to map the environment generate scattered data sets in time and space. These data sets must be either post-processed or processed online for construction of three dimensional terrain maps.

Fig.1.1 and Fig.1.2 depict a map building vehicle and trailer developed by Prof.John Ferris and his students at Virginia Tech. The system generates experimental observations in the form of data that is scattered in time and space. These data sets have extremely high dimensionality. Roughly 180 million scattered data points are collected per minute of data acquisition, which corresponds to a data file of roughly $\mathcal{O}(1GB)$ in size. Current algorithms and software developed in-house at Virginia Tech post-process the scattered data to generate road and terrain maps. This offline batch computing problem can take many days of computing time to complete. It remains a challenging task to derive a theory and associated algorithms that would enable adaptive or online estimation of terrain maps from such high dimensional, scattered measurements.

Figure 1.1: Vehicle Terrain Measurement System, Virginia Tech

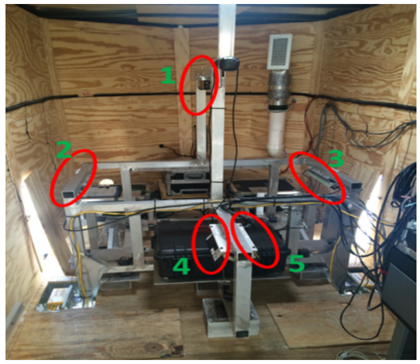


Figure 1.2: Experimental Setup with LMI 3D GO-Locator Lasers

1.1.2 Unsteady Aerodynamics and Flapping Flight

Biomimetics and bio-inspired robotics has been generated huge interest among the researchers in the past few decades. Particularly, understanding and attempting to mimic bird flight has helped push the boundaries of robotics in various aspects. The ultimate goal is to be able to mimic the flight abilities of a bird and use the merits of flapping wing to improve the agility and performance while flying. Due to the inherent nature of unsteady aerodynamics encountered in flapping wing flights, designing such vehicles becomes challenging task. In the early days of flapping wing flight research [80], [35] the wings were modeled as single rigid body without any articulations, to study the aerodynamic effects in flapping/oscillating rigid plates or membranes. This simplified approach helped researchers understand the effect of vortex formation and shedding, on drag and lift forces during flapping flight. Later, with detailed studies in bio-mechanics of bird flight, the new flapping mechanism models incorporated multi-body articulations to mimic the effect of bones, joints and muscles. These models aided assessment of flight stability and control. The aerodynamic forces were modeled as functions of state variables, with the underlying assumption of quasi-steady flow. The infinite dimensional coupled problem (Navier Stokes and multi-body dynamics) was reduced to a set of finite dimensional ordinary differential equations(ODEs). Since the flapping wing flight is affected by unsteady aerodynamics, these are not represented in the quasi-steady models of flapping flight mechanisms. Tobak et al. [79] represented the aerodynamic forces as a functional of state variables, which helped incorporate the unsteady effects. This emphasizes the fact that aerodynamic forces not only can depend on instantaneous state but also on the history of the state variables. Numerically intensive computational fluid dynamics (CFD) presents a precise method to simulate and study the unsteady lift and drag aerodynamic forces. Generally CFD methods exploit high dimensional models that incorporate computationally expensive moving boundary techniques for the Navier-Stokes equations. They are powerful tools to explain detailed characteristics of the aerodynamic forces. One of the characteristics that has inspired the approach here is the history dependence of the aerodynamic lift and drag functions. We refer the interested reader to [15] to study this phenomena in detail. Although CFD methods are advantageous in terms of their accuracy, they suffer from curse of dimensionality which makes them unfavorable choice for online control applications.

1.2 Literature Review : Adaptive Estimation

In chapter 3 we introduce a novel theory and associated algorithms that are amenable to observations that take the form of scattered data. Although the study of infinite dimensional distributed parameter systems can be substantially more difficult than the study of ODEs, a key result developed here is that stability and convergence of the approach can be established succinctly in many cases. Much of the complexity [6, 8] associated with construction of Gelfand triples or the analysis of infinitesimal generators and semigroups that define a DPS can be avoided for many examples of the systems in this dissertation. While our formulation is motivated by this particular applications, it is a general construction for framing and generalizing some conventional approaches for online adaptive estimation. This framework introduces sufficient conditions that guarantee convergence of estimates in spatial domain to the unknown function f appearing in the governing systems of ODEs. In contrast, nearly all conventional strategies [58, 40, 71, 34] consider stability and convergence in time alone for some fixed finite dimensional. The general theory derived in this work has been motivated in part by the terrain mapping application, but also by recent research in a number of fields related to estimation of nonlinear functions.

In this subsection we briefly review some of the recent research in probabilistic or Bayesian mapping methods, nonlinear approximation and learning theory, statistics, and nonlinear regression.

1.2.1 Bayesian and Probabilistic Mapping

Many popular known techniques adopt a probabilistic approach towards solving the localization and mapping problem in robotics. See [25, 78, 32, 4] for an overview of state-of-the-art in this field. The algorithms used to solve this problem fundamentally rely on Bayesian estimation techniques such as particle filters, Kalman filters and other variants of these methods [78, 32, 4]. The computational efforts required to implement these algorithms can be substantial since they involve constructing and updating maps while simultaneously tracking the relative locations of agents with respect to the environment. Over the last three decades significant progress has been made on various frontiers because of high-end sensing capabilities, faster data processing hardware, and efficient computational algorithms [25, 24]. However, the usual Kalman filter based approaches implemented in these applications often are required to address the inconsistency problem in estimation that arise from uncertainties in state estimates [73, 42]. Furthermore, it is well acknowledged among the community that these methods suffer from a major drawback of '*closing the loop*'. This refers to the ability to adaptively update the information if it is revisited. Such a capability for updating information demands huge memory to store the high resolution and high bandwidth data. Moreover, it is highly nontrivial to guarantee that the uncertainties of functions estimates would converge to a lower bound at sub optimal rates, since matching these rates and bounds significantly constrain the evolution of states along infeasible trajectories. While probabilistic methods, and in particular Bayesian estimation techniques, for the construction of terrain maps have flourished over the past few decades, relatively few approaches have appeared for establishing deterministic theoretical error bounds in the spatial domain of the unknown function representing the terrain.

1.2.2 Approximation and Learning Theory

Approximation theory has a long history, but the subtopics of most relevance to this dissertation include recent studies in multiresolution analysis (MRA), radial basis function (RBF) approximation and learning theory. The study of MRA techniques became popular in the late 1980's and early 1990's, and it has flourished since that time. We use only a small part of the general theory of MRAs here, and we urge the interested reader to consult one of the excellent treatises on this topic for a full account. References [55, 54, 17, 23] are good examples of such detailed treatments. We briefly summarize the pertinent aspects of MRA in Section 2.2.2. Our interest in multiresolution analysis arises since these methods can be used to develop multiscale kernels for RKHS, as summarized in [62, 61]. We only consider approximation spaces defined in terms of the scaling functions in this dissertation. Specifically, with a parameter $s \in \mathbb{R}^+$ measuring smoothness, we use $s-$regular MRAs to define admissible kernels for the reproducing kernels that embody the online and adaptive estimation strategies. When the MRA bases are smooth enough, the RKHS kernels derived from a MRA can be shown to be equivalent to a scale of Sobolev spaces having well documented approximation properties. The B-spline bases in the numerical examples yield RKHS embeddings with good condition numbers. We briefly summarize the pertinent aspects of MRA in Section 2.2.2.

1.2.3 Learning Theory and Nonlinear Regression

The methodology defined for online adaptive estimation can be viewed as similar in philosophy to the recent efforts that synthesize learning theory and approximation theory. In [22, 46, 11, 77], independent and identically distributed observations of some unknown function are collected, and they are used to define an estimator of that unknown function. Sharp estimates of error, guaranteed to hold in probability spaces, are possible using tools familiar from learning theory and thresholding in approximation spaces. The approximation spaces are usually defined terms of subspaces of an MRA. However, there are a few key differences between the these efforts in nonlinear regression and learning theory and this paper. The learning theory approaches to estimation of the unknown function depend on observations of the function itself. In contrast, the adaptive online estimation framework here assumes that observations are made of the estimator states, not directly of the unknown function itself. The learning theory methods also assume a discrete measurement process, instead of the continuous measurement process that characterizes online adaptive estimation. On the other hand, the methods based on learning theory derive sharp function space rates of convergence of the estimates of the unknown function. Such estimates are not available in conventional online adaptive estimation methods. Typically, convergence in adaptive estimation strategies is guaranteed in time in a fixed finite dimensional space. One of the significant contributions of this dissertation is to construct sharp convergence rates in function spaces, similar to approaches in learning theory, of the unknown function using online adaptive estimation.

1.2.4 Online Adaptive Estimation and Control

Since the approach here generalizes a standard strategy in online adaptive estimation and control theory, we review this class of methods in some detail. This summary will be crucial in understanding the nuances of the proposed technique and in contrasting the sharp estimates of error available in the new strategy to those in the conventional approach. Many popular textbooks study online or adaptive estimation within the context of adaptive control theory for systems governed by ordinary differential equations [71, 40, 34]. The theory has been extended in several directions, each with its subtle assumptions and associated analyses.

Adaptive estimation and control theory has been refined for decades, and significant progress has been made in deriving convergent estimation and stable control strategies that are robust with respect to some classes of uncertainty. The efforts in [6, 8] are relevant to this paper, where the authors generalize some of adaptive estimation and model reference adaptive control (MRAC) strategies for ODEs so that they apply to deterministic infinite dimensional evolution systems. In addition, [29, 30, 31, 63] also investigate adaptive control and estimation problems under various assumptions for classes of stochastic and infinite dimensional systems. Recent developments in $\mathcal{L}^1$ control theory as presented in [36], for example, utilize adaptive estimation and control strategies in obtaining stability and convergence for systems generated by collections of nonlinear ODEs.

We consider a model problem in which the plant dynamics are generated by the nonlinear ordinary differential equations

$$\dot{x}(t) = Ax(t) + Bf(x(t)), \qquad x(0) = x_0 \tag{1.1}$$

with state $x(t) \in \mathbb{R}^d$, the known Hurwitz system matrix $A \in \mathbb{R}^{d\times d}$, the known control influence matrix $B \in \mathbb{R}^d$, and the unknown function $f : \mathbb{R}^d \to \mathbb{R}$. Although this model problem is an exceedingly simple prototypical example studied in adaptive estimation and control of ODEs [71, 40, 34], it has proven to be an effective case study in motivating alternative formulations such as in [36] and will suffice to motivate the current approach. Of course, much more general plants are treated in standard methods [71, 40, 34, 58] and can be attacked using the strategy that follows. This structurally simple problem is chosen so as to clearly illustrate the essential constructions of the RKHS embedding method while omitting the nuances associated with general plants. A typical adaptive estimation problem can often be formulated in terms of an estimator equation and a learning law. One of the simplest estimators for this model problem takes the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{f}(t, x(t)), \qquad \hat{x}(0) = x_0 \tag{1.2}$$

where $\hat{x}(t)$ is an estimate of the state $x(t)$ and $\hat{f}(t, x(t))$ is time varying estimate of the unknown function f that depends on measurement of the state $x(t)$ of the plant at time t. When the state error $\tilde{x} := x - \hat{x}$ and function estimate error $\tilde{f} := f - \hat{f}$ are defined, the state error equation is simply

$$\dot{\tilde{x}}(t) = A\tilde{x}(t) + B\tilde{f}(t, x(t)), \qquad \tilde{x}(0) = \tilde{x}_0. \tag{1.3}$$

The goal of adaptive or online estimation is to determine a learning law that governs the evolution of the function estimate $\hat{f}$ and guarantees that the state estimate $\hat{x}$ converges to the true state x, $\tilde{x}(t) = x(t) - \hat{x}(t) \to 0$ as $t \to \infty$. Perhaps additionally, it is hoped that the function estimates $\hat{f}$ converge to the unknown function f, $\tilde{f}(t) = f(t) - \hat{f}(t) \to 0$ as $t \to \infty$. The choice of the learning law for the update of the adaptive estimate $\hat{f}$ depends intrinsically on what specific information is available about the unknown function f. It is most often the case for ODEs that the estimate $\hat{f}$ depends on a finite set of unknown parameters $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$. The learning law is then expressed as an evolution law for the parameters $\hat{\alpha}_i$, $i = 1, \ldots, n$. The discussion that follows emphasizes that this is a very specific underlying assumption regarding the information available about unknown function f. Much more general prior assumptions are possible.

Classes of Uncertainty in Adaptive Estimation

The adaptive estimation task seeks to construct a learning law based on the knowledge that is available regarding the function f. Different methods for solving this problem have been developed depending on the type of information available about the unknown function f. The uncertainty about f is often described as forming a continuum between structured and unstructured uncertainty. In the most general case, we might know that f lies in some compact set $\mathcal{C}$ of a particular Hilbert space of functions H over a subset $\Omega \subseteq \mathbb{R}^d$. This case, that reflects in some sense the least information regarding the unknown function, can be expressed as the condition that $f \in \{g \in \mathcal{C} | \mathcal{C} \subset H\}$, for some compact set of functions $\mathcal{C}$ in a Hilbert space of functions H. In approximation theory, learning theory, or non-parametric estimation problems this information is sometimes referred to as the *prior*, and choices of H are commonly known as the hypothesis space. The selection of the hypothesis space H and set $\mathcal{C}$ often reflect the approximation, smoothness, or compactness properties of the unknown function [22]. This example may in some sense utilize only limited or minimal information regarding the unknown function f, and in this case we may refer to the uncertainty as unstructured. Numerous variants of conventional adaptive estimation admit additional knowledge about the unknown function. In most conventional cases the unknown function f is assumed to be given in terms of some fixed set of parameters. This situation is similar in philosophy to problems of parametric estimation which restrict approximants to classes of functions that admit representation in terms of a specific set of parameters. Suppose the finite dimensional basis $\{\phi_k\}_{k=1,\ldots,n}$ is known for a particular finite dimensional subspace $H_n \subseteq H$ in which the function lies, and further that the uncertainty is expressed as the condition that there is a unique set of unknown coefficients $\{\alpha_i^*\}_{i=1,\ldots,n}$ such that $f := f^* = \sum_{i=1,\ldots,n} \alpha_i^* \phi_i \in H_n$. Consequently, conventional approaches may restrict the adaptive estimation technique to construct an estimate with knowledge that f lies in the set

$$f \in \left\{ g \in H_n \subseteq H \,\middle|\, g = \sum_{i=1,\ldots,n} \alpha_i \phi_i \text{ with } \alpha_i \in [a_i, b_i] \subset \mathbb{R} \text{ for } i = 1, \ldots, n \right\} \tag{1.4}$$

This is an example where the uncertainty in the estimation problem may be said to be structured. The unknown function is parameterized by the collection of coefficients $\{\alpha_i^*\}_{i=1,\ldots,n}$. In this case the compact set the $\mathcal{C}$ is a subset of H_n. As we discuss in sections 2.2, and 3.2, the RKHS embedding approach can be characterized by the fact that the uncertainty is more general and even unstructured, in contrast to conventional methods. In this dissertation we assume uncertainty to be unstructured.

1.3 Literature Review : History Dependent Differential Equations

It is typical in texts that introduce the fundamentals of modeling, stability, and control of robotic systems to assume that the underlying governing equations consist of a set of coupled nonlinear ordinary differential equations. This is a natural assumption when methods of analytical mechanics are used to derive the governing equations for systems composed of rigid bodies connected by ideal joints. A quick perusal of the textbooks [76], [74], or [52], for example, and the references therein gives a good account of the diverse collection of approaches that have been derived for this class of robotic system over the past few decades. These methods have been subsequently refined by numerous authors. Over roughly the same period, the technical community has shown a continued interest in systems that are governed by nonlinear, functional differential equations. These methods that helped to define the direction of initial efforts in the study of well-posedness and stability include [56], [48],[49], and their subsequent development is expanded in [28], [68], [69]. More recently, specific control strategies for classes of functional differential equations have appeared in [70], [39], and [38].

The research described in some cases above deals with quite general plant models. These can include classes of delay equations and general history dependent nonlinearities. One rich collection of history dependent models includes hysteretically nonlinear systems. General discussions of nonlinear hysteresis models can be found in [81] or [10], and some authors have studied the convergence and stability of systems with nonlinear hysteresis. For example, a synthesis of controllers for single-input / single-output functional differential equations is presented in [70] and [38], and these efforts include a wide class of scalar hysteresis operators.

The success of adaptive control strategies in classical manipulator robotics, as exemplified by [76], [52], [74], can be attributed to a large degree to the highly structured form of the governing system of nonlinear ordinary differential equations. As is well-known, much of the body of work in adaptive control for robotic systems relies on traditional linear-in-parameters assumptions.

We illustrate the class of models that are considered by outlining a variation on two familiar problems encountered in robotic manipulator dynamics, estimation, and control. Consider the task of developing a model and synthesizing a controller for a flapping wing, test robot that will be used to study flapping aerodynamics in a wind tunnel. See [3] for such a system that has been developed by researchers at Brown University over the past few years. Dynamics for a ground based flapping wing robot can be derived using analytical mechanics in a formulation that is tailored to the structure of a serial kinematic chain [52], [76], [74].

The equations of motion take the form

$$M(q(t))\ddot{q}(t) + C(q(t), \dot{q}(t))\dot{q}(t) + \frac{\partial V}{\partial q} = Q_a(t, \mu) + \tau(t) \tag{1.5}$$

where $M(q)(t)\in\mathbb{R}^{N\times N}$ is the generalized inertia or mass matrix, $C(q(t), \dot{q}(t))\in\mathbb{R}^{N\times N}$ is a nonlinear matrix that represents Coriolis and centripetal contributions, V is the potential energy, $Q_a(t, \mu)\in\mathbb{R}^N$ is a vector of generalized aerodynamic forces, and $\tau(t)\in\mathbb{R}^N$ is the actuation force or torque vector. The generalized forces $Q(t, \mu)$ due to aerodynamic loads are assumed to be expressed in terms of history dependent operators. For the current discussion, it suffices to note that the aerodynamic contributions are unknown, nonlinear, unsteady, and notoriously difficult to characterize.

We consider two specific sets of equations in this paper that are derived from the robotic Equations 1.5, both of which have similar form. We are interested in online identification problems in which we seek to find the final state and distributed parameters from observations of the states of the evolution equation. We are also interested in control synthesis where we choose the input to drive the system to some desired configuration, or to track a given input trajectory. To simplify our discussion, and following the standard practice for many control synthesis problems for robotics, we choose the original control input to be a partial feedback linearizing control that states the control problem in a standard form. In the case of online identification,it is conventional to choose the input $\tau = M(q)(u - G_1\dot{q} - G_0 q) - (C(q, \dot{q})\dot{q} +$

$\frac{\partial V}{\partial q}(q))$ so that the governing equations take the form

$$\frac{d}{dt}\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -G_0 & -G_1 \end{bmatrix}\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix}(M^{-1}(q)Q_a(t,\mu) + u(t)). \tag{1.6}$$

in terms of a new input u. The goal in the online identification problem is to learn the parameters μ and limiting values $q_\infty, \dot{q}_\infty$ from knowledge of the inputs and states $(u, q, \dot{q})$. We are also interested in tracking control problems. When the desired trajectory is given by q_d, we choose the input

$$\tau = M(q)(u + \ddot{q}_d - G_1\dot{e} - G_0 e) - (C(q,\dot{q})\dot{q} + \frac{\partial V}{\partial q}(q))$$

and the equations governing the tracking error $e := q - q_d$ take the form

$$\frac{d}{dt}\begin{bmatrix} e(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -G_0 & -G_1 \end{bmatrix}\begin{bmatrix} e(t) \\ \dot{e}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix}(M^{-1}(e + q_d)Q_a(t,\mu) + u(t)) \tag{1.7}$$

In either of the above two cases, we will show in the next section that the equations can be written in the general form

$$\dot{X}(t) = AX(t) + B((\mathcal{H}X)(t) \circ \mu + u(t)). \tag{1.8}$$

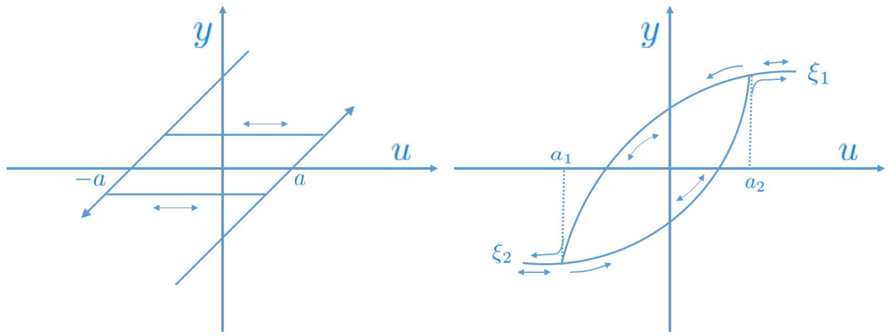
where $A \in \mathbb{R}^{m\times m}$ is the system matrix, $B \in \mathbb{R}^{m\times q}$ is the control input matrix, $u(t) \in \mathbb{R}^q$ is the corresponding input, and $(\mathcal{H}X)(t)$ is a history dependent operator that acts on the distributed parameter μ.

1.3.1 Hysteresis Models

Hysteresis is seen to occur in several day-to-day phenomena. Such systems to exhibit history dependent behavior. Hysteresis by definition is rate independent nonlinearity. The unsteady aerodynamics of a flapping wing mentioned in previous section exhibit hysteresis behavior for the observed aerodynamic loads as function of angle of attack. Therefore, to approximate the hysteresis behavior of such unknown aerodynamics loads, simple scalable mathematical models of hysteresis phenomenon are utilized. Research related to hysteresis models has been widely studied in past and collection of various models are studied in [37], [81], [10] and references therein serve as excellent treatise on these topics. The construction of history dependent operators defined using the generalized play operator is described in Chapter 4. We review some elementary hysteresis models below.

Backlash Hysteresis

Backlash hysteresis describes the play between two mechanical elements. The input versus output graph depicts the hysteresis loop corresponding to the play phenomenon. Formally the backlash operator can be defined as follows- Let $y(t)$ be the output and $u(t)$ the input at time t. If $\dot{u}(t)$ is positive then the output $y(t)$ corresponds to the loading curve described by the line with negative y-intercept in Figure 1.3a. On the other hand if $\dot{u}(t)$ is negative then the output $y(t)$ corresponds to the unloading curve described by the line with positive y-intercept in Figure 1.3a. If $\dot{u}(t) = 0$, i.e input remains constant, then the corresponding



(a) Backlash Hysteresis (b) Relay Hysteresis

Figure 1.3: Hysteresis Models

output remains constant too.

Relay Hysteresis

In the relay hysteresis the relationship between input and output is determined by two input threshold values $a_1 < a_2$. The output $y(t)$ varies on one of two fixed curves $\xi_1 : [a_1, \infty) \to \mathbb{R}$ and $\xi_2 : (-\infty, a_2] \to \mathbb{R}$. For an input $y(t)$, depending on which values among a_1 or a_2 was last visited, the output corresponds the ordinates of the respective curve ξ_1 or ξ_2.

The generalized play operator and the corresponding hysteresis kernel described in subsequent chapters can be intuitively thought of as a suitable combination of rules defined for the above two hysteresis operators. In this dissertation, we choose a typical kernel to be a special case of a generalized play operator [81]. For a piecewise linear function on $[0, t]$ and the an output function is defined by a recursion. Moreover, The recursion above depends on the choice of the left and right bounding ridge functions. Details are outlined in Chapter 4 in Section 4.1.

Before we conclude the first chapter, we would like to summarize the contributions of the dissertation in the next section.

1.4 Contributions of Dissertation

This dissertation contributes to the state-of-the-art understanding of adaptive estimation and history dependent formulations in following ways:

1. Adaptive Estimation using Reproducing Kernel Hilbert Spaces

    (a) We extend the conventional, general framework for online adaptive estimation problems for systems governed by uncertain nonlinear ordinary differential equations. The central feature of the theory introduced here represents the unknown function as a member of a reproducing kernel Hilbert space (RKHS) and defines a distributed parameter system (DPS) that governs state estimates and estimates of the unknown function.

    (b) We establish the well-posedness of the problem in this formulation. We 1) derive the sufficient conditions for the existence, uniqueness and stability of the infinite dimensional online estimation problem, 2) derive the existence and stability of finite dimensional approximations of the infinite dimensional estimates, and 3) determine the sufficient conditions that ensure convergence of finite dimensional approximations to the infinite dimensional online estimates.

   (c) A new condition for persistency of excitation in a RKHS in terms of evaluation functionals in an RKHS is introduced that enables the proof of convergence of the finite dimensional approximations of the unknown function in the RKHS.

   (d) We numerically validate the above framework using an example that has applications in terrain mapping.

2. Online Estimation and Control for History Dependent Differential Equations

   (a) We pose the problem of online estimation and adaptive control for a class of history dependent, functional differential equations that have application to some common mechanical systems.

   (b) The functional differential equations are constructed using integral operators that depend on distributed parameters.

   (c) We establish the well-posedness for the governing equations. That is, we derive existence and uniqueness results for the class of fully actuated robotic systems with history dependent forces in their governing equation of motion.

   (d) We derive rates of approximation for the class of history dependent operators in this paper, and provide sufficient conditions to guarantee that finite dimensional approximations of the online estimation equations converge to the solution of the infinite dimensional, distributed parameter system.

   (e) We develop a adaptive sliding mode control strategy for the history dependent functional differential equations and numerically implement it on a simplified pitch-plunge wing model.

(f) Further, some necessary conditions for convergence of parameter estimates in this setup are briefly discussed and their ramifications on integral operators and the hysteresis kernel are elucidated.

Chapter 2

Mathematical Background

This chapter presents the mathematical preliminaries and necessary background for subsequent chapters. We start with elementary review of preliminary notions in analysis. Later, in Section 2.2 we review some topics from approximation theory, especially those pertaining to kernel based approximation methods. In the end the approximation spaces used in construction of class of history dependent operator are presented.

2.1 Preliminaries

Let us begin by attempting to build an elementary understanding for some common notions encountered in analysis of several topics presented in the dissertation. Evidently, one might see that a recurring theme in this dissertation is that of approximation and estimation. To be able to gauge the how well a certain function is being approximated or estimated, we need to define a notion to relate the closeness of two functions. The notion is defined using a metric and the pair that assigns the distance function to the element in that space is called a metric space. Since we are discussing here a notion of distance it is easy to convince ourselves that this function has to be non-negative and zero if and only if the two points are identical. Moreover, the distance function has to be symmetric and satisfy the triangle inequality, much like the distances in the Euclidean setting. A vector Space is collection of objects that is closed under vector addition and scalar multiplication. The associated metric on this space is derived from the norm and this assigns a notion of length to the vector in the vector space. The complete normed vector space is called as the Banach space. Further, the notion of an angle between the vectors is defined by an inner product. Again, in the familiar Euclidean setting it is the dot product between two vectors. The complete inner product space is called as the Hilbert space. Note that we can derive the norm from a inner product i.e

$$\|x\|^2 = \langle x, x \rangle$$

where $\|\cdot\|$ is the norm and $\langle \cdot, \cdot \rangle$ is an inner product. A Hilbert space is also a Banach space but the converse need not be true.

In this dissertation, significant emphasis is put on establishing the well-posedness of a problem. Particularly, the sufficient conditions for existence, uniqueness and stability are derived. Often,if these issues are not studied, they can lead to faulty numerical implementations or dubious results. We briefly summarize the analytical tools that are employed in this dissertation to investigate well-posedness in subsequent chapters.

2.1.1 Banach Fixed Point Theorem

Banach Fixed Point Theorem is used in Chapter 4 in the proof of Theorem 7 and 1

Definition 1. *Let $(\mathcal{X}, \|\cdot\|)$ be normed vector space. Then the transformation $T : \mathcal{X} \to \mathcal{X}$ is said to be* contraction *if it is Lipschitz continuous with some Lipschitz constant $\alpha < 1$.*

$$\|T(x) - T(y)\| \leq \alpha \|x - y\| \quad \forall x, y \in \mathcal{X}$$

Theorem 1. *If $T : \mathcal{X} \to \mathcal{X}$ is a contraction mapping defined on the Banach space $(\mathcal{X}, \|\cdot\|)$ then T has a unique fixed point x^*. For each point x_o in $\mathcal{X}$, the orbit $(T^n(x_0))$ is guaranteed to converge to that fixed point.*

In the context of differential equations, T can be considered as the integral operator that maps continuous functions into continuous functions. The solution of the differential equation is the desired fixed point to which we would like to converge. The usual approach to prove existence and uniqueness of solutions is two pronged: 1) first, establish that the integral operator is indeed a self-mapping over the Banach space $\mathcal{X}$, and 2) then derive the sufficient condition for T to be contraction mapping in $\mathcal{X}$. These steps will eventually guarantee sufficient conditions for local existence and uniqueness of solutions over an interval $(t_0, t_0 + \delta)$ where δ may depend on the Lipschitz constant.

2.1.2 Semiflow and Semigroups

Definition 2. *A family of operators* $\{S(t)|t \in \mathbb{R}^+\}$ *(with each* $S(t) : \mathcal{X} \to \mathcal{X}$*) is a continuous semigroup provided (i)* $S(0) = I_d$*, (ii)* $S(t)S(\tau) = S(\tau + t)$, $\forall t, \tau \in \mathbb{R}^+$*, and lastly (iii) the mapping :* $(t, y) \to S(t)y$ *is continuous.*

This operator characterizes an autonomous process which is also known as semidynamical system or semiflow. To illustrate this with an example, let us consider the ordinary differential equation

$$\dot{x} = Ax, \qquad x(0) = x_0, \qquad \text{where} \qquad x \in \mathbb{R}^d.$$

The solution to above equation is given as

$$x(t) = e^{At}x_0.$$

It is easy to verify that e^{At} generates a C_0-semigroup in $\mathcal{X} \in \mathbb{R}^d$ since it satisfies all the properties of the semigroup operator. This notion can be extended to nonlinear ordinary differential equations that model the evolution of states in the Banach space. This idea is also used to establish existence and uniqueness of mild solution for the governing equations formulated in Chapter 3.

2.1.3 Zorn's Lemma

Definition 3. $\mathfrak{X}$ *is a partially ordered nonempty set in which every totally ordered subset of* $\mathfrak{X}$ *has an upper bound, then* $\mathfrak{X}$ *contains at least on maximal element.*

A partial order is formal way of generalizing the notion of arrangements of elements by some relation over all the elements in the set $\mathfrak{X}$. Without going further down the rabbit hole, the key point is that Zorn's lemma enables existence of maximal solutions of evolution equations in Chapters 3 and 4. In context of differential equations the existence and uniqueness of local solutions derived from fixed point theorem can be extended to the maximal domain. It is therefore a handy tool that is used in Chapter 3 and 4.

The section presents detailed discussion of various topics in approximation theory. These results are required in the proof and discussion of Theorems 1 to 11.

2.2 Topics in Approximation Theory

2.2.1 Reproducing Kernel Hilbert Spaces (RKHS)

Estimation techniques for distributed parameter systems have been previously studied in [5], and further developed to incorporate adaptive estimation of parameters in certain infinite dimensional systems by [6] and the references therein. These works also present the necessary conditions required to achieve parameter convergence during online estimation in the general infinite dimensional setting. But both of these approaches rely on delicate semigroup analysis of Gelfand triples. The approach herein is much simpler and amenable to a broad class of applications of interest in classical adaptive or online estimation. It appears to be a relatively simpler, practical approach to generalize conventional methods for ODEs, without the complexity of using Gelfand triples explicitly. We consider estimation problems that are cast in terms of the unknown function $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$, and our approximations will assume that this function is an element of a reproducing kernel Hilbert space.

The key attribute of the approach is that the unknown function representing the terrain is viewed as an element of a RKHS. The RKHS is constructed in terms of a kernel function $k(\cdot,\cdot) : \Omega \times \Omega \to \mathbb{R}$ where $\Omega \subseteq \mathbb{R}^d$ is the domain over which scattered measurements are made. The kernel k can often be used to define a collection of radial basis functions (RBFs) $k_x(\cdot) := k(x,\cdot)$, each of which is said to be centered at some point $x \in \Omega$. For example, these RBFs might be exponentials, wavelets, or thin plate splines [82]. By embedding the unknown function that represents the terrain in a RKHS, the new formulation generates a system that constitutes a distributed parameter system. The unknown function, representing map terrain, is the infinite dimensional distributed parameter.

The kernel $k(\cdot,\cdot) : \Omega \times \Omega \to \mathbb{R}$ that defines the RKHS provides a natural collection of bases for approximate estimates of the solution that are based directly on some subset of scattered measurements $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$. It is typical in applications to select the centers $\{x_i\}_{i=1}^n$ that locate the basis functions from some sub-sample of the locations at which the scattered data is measured. Thus, while we do not study the nuances of such methods, in this work, the formulation provides a natural framework to pose so-called “basis adaptive methods” such as in [26] and the references therein.

One way to define a reproducing kernel Hilbert space relies on demonstrating the boundedness of evaluation functionals, but we briefly summarize a constructive approach that is helpful in applications and in understanding computations such as in our numerical examples.

Let $\mathbb{R}$ denote the real numbers, $\mathbb{N}$ the positive integers, $\mathbb{N}_0$ the non-negative integers, and $\mathbb{Z}$ the integers. We follow the convention that $a \gtrsim b$ means that there is a constant c, independent of a or b, such that $b \leq ca$. When $a \gtrsim b$ and $b \gtrsim a$, we write $a \approx b$. Several function spaces are used in this paper. The p-integrable Lebesgue spaces are denoted $L^p(\Omega)$ for $1 \leq p \leq \infty$, and $C^s(\Omega)$ is the space of continuous functions on Ω all of whose derivatives of the order less than or equal to s are continuous. The space $C^s_b(\Omega)$ is the normed vector subspace of $C^s(\Omega)$ that consists of all $f \in C^s(\Omega)$ whose derivatives of order less than or equal to s are bounded. The space $C^{s,\lambda}(\Omega) \subseteq C^s_b(\Omega) \subseteq C^s(\Omega)$ is the collection of functions with derivatives $\frac{\partial^{|\alpha|} f}{\partial x^{|\alpha|}}$ that are λ-Holder continuous. That is, they satisfy

$$\|f(x) - f(y)\| \leq C\|x - y\|^\lambda.$$

The Sobolev space of functions in $L^p(\Omega)$ that have weak derivatives of the order less than equal to r that lie in $L^p(\Omega)$ is denoted $H^r_p(\Omega)$.

A reproducing kernel Hilbert space is constructed in terms of a symmetric, continuous, and positive definite function $k : \Omega \times \Omega \to \mathbb{R}$, where positive definiteness requires that there is a fixed positive constant $c > 0$ such that for any finite collection of points $\{x_i\}_{i=1}^n \subseteq \Omega$

$$\sum_{i,j=1}^{n} k(x_i, x_j)\alpha_i\alpha_j \gtrsim c\|\alpha\|^2_{\mathbb{R}^n}$$

for all $\alpha = \{\alpha_1, \ldots, \alpha_n\}^T$. For each $x \in \Omega$, we denote the function $k_x := k_x(\cdot) = k(x, \cdot)$ and refer to k_x as the kernel function centered at x. In many typical examples [82], k_x can be interpreted literally as a radial basis function centered at $x \in \Omega$. For other definitions of kernels, such as those defined in terms of a multiresolution analysis, the center x can be understood as a representative point in the support of k. For any kernel functions k_x and k_y centered at $x, y \in \Omega$, we define the inner product $(k_x, k_y)_H := k(x, y)$. The RKHS H is then defined as the completion of all finite sums extracted from the set $\{k_x | x \in \Omega\}$, i.e, it is closed linear span of $\{k_x\}_{x \in \Omega}$. It is well known that this construction guarantees the boundedness of the evaluation functionals $E_x : H \to \mathbb{R}$. In other words for each $x \in \Omega$ there is a constant c_x such that

$$|E_x f| = |f(x)| \leq c_x \|f\|_H$$

for all $f \in H$. The reproducing property of the RKHS H plays a crucial role in the analysis here, and it states that

$$E_x f = f(x) = (k_x, f)_H$$

for $x \in \Omega$ and $f \in H$. We will also require the adjoint operator $E_x^* : \mathbb{R} \to H$, which can be calculated directly by noting that

$$(E_x f, \alpha)_{\mathbb{R}} = (f, \alpha k_x)_H = (f, E_x^* \alpha)_H$$

for $\alpha \in \mathbb{R}$, $x \in \Omega$ and $f \in H$. Hence, $E_x^* : \alpha \mapsto \alpha k_x \in H$.

Finally, we will have particular interest for case in which it is possible to show that the RKHS H is a subset of $C(\Omega)$, which implies that the associated injection $i : H \to C(\Omega)$ is uniformly bounded.

Example: The Exponential Kernel

A popular example of an RKHS, one that will be used in the numerical examples, is constructed from the family of exponentials $k(x,y) := e^{-\|x-y\|^2/\sigma^2}$ where $\sigma > 0$. Suppose that $\tilde{C} = \sqrt{\sup_{x\in\Omega} k(x,x)} < \infty$. Smale and Zhou in [75] argue that

$$|f(x)| = |E_x(f)| = |(k_x, f)_H| \leq \|k_x\|_H \|f\|_H$$

for all $x \in \Omega$ and $f \in H$, and since $\|k_x\|^2 = |k(x,x)| \leq \tilde{C}^2$, it follows that the embedding $i : H \to C(\Omega)$ is bounded,

$$\|f\|_{C(\Omega)} := \|i(f)\|_{C(\Omega)} \leq \tilde{C}\|f\|_H.$$

For the exponential kernel above, $\tilde{C} = 1$. Let $C^s(\Omega)$ denote the space of functions on Ω all of whose partial derivatives of order less than or equal to s are continuous. The space $C_b^s(\Omega)$ is the subspace $C^s(\Omega)$ endowed with the norm

$$\|f\|_{C_b^s(\Omega)} := \max_{|\alpha|\leq s} \left\| \frac{\partial^{|\alpha|} f}{\partial x^\alpha} \right\|_{C(\Omega)},$$

with the summation taken over multi-indices $\alpha := \{\alpha_1, \ldots, \alpha_d\} \in \mathbb{N}^d$, $\partial x^\alpha := \partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}$, and $|\alpha| = \sum_{i=1,\ldots,d} \alpha_i$. Observe that the continuous functions in $C^s(\Omega)$ need not be bounded even if Ω is a bounded open domain. The space $C^{s,\lambda}(\Omega)$ is the subspace of functions f in $C^s(\Omega)$ for which all of the partial derivatives $\frac{\partial f^{|\alpha|}}{\partial x^\alpha}$ with $|\alpha| \leq s$ are λ-Holder continuous. The norm of $C^{s,\lambda}(\Omega)$ for is given by

$$\|f\|_{C^{s,\lambda}(\Omega)} = \|f\|_{C^s(\Omega)} + \max_{0\leq\alpha\leq s} \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{\left| \frac{\partial^{|\alpha|} f}{\partial x^{|\alpha|}}(x) - \frac{\partial^{|\alpha|} f}{\partial x^{|\alpha|}}(y) \right|}{|x-y|^\lambda}$$

Also, reference [75] notes that if $k(\cdot,\cdot) \in C_b^{2s,\lambda}(\Omega \times \Omega)$ with $0 < \lambda < 2$ and Ω is a closed domain, then the inclusion $H \rightarrow C_b^{s,\lambda/2}(\Omega)$ is well defined and continuous. That is the mapping $i : H \rightarrow C_b^{s,\lambda/2}$ defined via $f \mapsto i(f) := f$ satisfies

$$\|f\|_{C_b^{s,\lambda/2}(\Omega)} \lesssim \|f\|_H.$$

In fact reference [75] shows that

$$\|f\|_{C_b^s(\Omega)} \leq 4^s \|k\|^{1/2}_{C_b^{2s}(\Omega\times\Omega)} \|f\|_H.$$

The overall important conclusion to draw from the summary above is that there are many conditions that guarantee that the embedding $H \hookrightarrow C(\Omega)$ is continuous. This condition will play a central role in devising simple conditions for existence of solutions of the RKHS embedding technique.

2.2.2 Multiscale Kernels Induced by s-Regular Scaling Functions

A multiresolution analysis defines a family of nested approximation spaces $\{H_j\}_{j\in\mathbb{N}} \subseteq H$ of an abstract space H in terms of a single function ϕ, the scaling function. The approximation space H_j is defined in terms of bases that are constructed from dilates and translates $\{\phi_{j,k}\}_{k\in\mathbb{Z}^d}$ with $\phi_{j,k}(x) := 2^{jd/2}\phi(2^j x - k)$ for $x \in \mathbb{R}^d$ of this single function ϕ. It is for this reason that these spaces are sometimes referred to as shift invariant spaces. While the MRA is ordinarily defined only in terms of the scaling functions, the theory provides a rich set of tools to derive bases $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$, or wavelets, for the complement spaces $W_j := V_{j+1} - V_j$. The multiresolution analysis of $L_2(\mathbb{R}^d)$ generated by $\phi \in C^s(\mathbb{R}^d)$ is s-regular for $s \in \mathbb{N}_0$ if for each integer $p \in \mathbb{N}$ and multiindex $a = (a_1, \ldots, a_d) \in \mathbb{N}_0^d$ with $\sum_{i=1}^{d} a_i \leq s$, there is a constant $b_{ap} > 0$ such that

$$|\partial^a \phi(u)| \leq b_{ap}(1 + \|u\|)^{-p}$$

for all $u \in \mathbb{R}^d$. It is known [55, 62, 61] that when $r < s$, the Sobolev space $H_2^r(\mathbb{R}^d)$ of functions that have r weak derivatives in $L_2(\mathbb{R}^d)$ can be characterized as

$$H_2^r(\mathbb{R}^n) := \left\{ f \in L_2(\mathbb{R}^d) \middle| \ \|f\|_{H_2^r}^2 < \infty \right\}$$

with the norm $\|\cdot\|_{H_2^r}$ defined as

$$\|\cdot\|_{H_2^r}^2 := \sum_{j \in \mathbb{N}_0} 2^{2rj} \|(\Pi_j - \Pi_{j-1}) f\|_{L_2(\mathbb{R}^d)}^2$$

where $\Pi_{-1} = 0$ and $\Pi_j : L_2(\mathbb{R}^d) \to V_j$ is the orthogonal projection of $L_2(\mathbb{R}^d)$ onto V_j.

The characterization of the norm of the Sobolev space $H_2^r := H_2^r(\mathbb{R}^d)$ has appeared in many monographs that discuss multiresolution analysis [55, 54, 21]. It is also possible to define the Sobolev space $H_2^r(\mathbb{R}^d)$ as the Hilbert space constructed from a reproducing kernel $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that is defined in terms of an s-regular scaling function ϕ of an multi-resolution analysis (MRA) [55, 21]. For $\frac{d}{2} < r < s$, we define the kernel k in this case as

$$\begin{aligned} k(u, v) :&= \sum_{j=0}^{\infty} 2^{j(d-2r)} \sum_{k \in \mathbb{Z}^d} \phi(2^j u - k)\phi(2^j v - k) \\ &= \sum_{j=0}^{\infty} 2^{-2rj} \sum_{k \in \mathbb{Z}^d} \phi_{j,k}(u)\phi_{j,k}(v). \end{aligned}$$

It should be noted that the requirement $d/2 < r$ implies the coefficient $2^{j(d-2r)}$ above is decreasing as $j \to \infty$ and ensures the summation converges. As discussed in Section 2.2 and in reference [62, 61], the RKHS $H_k^r(\mathbb{R}^d)$ is constructed as the closure of the finite linear span of the set of function $\{k_u\}_{u\in\Omega}$ with $k_u(\cdot) := k(u, \cdot)$. Under the assumption that $\frac{d}{2} < r < s$, the Sobolev space $H_2^r(\mathbb{R}^d)$ can also be related to the Hilbert space $H_k^r(\mathbb{R}^d)$ defined as

$$H_k^r(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \mid (f,f)_{k,r}^{\frac{1}{2}} = \|f\|_{k,r} < \infty \right\}$$

with the inner product $(\cdot,\cdot)_{k,r}$ on $H_k^r(\mathbb{R}^d)$ defined as

$$(f,f)_{k,r} := \|f\|_{k,r}^2 := \inf\left\{ \sum_{j=0}^{\infty} 2^{j(2r-d)} \|f_j\|_{V_j}^2 \middle| f_j \in V_j, f = \sum_{j=0}^{\infty} f_j \right\}$$

with $\|f\|_{V_j}^2 = \sum_{k\in\mathbb{Z}^d} c_{j,k}^2$ for $f_j(u) = \sum_{k\in\mathbb{Z}^d} c_{j,k}\phi(2^j u - k) \in V_j := \operatorname{span}_{\substack{j\in\mathbb{N}_0 \\ k\in\mathbb{Z}^d}} \phi(2^j x - k)$ and $j \in \mathbb{N}_0$. Note that the characterization above of $H_k^r(\mathbb{R}^d)$ is expressed only in terms of the scaling functions $\phi(2^j - k)$ for $j \in \mathbb{N}_0$ and $k \in \mathbb{Z}^d$. The functions ϕ and ψ need not define an orthonormal multiresolution analysis in this characterization, and the bases $\psi_{j,k}$ for the complement spaces $W_j = V_j - V_{j-1}$ are not used. References [62, 61] show that when $d/2 < r < s$, we have the norm equivalence

$$H_k^r(\mathbb{R}^d) \approx H_2^r(\mathbb{R}^d). \tag{2.1}$$

It is well known that when a bounded domain Ω is well aligned with scaling of functions $\phi_{j,k} := 2^{\frac{jd}{2}}\phi(2^j - k)$, the above analysis also holds over the well aligned, bounded domain Ω. This means that we can construct RKHS spaces over Ω from by the restriction of bases to Ω of an MRA. From Sobolev's Embedding Theorem [67], whenever $r > d/2$ we have the embedding

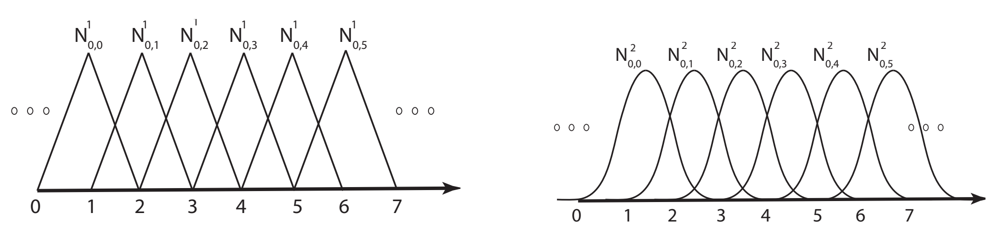
$$H_2^r(\Omega) \hookrightarrow C_b^{r-d/2}(\Omega) \subset C^{r-d/2}(\Omega)$$

In fact, by choosing the s-regular MRA with s and r large enough, we have the embedding $H_2^r(\Omega) \hookrightarrow C(\Omega)$ when $\Omega \subseteq \mathbb{R}^d$ [67].

One of the simplest examples that meet the conditions of this section includes the normalized B-splines of order $r > 0$. We denote by N^r the normalized B-spline of order r with integer knots and define its translated dilates by $N_{j,k}^r := 2^{jd/2} N^r(2^{jd}x - k)$ for $k \in \mathbb{Z}^d$ and $j \in \mathbb{N}_0$. In this case the kernel is written in the form

$$k(u,v) := \sum_{j=0}^{\infty} 2^{-2rj} \sum_{k \in \mathbb{Z}^d} N_{j,k}^r(u) N_{j,k}^r(v).$$

Figure 2.1 depicts the translated dilates of the normalized B-splines of order 1 and 2 respectively.



B-splines N^1 B-splines N^2

Figure 2.1: Translated Dilates of Normalized B-Splines

2.2.3 Approximation Spaces $\mathcal{A}_2^\alpha$

The approximation framework in this dissertation is based on a straightforward implementation of approximation spaces discussed in detail in [23] or [21], and further developed by Dahmen

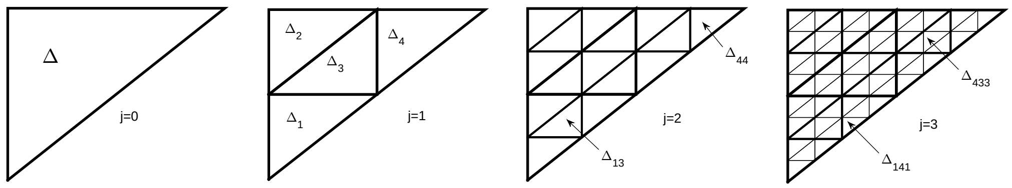


Figure 2.2: Regular refinement process for domain Δ

in [16]. We will see that approximation of the class of history dependent operators under consideration exploit in Chapter 4 a well-known connection between the class of Lipschitz functions and certain approximation spaces as described in [21]. We consider the regular refinement shown in Figure 2.3 where $\Delta_{i_1 i_2}$ is the i_2 child of Δ_{i_1}. In general $\Delta_{i_1 i_2 \ldots i_m i_{m+1}}$ is the $(m+1)^{st}$ child of $\Delta_{i_1 i_2 \ldots i_m}$. The multiscaling function $\phi_{j,k}$ is defined as

$$\phi_{j,k}(x) = \frac{1_{\Delta_{i_1 i_2 \ldots i_j}}(x)}{\sqrt{m(\Delta_{i_1 i_2 \ldots i_j})}}$$

where j refers to the level of refinement in the grid and $m(\cdot)$ is Lebesgue measure.

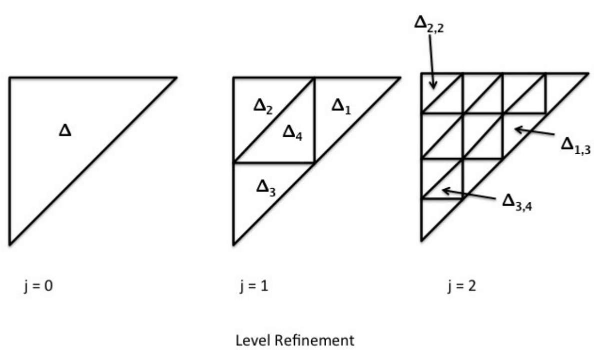


Figure 2.3: j level refinement

Since the history dependent operators $(\mathcal{H}X)(t)$ in Chapter 4 act on the infinite dimensional space $P = P_1 \times \cdots \times P_\ell$ of functions $\mu = (\mu_1, \ldots, \mu_\ell)$, we need approximations of these operators for computations and applications. In the discussion that follows we choose each function $\mu_i \in P_i := L^2(\Delta)$ where the domain $\Delta \subset \mathbb{R}^2$ is defined as

$$\Delta := \left\{ (s_1, s_2) \in \mathbb{R}^2 \middle| \underline{s} \leq s_1 \leq s_2 \leq \overline{s} \right\}.$$

with $\underline{s}, \overline{s} > 0$. The modification of the construction that follows for different domains Δ_i for the functions $\mu_i \in L^2(\Delta_i)$ is trivial, but notationally tedious, and we leave the more general case to the reader. Given the domain Δ we introduce a regular refinement depicted in Figure 2.3 and disscused in more detail in Appendix A. The set Δ is subdivided into $\Delta_1, \Delta_2, \Delta_3, \Delta_4$ as shown, and each Δ_i is subdivided into $\Delta_{i1}, \Delta_{i2}, \Delta_{i3}, \Delta_{i4}$. Further subdivision recursively introduces the sets $\Delta_{i_1 \ldots i_j}$ for $i_j = 1, \ldots, 4$ that are the children of $\Delta_{i_1 \ldots i_{j-1}}$.

The characteristic functions $1_{\Delta_1}, \ldots, 1_{\Delta_4}$ define a collection of multiscaling functions $\phi^1, \ldots, \phi^4$ as defined in [44]. We define the space of piecewise constant functions V_j on grid refinement level j to be the span of the characteristic functions of the sets $\Delta_{i_1, \ldots, i_j}$, so that the dimension of V_j is 4^j. We denote by $\{\phi_{j,k}\}_{k=1,\ldots,4^j}$ the orthonormal basis obtained from these characteristic functions on a particular grid level, each normalized so that $(\phi_{j,k}, \phi_{j,\ell})_{L^2(\Delta)} = \delta_{k,l}$. Each of the basis functions $\phi_{j,k}$ will be proportional to the characteristic function $\phi^\ell(2^j x + d)$ for some $\ell \in, 1, 2, 3, 4$ and displacement vector d. It is straightforward in this case [44] to define 3 piecewise constant multiwavelets ψ^1, ψ^2, ψ^3 that that are used to define functions $\psi_{j,m}$ for $m = 1, \ldots, 3 \times 4^j$ that span the complement spaces $W_j = \text{span}\,\{\psi_{j,m}\}_{m=1,\ldots,3\cdot 4^j} := V_{j+1} - V_j$

$$. \quad \underbrace{V_j}_{4^j functions} = \underbrace{V_{j-1}}_{4^{j-1} functions} \bigoplus \underbrace{W_{j-1}}_{3\times 4^{j-1} functions} \quad .$$

It is a straightforward exercise to define $L^2(\Delta)-$orthonormal wavelets that span W_j for each

$j \in \mathbb{N}_0$, but the nomenclature is lengthy. Since we do not use the wavelets specifically, the details are omitted. Each function $\psi_{j,m}$ is proportional to one of the three scaled and translated multiwavelet functions and satisfies the orthonormality conditions

$$(\psi_{j,k}, \psi_{m,\ell}) = \delta_{j,m}\delta_{k,\ell} \qquad \text{for all } j, k, m, \ell,$$

$$(\psi_{j,k}, \phi_{m,\ell}) = 0 \qquad \text{for } j \geq m \text{ and all } k, \ell.$$

In the next step, we denote the orthogonal projection onto the span of the piecewise constants defined on a grid of resolution level j by Π_j so that

$$\Pi_j : P \to V_j.$$

Finally, we define the approximation space $\mathcal{A}_2^\alpha$ in terms of the projectors Π_j as

$$\mathcal{A}_2^\alpha := \left\{ f \in P \middle| \|f\|_{\mathcal{A}_2^\alpha} := \left(\sum_{j=0}^{\infty} 2^{2\alpha j} \|(\Pi_j - \Pi_{j-1})f\|_P^2 \right)^{1/2} \right\}.$$

Note that this is a special case of the more general analysis in [16]. We define our approximation method in Chapter 4 in terms of one point quadratures defined over the triangles $\Delta_{i_1 \dots i_j}$ that constitute the grid of level j that defines V_j. For notational convenience, we collect all triangles at a fixed level j in the singly indexed set

$$\{\Delta_{j,k}\}_{k \in \Lambda_j} := \left\{\Delta_{i_1 \dots i_j}\right\}_{i_1, \dots, i_j \in 1,2,3,4}$$

where $\Lambda_j := \{k \in \mathbb{N} \quad |1 \leq k \leq 4^j\}$, and the quadrature points are chosen such that $\xi_{j,k} \in \Delta_{j,k}$ for $k = 1, \cdots, \Lambda_j$.

Chapter 3

Adaptive Estimation using Reproducing Kernel Hilbert Spaces

In this chapter we define and elaborate on an extension to conventional online adaptive estimation problems for systems governed by unknown nonlinear ordinary differential equations. The central feature of the theory introduced here represents the unknown function as a element of a reproducing kernel Hilbert space (RKHS) and defines a distributed parameter system (DPS) that governs state estimates and estimates of the unknown function. In this chapter, we

1. derive sufficient conditions for the existence and stability of the infinite dimensional online estimation problem,
2. derive existence and stability of finite dimensional approximations of the infinite dimensional

estimates, and

3. determine sufficient conditions that ensure convergence of finite dimensional approximations to the infinite dimensional online estimates.

We focus our attention on two choices of the RKHS, those that are generated by exponential functions and those that are generated by multiscale kernels defined from a multiresolution analysis. The chapter concludes with an example of an RKHS adaptive estimation problem for a simple model of map building from vehicles. The numerical example demonstrates the rate of convergence for finite dimensional models constructed from RBF bases that are centered at a subset of scattered observations.

3.1 Adaptive Estimation

3.1.1 Adaptive Estimation in $\mathbb{R}^d \times \mathbb{R}^n$

The development of adaptive estimation strategies when the uncertainty takes the form in Equation 1.4 represents, in some sense, an iconic approach in the adaptive estimation and control community. Entire volumes [71, 40, 34, 59] contain numerous variants of strategies that can be applied to solve adaptive estimation problems in which the uncertainty takes the form in Equation 1.4. One canonical approach to such an adaptive estimation problem is governed by three coupled equations: the plant dynamics Equation 3.1, estimator Equation 3.2, and the learning rule. In conventional online estimation, we organize the basis functions as $\phi := [\phi_1, \ldots, \phi_n]^T$ and the true parameters as $\alpha^{*^T} = [\alpha_1^*, \ldots, \alpha_n^*]$, $\hat{\alpha}^T = [\hat{\alpha}_1, \ldots, \hat{\alpha}_n]$. A common gradient based learning law yields the governing equations that incorporate the plant dynamics, estimator equation, and the learning rule, respectively,

$$\dot{x}(t) = Ax(t) + B\alpha^{*^T}\phi(x(t)), \tag{3.1}$$

$$\dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{\alpha}^T(t)\phi(x(t)), \tag{3.2}$$

$$\dot{\hat{\alpha}}(t) = \Gamma^{-1}\phi B^T P(x - \hat{x}), \tag{3.3}$$

where $\Gamma \in \mathbb{R}^{n\times n}$ is symmetric and positive definite. We suppose that A is Hurwitz The symmetric positive definite matrix $P \in \mathbb{R}^{d\times d}$ is the unique solution of Lyapunov's equation $A^T P + PA = -Q$, for some selected symmetric positive definite $Q \in \mathbb{R}^{d\times d}$. Usually the above equations are summarized in terms the two error equations

$$\dot{\tilde{x}}(t) = A\tilde{x} + B\phi^T(x(t))\tilde{\alpha}(t) \tag{3.4}$$

$$\dot{\tilde{\alpha}}(t) = -\Gamma^{-1}\phi(x(t))B^T P\tilde{x}. \tag{3.5}$$

with $\tilde{\alpha} := \alpha^* - \hat{\alpha}$ and $\tilde{x} := x - \hat{x}$. Equations 3.4, 3.5 can also be written as

$$\begin{Bmatrix} \dot{\tilde{x}}(t) \\ \dot{\tilde{\alpha}}(t) \end{Bmatrix} = \begin{bmatrix} A & B\phi^T(x(t)) \\ -\Gamma^{-1}\phi(x(t))B^T P & 0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{\alpha}(t) \end{Bmatrix}. \tag{3.6}$$

This equation defines an evolution on $\mathbb{R}^d \times \mathbb{R}^n$ and has been studied in great detail in [58, 60, 2]. Standard texts such as [71, 40, 34, 59] describes alternatives for the conventional online adaptive estimation problem using projection, least squares methods and other popular approaches.

3.1.2 Adaptive Estimation in $\mathbb{R}^d \times H$

We study the method of RKHS embedding that interprets the unknown function f as an element of the RKHS H, without any *a priori* selection of the particular finite dimensional subspace used for estimation of the unknown function. The counterparts to Equations 3.1, 3.2, 3.3 are the plant, estimator, and learning laws

$$\dot{x}(t) = Ax(t) + BE_{x(t)}f, \tag{3.7}$$

$$\dot{\hat{x}}(t) = A\hat{x}(t) + BE_{x(t)}\hat{f}(t), \tag{3.8}$$

$$\dot{\hat{f}}(t) = \Gamma^{-1}(BE_{x(t)})^* P(x(t) - \hat{x}(t)), \tag{3.9}$$

where as before $x, \hat{x} \in \mathbb{R}^d$, but f and $\hat{f}(t) \in H$, $E_\xi : H \to \mathbb{R}^d$ is the evaluation functional that is given by $E_\xi : f \mapsto f(\xi)$ for all $\xi \in \mathbb{R}^d$ and $f \in H$, and $\Gamma \in \mathcal{L}(H, H)$ is a self adjoint, positive definite linear operator. The error equation analogous to Equation 3.6 is then given by

$$\begin{Bmatrix} \dot{\tilde{x}}(t) \\ \dot{\tilde{f}}(t) \end{Bmatrix} = \begin{bmatrix} A & BE_{x(t)} \\ -\Gamma^{-1}(BE_{x(t)})^* P & 0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix}, \tag{3.10}$$

which defines an evolution on $\mathbb{R}^d \times H$, instead of on $\mathbb{R}^d \times \mathbb{R}^n$.

3.2 Existence, Uniqueness and Stability

In the adaptive estimation problem that is cast in terms of a RKHS H, we seek a solution $X = (\tilde{x}, \tilde{f}) \in \mathbb{R}^d \times H \equiv \mathbb{X}$ that satisfies Equation 3.10. In general $\mathbb{X}$ is an infinite dimensional state space for this estimation problem, which can in principle substantially complicate the analysis in comparison to conventional ODE methods. We first establish that the adaptive estimation problem in Equation 3.10 is well-posed. The result that is derived below is not the most general possible, but rather has been emphasised because its conditions are simple and easily verifiable in many applications.

Theorem 2. *Suppose that $x \in C([0,T];\mathbb{R}^d)$ and that the embedding $i : H \hookrightarrow C(\Omega)$ is uniform in the sense that there is a constant $C > 0$ such that for any $f \in H$,*

$$\|f\|_{C(\Omega)} \equiv \|if\|_{C(\Omega)} \leq C\|f\|_H. \tag{3.11}$$

For any $T > 0$ there is a unique mild solution $(\tilde{X}, \tilde{f}) \in C([0,T],\mathbb{X})$ to Equation 3.10 and the map $X_0 \equiv (\tilde{x}_0, \tilde{f}_0) \mapsto (\tilde{x}, \tilde{f})$ is Lipschitz continuous from $\mathbb{X}$ to $C([0,T],\mathbb{X})$.

Proof. We can split the governing Equation 3.10 into the form

$$\begin{Bmatrix} \dot{\tilde{x}}(t) \\ \dot{\tilde{f}}(t) \end{Bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A_0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix} + \begin{bmatrix} 0 & BE_{x(t)} \\ -\Gamma^{-1}(BE_{x(t)})^* P & -A_0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix}, \tag{3.12}$$

and write it more concisely as

$$\dot{\tilde{X}} = \mathbb{A}\tilde{X}(t) + \mathbb{F}(t, \tilde{X}(t)) \tag{3.13}$$

where the operator $A_0 \in \mathcal{L}(H, H)$ is arbitrary. Then,

$$
\begin{aligned}
\left\| \begin{bmatrix} 0 & BE_{x(t)} \\ -\Gamma^{-1}(BE_{x(t)})^* P & -A_0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix} \right\|_{\mathbb{X}}^2 &= \|BE_{x(t)} f\|_{\mathbb{R}^n}^2 + \|(BE_{x(t)})^* P x - A_0 f\|_H^2 \\
&\le \|B\|^2 \|E_{x(t)}\|^2 \|f\|_H^2 + 2\|(BE_{x(t)})^* P x\|_H^2 + 2\|A_0 f\|_H^2 \\
&\le \|B\|^2 C_k^2 \|f\|_H^2 + 2\|B\|^2 C_k^2 \|x\|_H^2 + 2\|A_0\|^2 \|f\| \\
&= \left(\|B\|^2 C_k^2 + 2\|A_0\|^2\right) \|f\|_H^2 + \left(2\|B\|^2 C_k^2\right) \|x\|_H^2.
\end{aligned}
$$

It is immediately clear that $\mathbb{A}$ is the infinitesimal generator of C_0 semigroup on $\mathbb{X} \equiv \mathbb{R}^d \times H$ since $\mathbb{A}$ is bounded on $\mathbb{X}$. In addition, we see the following:

1. The function $\mathbb{F} : \mathbb{R}^+ \times \mathbb{X} \to \mathbb{X}$ is uniformly globally Lipschitz continuous: there is a constant $L > 0$ such that

   $$\|\mathbb{F}(t, X) - \mathbb{F}(t, Y)\| \le L\|X - Y\|$$

   for all $X, Y \in \mathbb{X}$ and $t \in [0, T]$.

2. The map $t \mapsto \mathbb{F}(t, X)$ is continuous on $[0, T]$ for each fixed $X \in \mathbb{X}$.

By Theorem 1.2, p.184, in reference [65], there is a unique mild solution

$$\tilde{X} = \{\tilde{x}, \tilde{f}\}^T \in C([0, T]; \mathbb{X}) \equiv C([0, T]; \mathbb{R}^d \times H).$$

In fact the map $\tilde{X}_0 \mapsto X$ is Lipschitz continuous from $\mathbb{X} \to C([0, T]; \mathbb{X})$. □

The proof of stability of the equilibrium at the origin of the RKHS Equation 3.10 closely resembles the Lyapunov analysis of Equation 3.6; the extension to consideration of the infinite dimensional state space $\mathbb{X}$ is required. It is useful to carry out this analysis in some detail to see how the adjoint $E_x^* : \mathbb{R} \to H$ of the evaluation functional $E_x : H \to \mathbb{R}$ plays a central and indispensable role in the study of the stability of evolution equations on the RKHS.

Theorem 3. *Suppose that the RKHS Equations 3.10 have a unique solution in* $C([0,\infty); H)$ *for every initial condition* X_0 *in some open ball* $B_r(0) \subseteq \mathbb{X}$. *Then the equilibrium at the origin is Lyapunov stable. Moreover, the state error* $\tilde{x}(t) \to 0$ *as* $t \to \infty$.

Proof. Define the Lyapunov function $V : \mathbb{X} \to \mathbb{R}$ as

$$V \begin{Bmatrix} \tilde{x} \\ \tilde{f} \end{Bmatrix} = \frac{1}{2}\tilde{x}^T P \tilde{x} + \frac{1}{2}(\Gamma \tilde{f}, \tilde{f})_H.$$

This function is norm continuous and positive definite on any neighborhood of the origin since $V(X) \geq \|X\|_{\mathbb{X}}^2$ for all $X \in \mathbb{X}$. For any X, and in particular over the open set $B_r(0)$, the derivative of the Lyapunov function V along trajectories of the system is given as

$$\begin{aligned} \dot{V} &= \frac{1}{2}(\dot{\tilde{x}}^T P \tilde{x} + \tilde{x}^T P \dot{\tilde{x}}) + (\Gamma \tilde{f}, \dot{\tilde{f}})_H \\ &= -\frac{1}{2}\tilde{x}^T Q \tilde{x} + (\tilde{f}, E_x^* B^* P \tilde{x} + \Gamma \dot{\tilde{f}})_H = -\frac{1}{2}\tilde{x}^T Q \tilde{x}, \end{aligned}$$

since $(\tilde{f}, E_x^* B^* P \tilde{x} + \Gamma \dot{\tilde{f}})_H = 0$. Let ϵ be some constant such that $0 < \epsilon < r$. Define $\gamma(\epsilon)$ and Ω_γ according to

$$\gamma(\epsilon) = \inf_{\|X\|_{\mathbb{X}} = \epsilon} V(X),$$

$$\Omega_\gamma = \{X \in \mathbb{X} | V(X) < \gamma\}.$$

We can picture these quantities as shown in Fig. 3.1 and Fig. 3.2. But $\Omega_\gamma = \{X \in$

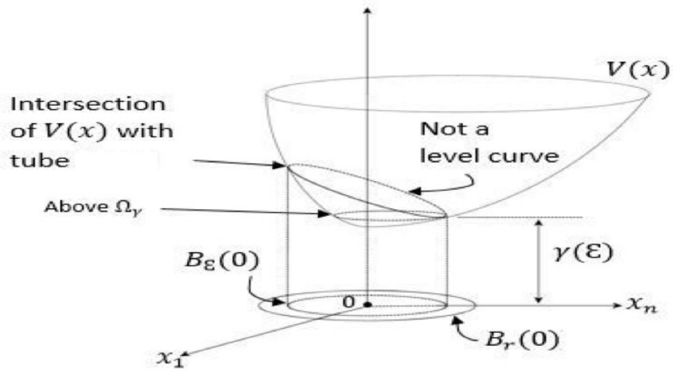


Figure 3.1: Lyapunov function, $V(x)$

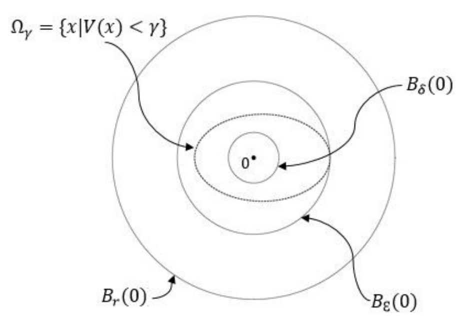


Figure 3.2: Stability of the equilibrium

$\mathbb{X}|V(X) < \gamma\}$ is an open set since it is the inverse image of the open set $(-\infty, \gamma) \subset \mathbb{R}$ under the continuous mapping $V : \mathbb{X} \to \mathbb{R}$. The set Ω_γ therefore contains an open neighborhood of each of its elements. Let $\delta > 0$ be the radius of such an open ball containing the origin with $B_\delta(0) \subset \Omega_\gamma$. Since $\overline{\Omega}_\gamma := \{X \in \mathbb{X}|V(X) \leq \gamma\}$ is a level set of V and V is non-increasing, it is a positive invariant set. Given any initial condition $x_0 \in B_\delta(0) \subseteq \Omega_\gamma$, we know that the trajectory $x(t)$ starting at x_0 satisfies $x(t) \in \overline{\Omega}_\gamma \subseteq \overline{B_\epsilon(0)} \subseteq B_r(0)$ for all $t \in [0, \infty)$. The equilibrium at the origin is stable.

The convergence of the state estimation error $\tilde{x}(t) \to 0$ as $t \to \infty$ can be based on Barbalat's lemma by modifying the conventional arguments for ODE systems. Since $\frac{d}{dt}(V(X(t))) = -\frac{1}{2}\tilde{x}^T(t)Q\tilde{x} \leq 0$, $V(X(t))$ is non-increasing and bounded below by zero. There is a constant $V_\infty := \lim_{t\to\infty} V(X(t))$, and we have

$$V(X_0) - V_\infty = \int_0^\infty \tilde{x}^T(\tau)Q\tilde{x}d\tau \gtrsim \|\tilde{x}\|^2_{L^2((0,\infty);\mathbb{R}^d)}.$$

Since $V(X(t)) \leq V(X_0)$, we likewise have $\|\tilde{x}\|_{L^\infty(0,\infty)} \lesssim V(X_0)$ and $\|\tilde{f}\|_{L^\infty((0,\infty);H)} \lesssim V(X_0)$. The equation of motion enables a uniform bound on $\dot{\tilde{x}}$ since

$$\begin{aligned}
&\|\dot{\tilde{x}}(t)\|_{\mathbb{R}^d} \leq \|A\|\|\tilde{x}(t)\|_{\mathbb{R}^d} + \|B\|\|E_{x(t)}\tilde{f}(t)\|_{\mathbb{R}^d}, \\
&\leq \|A\|\|\tilde{x}(t)\|_{\mathbb{R}^d} + \tilde{C}\|B\|\|\tilde{f}(t)\|_H, \qquad (3.14)\\
&\leq \|A\|\|\tilde{x}\|_{L^\infty((0,\infty);\mathbb{R}^d)} + \tilde{C}\|B\|\|\tilde{f}\|_{L^\infty((0,\infty),H)}.
\end{aligned}$$

Since $\tilde{x} \in L^\infty((0,\infty);\mathbb{R}^d))\cap L^2((0,\infty);\mathbb{R}^d)$ and $\dot{\tilde{x}} \in L^\infty((0,\infty);\mathbb{R}^d)$, we conclude by generalizations of Barbalat's lemma [33] that $\tilde{x}(t) \to 0$ as $t \to \infty$. □

It is evident that Theorem 3 yields results about stability and convergence over the RKHS of the state estimate error to zero that are analogous to typical results for conventional ODE systems. As expected, conclusions for the convergence of the function estimates $\hat{f}$ to f are more difficult to generate, and they rely on persistency of excitation conditions that are suitably extended to the RKHS framework which is discussed in Chapter 5.

3.3 Finite Dimensional Approximations

The governing Equations 3.10 define an evolution on $\mathbb{X} = \mathbb{R}^d \times H$. A system of approximate governing equations can be derived using the weak form of the governing equations that seek $(\tilde{x}, \tilde{f}) \in C([0,T]; \mathbb{X})$ such that

$$\frac{d}{dt}\left(\begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix}, \begin{Bmatrix} y \\ g \end{Bmatrix} \right)_{\mathbb{X}} = \left(\begin{bmatrix} A & BE_{x(t)} \\ -\Gamma^{-1}E^*_{x(t)}B^*P & 0 \end{bmatrix} \begin{Bmatrix} \tilde{x}(t) \\ \tilde{f}(t) \end{Bmatrix}, \begin{Bmatrix} y \\ g \end{Bmatrix} \right)_{\mathbb{X}} \tag{3.15}$$

for all $(y, g) \in \mathbb{X}$. We define approximations in term of the finite dimensional subspaces

$$H_k := \text{span}\{\phi_j\}_{j=1}^k,$$

$$\mathbb{X}_k := \mathbb{R}^d \times H_k.$$

The approximate equations in weak form seek $(\tilde{x}_k, \tilde{f}_k) \in C([0,T]; \mathbb{X}_k)$ such that

$$\frac{d}{dt}(\tilde{x}_k(t), y)_{\mathbb{R}^d} = (A\tilde{x}_k(t), y)_{\mathbb{R}^d} + (BE_{x(t)}\tilde{f}_k(t), y)_{\mathbb{R}^d} \tag{3.16}$$

$$\frac{d}{dt}(\tilde{f}_k(t), g)_H = (-\Gamma E^*_{x(t)}B^*P\tilde{x}_k(t), g)_H \tag{3.17}$$

for all $(y, g) \in \mathbb{X}_k$. These two equations can be rewritten using the orthogonal projector $\Pi_k : H \to H_k$ as

$$\begin{aligned} \frac{d}{dt}(\tilde{x}_k(t), y)_{\mathbb{R}^d} &= (A\tilde{x}_k(t), y)_{\mathbb{R}^d} + (BE_{x(t)}\Pi_k \tilde{f}_k(t), y)_{\mathbb{R}^d} \\ \frac{d}{dt}(\tilde{f}_k(t), g)_H &= \underbrace{(-\Gamma E^*_{x(t)} B^* P \tilde{x}_k(t), \Pi_k g)_H}_{(-\Pi_k \Gamma E^*_{x(t)} B^* P \tilde{x}_k(t), g)_H} \end{aligned}$$

for all $(y, g) \in \mathbb{X}_k$. The approximate equations can then be written in the condensed form as

$$\begin{Bmatrix} \dot{\tilde{x}}_k(t) \\ \dot{\tilde{f}}_k(t) \end{Bmatrix} = \begin{bmatrix} A & BE_{x(t)}\Pi_n^* \\ -\Pi_n \Gamma^{-1} E^*_{x(t)} B^* P & 0 \end{bmatrix} \begin{Bmatrix} \tilde{x}_k(t) \\ \tilde{f}_k(t) \end{Bmatrix}. \tag{3.18}$$

If we recall that $B : \mathbb{R} \to \mathbb{R}^d$, $B^* : \mathbb{R}^d \to \mathbb{R}$, $E_{x(t)} : H \to \mathbb{R}$, $E^*_{x(t)} : \mathbb{R} \to H$, $\Pi_n : H \to H_n$, $\Pi_n^* : H_n \to H$, it is then clear that

$$\begin{bmatrix} A & BE_{x(t)}\Pi_n^* \\ -\Pi_n \Gamma^{-1} E^*_{x(t)} B^* P & 0 \end{bmatrix} : \mathbb{X}_n \to \mathbb{X}_n. \tag{3.19}$$

3.3.1 Convergence of Finite Dimensional Approximations

The governing system in Equations 3.10 constitute a distributed parameter system since the functions $\tilde{f}(t)$ evolve in the infinite dimensional space H. In practice these equations must be approximated by some finite dimensional system. Let $\{H_n\}_{n\in\mathbb{N}_0} \subseteq H$ be a nested sequence of subspaces. Let Π_j be a collection of approximation operators $\Pi_j : H \to H_n$ such that $\lim_{j\to\infty} \Pi_j f = f$ for all $f \in H$ and $\sup_{j\in\mathbb{N}_0} \|\Pi_j\| \leq C$ for a constant $C > 0$. Perhaps the most evident example of such collection might choose Π_j as the H-orthogonal projection for a dense collection of subspaces H_n. It is also common to choose Π_j as a uniformly bounded family of quasi-interpolants [21]. We next construct a finite dimensional approximations $\hat{x}_j$ and $\hat{f}_j$ of the online estimation equations in

$$\dot{\hat{x}}_j(t) = A\hat{x}_j(t) + BE_{x(t)}\Pi_j^* \hat{f}_j(t), \tag{3.20}$$

$$\dot{\hat{f}}_j(t) = \Gamma_j^{-1}\left(BE_{x(t)}\Pi_j^*\right)^* P\tilde{x}_j(t) \tag{3.21}$$

with $\tilde{x}_j := x - \hat{x}_j$. It is important to note that in the above equation $\Pi_j : H \to H_n$, and $\Pi_j^* : H_n \to H$.

Theorem 4. *Suppose that* $x \in C([0,T], \mathbb{R}^d)$ *and that the embedding* $i : H \to C(\Omega)$ *is uniform in the sense that*

$$\|f\|_{C(\Omega)} \equiv \|if\|_{C(\Omega)} \leq C\|f\|_H. \tag{3.22}$$

Then for any $T > 0$*,*

$$\|\hat{x} - \hat{x}_j\|_{C([0,T];\mathbb{R}^d)} \to 0,$$

$$\|\hat{f} - \hat{f}_j\|_{C([0,T];H)} \to 0,$$

as $j \to \infty$*.*

Proof. Define the operators $\Lambda(t) := BE_{x(t)} : H \to \mathbb{R}^d$ for each $t \geq 0$, introduce the measures of state estimation error $\overline{x}_j := \hat{x} - \hat{x}_j$, and define the function estimation error $\overline{f}_j = \hat{f} - \hat{f}_j$. Note that $\tilde{x}_j := x - \hat{x}_j = x - \hat{x} + \hat{x} - \hat{x}_j = \tilde{x} + \overline{x}_j$. The time derivative of the error induced by approximation of the estimates can be expanded as follows:

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\left((\overline{x}_j,\overline{x}_j)_{\mathbb{R}^d}+(\overline{f}_j,\overline{f}_j)_H\right)=(\dot{\overline{x}}_j,\overline{x}_j)_{\mathbb{R}^d}+(\dot{\overline{f}}_j,\overline{f}_j)_H\\
&=(A\overline{x}_j+\Lambda\overline{f}_j,\overline{x}_j)_{\mathbb{R}^d}+\left(\left(\Gamma^{-1}-\Pi_j^*\Gamma_j^{-1}\Pi_j\right)\Lambda^*P\tilde{x},\overline{f}_j\right)_H-\left(\Pi_j^*\Gamma_j^{-1}\Pi_j\Lambda^*P\overline{x}_j,\overline{f}_j\right)_H\\
&\leq C_A\|\overline{x}_j\|^2_{\mathbb{R}^d}+\|\Lambda\|\|\overline{f}_j\|_H\|\overline{x}_j\|_{\mathbb{R}^d}+\|\Gamma^{-1}(I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j)\Lambda^*P\tilde{x}\|_H\|\overline{f}_j\|_H\\
&\qquad+\left\|\Pi_j^*\Gamma_j^{-1}\Pi_j\Lambda^*P\right\|\|\overline{x}_j\|\|\overline{f}_j\|\\
&\leq C_A\|\overline{x}_j\|^2_{\mathbb{R}^d}+\frac{1}{2}\|\Lambda\|\left(\|\overline{f}_j\|^2_H+\|\overline{x}_j\|^2_{\mathbb{R}^d}\right)+\frac{1}{2}\|\Pi_j^*\Gamma_j^{-1}\Pi_j\|\|\Lambda^*\|\|P\|\left(\|\overline{x}_j\|^2_{\mathbb{R}^d}+\|\overline{f}_j\|_H\right)\\
&\qquad+\frac{1}{2}\left(\Gamma^{-1}(I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j)\Lambda^*P\tilde{x}\|_H+\|\overline{f}_j\|^2_H\right)\\
&\leq\frac{1}{2}\|\Gamma^{-1}\|\|\Lambda^*\|\|P\|\|I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j\|^2\|\tilde{x}\|^2_{\mathbb{R}^d}+\left(C_A+\frac{1}{2}\|\Lambda\|+\frac{1}{2}C_B\|\Lambda^*\|\|P\|\right)\|\overline{x}_j\|^2_{\mathbb{R}^d}+\\
&\qquad+\frac{1}{2}\left(\|\Lambda\|+1+\frac{1}{2}C_B\|\Lambda^*\|\|P\|\right)\|\overline{f}_j\|^2_H
\end{aligned}
$$

We know that $\|\Lambda(t)\|=\|\Lambda^*(t)\|$ is bounded uniformly in time from the assumption that H is uniformly embedded in $C(\Omega)$. We next consider the operator error that manifests in the term $(\Gamma^{-1}-\Pi_j^*\Gamma_j^{-1}\Pi_j)$. For any $g\in H$ we have

$$
\begin{aligned}
\|(\Gamma^{-1}-\Pi_j^*\Gamma_j^{-1}\Pi_j)g\|_H&=\|\Gamma^{-1}(I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j)g\|_H\\
&\leq\|\Gamma^{-1}\|\|\left(\Pi_j+(I-\Pi_j)\right)(I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j)g\|_H\\
&\lesssim\|I-\Pi_j\|\|g\|_H.
\end{aligned}
$$

This final inequality follows since $\Pi_j(I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j)=0$ and $\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j\equiv\Gamma\Pi_j^*\left(\Pi_j\Gamma\Pi_j^*\right)^{-1}\Pi_j$ is uniformly bounded. We then can write

$$
\frac{d}{dt}\left(\|\overline{x}_j\|^2_{\mathbb{R}^d}+\|\overline{f}_j\|^2_H\right)\leq C_1\|I-\Gamma\Pi_j^*\Gamma_j^{-1}\Pi_j\|^2+C_2\left(\|\overline{x}_j\|^2_{\mathbb{R}^d}+\|\overline{f}_j\|^2_H\right)
$$

where $C_1, C_2 > 0$. We integrate this inequality over the interval $[0, T]$ and obtain

$$\begin{aligned}\|\overline{x}_j(t)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(t)\|^2_H \leq \|\overline{x}_j(0)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(0)\|^2_H \\ + C_1 T \|I - \Gamma \Pi_j^* \Gamma_j^{-1} \Pi_j\|^2 + C_2 \int_0^T \left(\|\overline{x}_j(\tau)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(\tau)\|^2_H \right) d\tau\end{aligned}$$

We can always choose $\hat{x}(0) = \hat{x}_j(0)$, so that $\overline{x}_j(0) = 0$. If we choose $\hat{f}_j(0) := \Pi_j \hat{f}(0)$ then,

$$\begin{aligned}\|\overline{f}_j(0)\| &= \|\hat{f}(0) - \Pi_j \hat{f}(0)\|_H, \\ &\leq \|I - \Pi_j\|_H \|\hat{f}(0)\|_H.\end{aligned}$$

The non-decreasing term can be rewritten as $C_1 T \|I - \Gamma \Pi_j^* \Gamma_j^{-1} \Pi_j\|^2 \leq C_3 \|I - \Pi_j\|^2_H$.

$$\|\overline{x}_j(t)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(t)\|^2_H \leq C_4 \|I - \Pi_j\|^2_H + C_2 \int_0^T \left(\|\overline{x}_j(\tau)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(\tau)\|^2_H \right) d\tau \tag{3.23}$$

Let $\alpha(t) := C_4 \|I - \Pi_j\|^2_H$ and applying Gronwall's inequality to Equation 3.23, we get

$$\|\overline{x}_j(t)\|^2_{\mathbb{R}^d} + \|\overline{f}_j(t)\|^2_H \leq \alpha(t) e^{C_2 T} \tag{3.24}$$

As $j \to \infty$ we get $\alpha(t) \to 0$, this implies $\overline{x}_j(t) \to 0$ and $\overline{f}_j(t) \to 0$. Therefore the finite dimensional approximation converges to the infinite dimensional states in $\mathbb{R}^d \times H$. □

Lastly, the finite dimensional matrix form of the learning law is given by

$$\dot{\alpha}_i = G_{ij}^{-1} \phi_j(x(t)) B^T P \tilde{x}.$$

where $G_{ij} = (\phi_i, \phi_j)_H$ is the Grammian matrix.

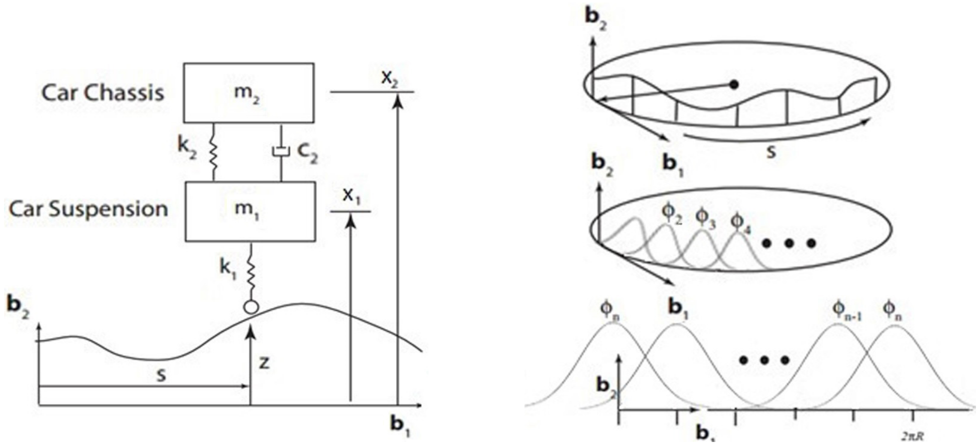


Figure 3.3: Experimental setup and definition of basis functions

3.4 Numerical Simulations

A schematic representation of a quarter car model consisting of a chassis, suspension and road measuring device is shown in Figure 3.3. In this simple model the displacement of the car suspension and chassis are denoted x_1 and x_2, respectively. The arc length s measures the distance along the track that vehicle follows. The equation of motion for the two DOF model has the form

$$M\ddot{x}(t) + C\dot{x}(t) + Kx(t) = Bf(s(t)) \tag{3.25}$$

with the mass matrix $M \in \mathbb{R}^{2\times 2}$, the stiffness matrix $K \in \mathbb{R}^{2\times 2}$, the damping matrix $C \in \mathbb{R}^{2\times 2}$, the control influence vector $b \in \mathbb{R}^{2\times 1}$ in this example. The road profile is denoted by the unknown function $f : \mathbb{R} \to \mathbb{R}$. For simulation purposes, the car is assumed to traverse a circular path of radius R, so that we restrict attention to periodic profiles $f : [0, R] \to \mathbb{R}$. To illustrate the methodology, we first assume that the unknown function, f is restricted to the class of uncertainty mentioned in Equation 1.4 and therefore can be approximated as

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i^* k_{x_i}(\cdot) \tag{3.26}$$

with n as the number of basis functions, α_i^* are the true unknown coefficients to be estimated, and $k_{x_i}(\cdot)$ are basis functions over the circular domain. Hence the state space equation can be written in the form

$$\dot{x}(t) = Ax(t) + B\sum_{i=1}^{n} \alpha_i^* k_{x_i}(s(t)). \tag{3.27}$$

where the state vector $x = [\dot{x}_1, x_1, \dot{x}_2, x_2]$, the system matrix $A \in \mathbb{R}^{4\times4}$, and control influence matrix $B \in \mathbb{R}^{4\times1}$. For the quarter car model shown in Figure 3.3 we derive the matrices,

$$A = \begin{bmatrix} \frac{-c_2}{m_1} & \frac{-(k_1+k_2)}{m_1} & \frac{c_2}{m_1} & \frac{k_2}{m_1} \\ 1 & 0 & 0 & 0 \\ \frac{-c_2}{m_2} & \frac{(k_2)}{m_2} & \frac{-c_2}{m_2} & \frac{-k_2}{m_2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \frac{k_1}{m_1} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Note that if we augment the state to be $\{x_1, x_2, x_3, x_4, s\}$ and append an ODE that specifies $\dot{s}(t)$ for $t \in \mathbb{R}^+$ the Equations 3.27 can be written in the form of Equations 1.1. Then the finite dimensional set of coupled ODE's for the adaptive estimation problem can be written in terms of the plant dynamics, estimator equation, and the learning law which are of the form shown in Equations 3.1, 3.2, and 3.3 respectively.

3.4.1 Synthetic Road Profile Data

The constants in the equation are initialized as follows: $m_1 = 0.5$ kg, $m_2 = 0.5$ kg, $k_1 = 50000$ N/m, $k_2 = 30000$ N/m, $c_2 = 200$ Ns/m, and $\Gamma = 0.001$. The radius of the path traversed is $R = 4$ m, the road profile to be estimated is assumed to have the shape $f(\cdot) = \kappa \sin(2\pi\nu(\cdot))$ with $\nu = 0.04$ Hz and $\kappa = 2$. Adaptive estimation problem is formulated for a synthetic road profile in the RKHS $H = \overline{\{k_x(\cdot) | x \in \Omega\}}$ with $k_x(\cdot) = e^{\frac{-\|x-\cdot\|^2}{2\sigma^2}}$. The radial basis functions, each with standard deviation of $\sigma = 50$, span over the range of 25^o with their centers s_i evenly separated along the arc length. It is important to note that we have chosen a scattered basis that can be located at any collection of centers $\{s_i\}_{i=1}^n \subseteq \Omega$ but the uniformly spaced centers are selected to illustrate the convergence rates. Figure 3.4 shows the finite dimensional

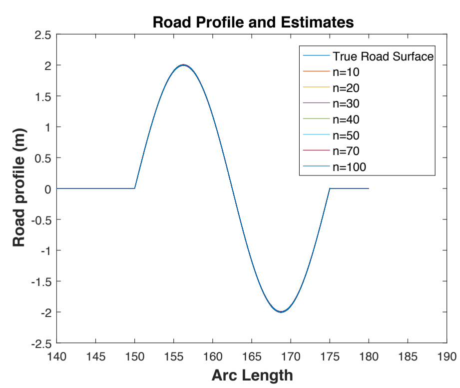


Figure 3.4: Road surface estimates for $n = \{10, 20, \cdots, 100\}$

estimates $\hat{f}$ of the road and the true road surface f for different number of basis kernels ranging from $n = \{10, 20, \cdots, 100\}$. The plots in Figure 3.5 show the rate of convergence of the L^2 error and the $C(\Omega)$ error with respect to the number of basis functions. The *log* along the axes in the figures refer to the natural logarithm unless explicitly specified.

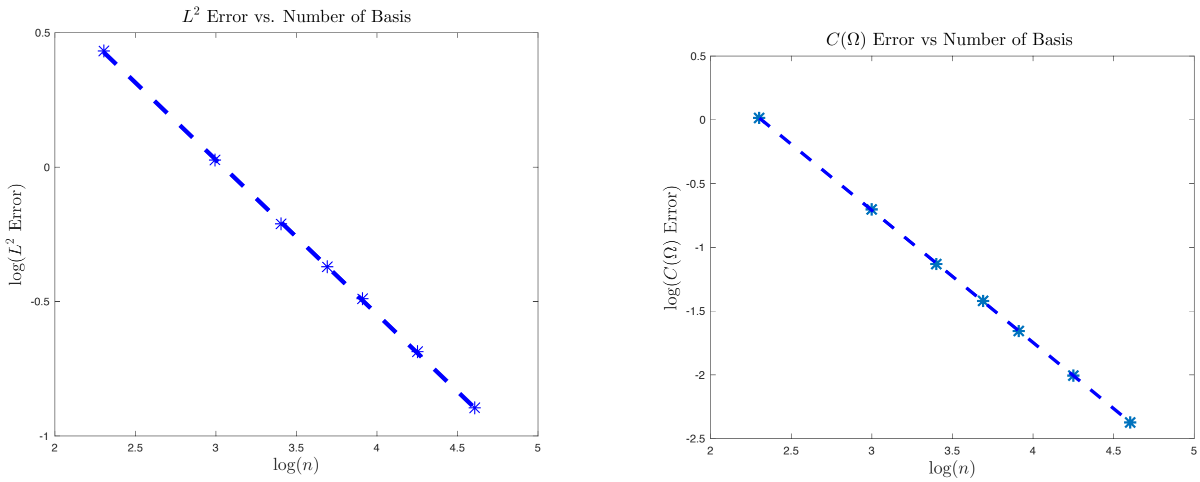
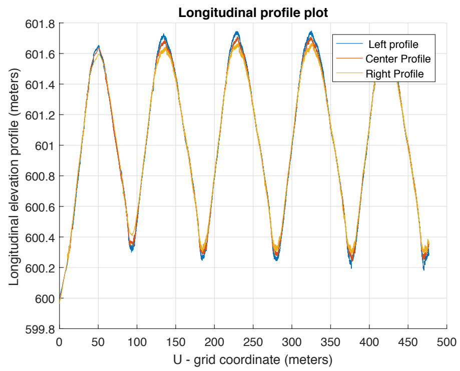


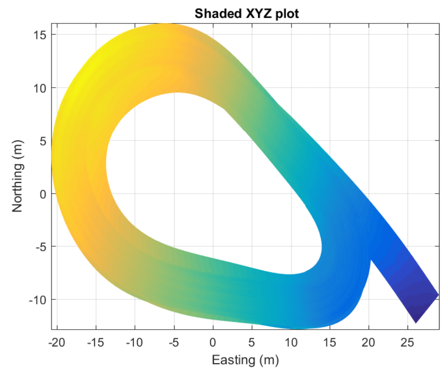
Figure 3.5: Convergence rates using Gaussian kernel for synthetic data

3.4.2 Experimental Road Profile Data

The road profile to be estimated in this subsection is based on the experimental data obtained from the Vehicle Terrain Measurement System shown in Figure 3.6. The constants in the estimation problem are initialized to the same numerical values as in previous subsection. In this section we first study the adaptive estimation problem is formulated in the RKHS $H = \overline{k_x(\cdot)|x \in \Omega\}}$ with $k_x(\cdot) = e^{\frac{-\|x-\cdot\|^2}{2\sigma^2}}$. The radial basis functions, each with standard deviation of $\sigma = 50$, span over the range of with a collection of centers located at $\{s_i\}_{i=1}^n \subseteq \Omega$, evenly separated along the arclength which is measured in meters. This is repeated for kernels defined using B-splines of first order and second order respectively. Figure 3.7 shows the finite dimensional estimates of the road and the true road surface f for a data representing a single lap around the circular track. The finite dimensional estimates $\hat{f}_n$ are plotted for different number of basis kernels ranging from $n = \{35, 50, \cdots, 140\}$ the simulations employ



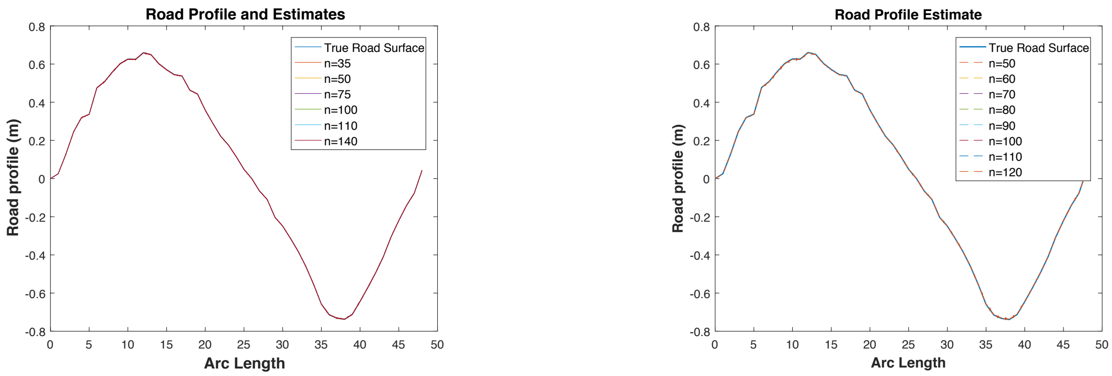
Longitudinal Elevation Profile.



Circular Path followed by VTMS.

Figure 3.6: Experimental Data From VTMS.

the Gaussian kernel as well as the second order B-splines. The x-axes plots the arclength measured in meters while the y-axes plots the corresponding elevation in the road profile measured in meters. The finite dimensional estimates $\hat{f}_n$ of the road profile and the true road profile f for data collected representing multiple laps around the circular track is plotted for the first order B-splines as shown in Figure 3.8. The plots in Figure 3.9 show the rate of convergence of the L^2 error and the $C(\Omega)$ error with respect to number of basis functions. It is seen that the rate of convergence for 2^{nd} order B-Spline is better as compared to other kernels used to estimate in these examples. This corroborates the fact that smoother kernels are expected to have better convergence rates. Also, the condition number of the Grammian matrix varies with n, as illustrated in Table 3.1 and Figure 3.10. This is an important factor to consider when choosing a specific kernel for the RKHS embedding technique since it is well known that the error in numerical estimates of solutions to linear systems is bounded above by the condition number. The implementation of the RKHS embedding method requires



Road surface estimates for Gaussian kernels   Road surface estimate for second-order B-splines

Figure 3.7: Road surface estimates for single lap

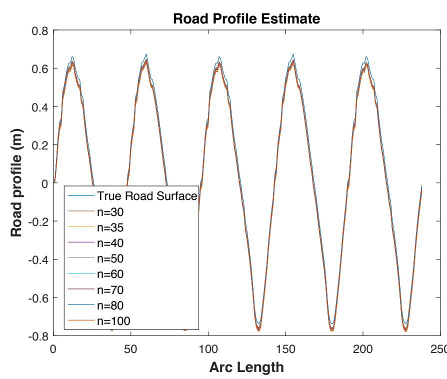


Figure 3.8: Road surface estimate using first-order B-splines

such a solution that depends on the Grammian matrix of the kernel bases at each time step. We see that the condition number of Grammian matrices for exponential basis is $\mathcal{O}(10^{16})$ greater than the corresponding condition numbers for splines. Since the sensitivity of the solutions of linear equations is bounded by the condition numbers, it is expected that the use of exponentials could suffer from a severe loss of accuracy as the dimensionality increases. The development of preconditioning techniques for Grammian matrices constructed from

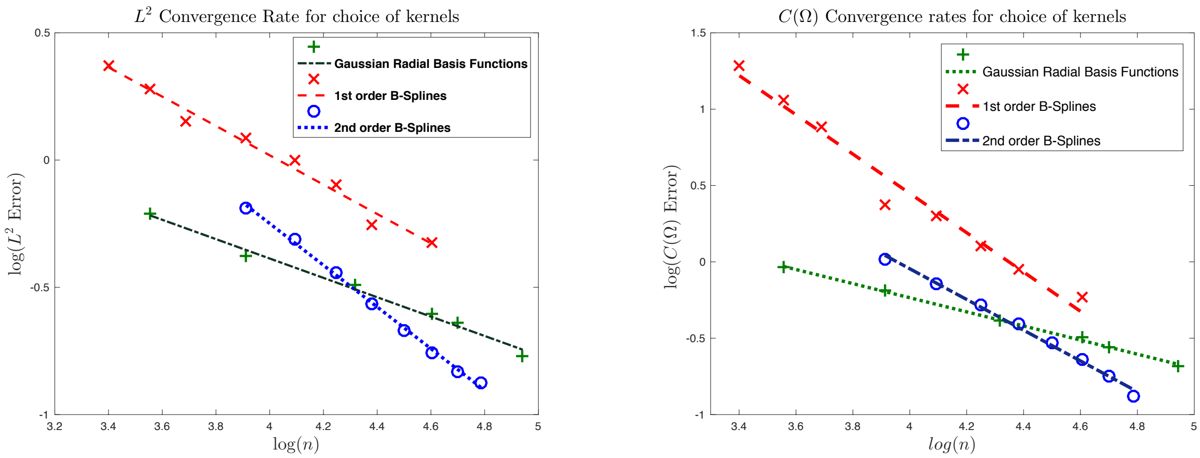


Figure 3.9: Convergence rates for different kernels

radial basis functions to address this problem is an area of active research [60].

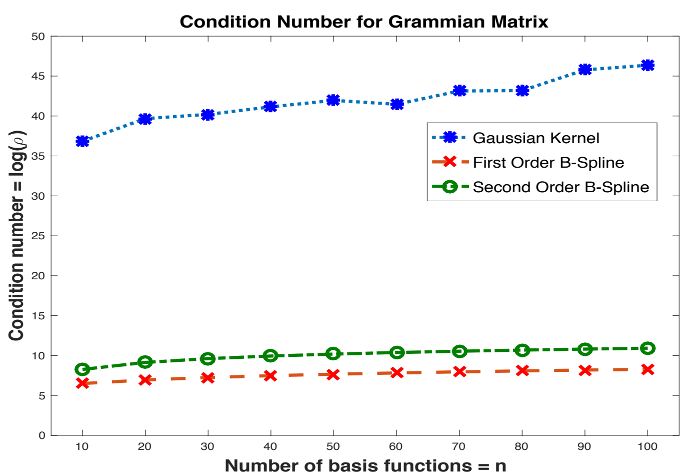


Figure 3.10: Condition Number of Grammian Matrix vs Number of Basis Functions

No. of Basis Functions	Condition No. (First order B-Splines) $\times 10^3$	Condition No.(Second order B-Splines) $\times 10^4$	Condition No.(Gaussian Kernels) $\times 10^{20}$
10	0.6646	0.3882	0.0001
20	1.0396	0.9336	0.0017
30	1.4077	1.5045	0.0029
40	1.7737	2.0784	0.0074
50	2.1388	2.6535	0.0167
60	2.5035	3.2293	0.0102
70	2.8678	3.8054	0.0542
80	3.2321	4.3818	0.0571
90	3.5962	4.9583	0.7624
100	3.9602	5.5350	1.3630

Table 3.1: Condition number of Grammian Matrix vs Number of Basis Functions

Chapter 4

Online Estimation and Control for a Class of History Dependent Systems

This chapter presents sufficient conditions for the convergence of online estimation methods, and the stability of adaptive control strategies for a class of history dependent, functional differential equations. The functional differential equations are constructed using integral operators that depend on distributed parameters. As a consequence, the resulting estimation and control equations are examples of distributed parameter systems whose states and distributed parameters evolve in finite and infinite dimensional spaces, respectively. Existence, and uniqueness are discussed for a class that includes fully actuated robotic systems with history dependent forces in their governing equation of motion. The rates of approximation for the class of history dependent operators are subsequently derived and sufficient conditions to guarantee that finite dimensional approximations of the online estimation equations converge to the solution of the infinite dimensional, distributed parameter system are presented. The convergence and stability of a sliding mode adaptive control strategy for the history dependent, functional differential equations is established using Barbalat's lemma.

4.1 A Class of History Dependent Operators

Methods for modeling history dependent nonlinearities can be formulated using a wide array of approaches. Analytical methods for the study of such systems can be based on ordinary or partial differential equations, differential inclusions, functional differential equations, delay differential equations, or operator theoretic approaches. See references [47],[81],[76]. This paper treats evolution equations that are constructed using a specific class of history dependent operators $\mathcal{H}$ that are defined in terms of integral operators constructed from history dependent kernels. These operators are studied in general in [47] and [81]. The history dependent operators in this dissertation are mappings

$$\mathcal{H} : C([0,T), \mathbb{R}^m) \rightarrow C([0,T), P^*)$$

where the T is the final time of an interval under consideration, m is the number of input functions, q is the number of output functions, P is a Hilbert space of distributed parameters, and its topological dual space is P^*. We limit our consideration to input−output relationships that take the form

$$y(t) = (\mathcal{H}X)(t) \circ \mu \tag{4.1}$$

for each $t \in [0,T)$ where $y(t) \in \mathbb{R}^q$, $(\mathcal{H}X)(t) \in P^*$, and $\mu \in P$. More specifically, we have

$$y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_q(t) \end{bmatrix} = \begin{Bmatrix} (\mathcal{H}_1 X)(t) \circ \mu \\ \vdots \\ (\mathcal{H}_q X)(t) \circ \mu \end{Bmatrix}$$

where for $t \in [0,T)$, each entry is given by $y_i(t) = (\mathcal{H}_i X)(t) \circ \mu$ with $\mu \in P$ and $(\mathcal{H}_i X)(t) \in P^*$.

The definition of $\mathcal{H}$ is carried out in several steps. All of our history dependent operators $\mathcal{H}$ are defined by a superposition or weighting of elementary hysteresis kernels κ_i that are continuous as mappings $\kappa_i : \Delta \times [0,T) \times C[0,T) \to C[0,T)$ for $i = 1, \dots, \ell$. First, define the operator $h_i : C[0,T) \to C([0,T), P^*)$

$$(h_i f)(t) \circ \mu_i := \iint_\Delta \kappa_i(s,t,f)\mu_i(s)ds \tag{4.2}$$

for $\mu_i \in P_i$ and $P = P_1 \times \cdots \times P_\ell$. When we consider problems such as in our motivating examples and numerical case studies, we must construct arrays H of history dependent operators where we define the diagonal matrix

$$(HX)(t) := \begin{Bmatrix} h_1(a(X))(t) & & 0 \\ & \ddots & \\ 0 & & h_\ell(a(X))(t) \end{Bmatrix}$$

for each $t \in [0,T)$ where $a : \mathbb{R}^m \to \mathbb{R}$ is some nonlinear smooth map. Finally, our applications to robotics require that we consider

$$(\mathcal{H}X)(t) = b(X(t))(HX)(t), \tag{4.3}$$

where $b : \mathbb{R}^m \to \mathbb{R}^{q\times\ell}$ is some nonlinear, smooth map. In terms of our entrywise definitions of the input–output mappings, we have

$$y_i(t) := \sum_{j=1}^{l} b_{ij}(X(t))[h_j(a(X))](t) \circ \mu_j \tag{4.4}$$

for $i = 1, \cdots, q$.

In the following discussion, let κ be a generic representation of any of the kernels κ_i for $i = 1, \ldots, \ell$. We choose a typical kernel $\kappa(s,t,f)$ to be a special case of a generalized play operator [81]. We suppose that f is a piecewise linear function on $[0,t]$ with $N+1$ breakpoints numbered $0 = t_0 < t_1 < \cdots < t_N = t$. The output function $t \mapsto \kappa(s,t,f)$, for a fixed $s = (s_1, s_2) \in \Delta \subset \mathbb{R}^2$ and piecewise linear $f : [0,t] \to \mathbb{R}$, is defined by the recursion where $\kappa^{n-1} := \kappa(s, t_{n-1}, f)$ and for $t \in [t_{n-1}, t_n]$ we have

$$\kappa(s,t,f) := \begin{cases} \max\{\kappa^{n-1}, \gamma_{s_2}(f(t))\} & f \text{ increasing on } [t_{n-1}, t_n], \\ \min\{\kappa^{n-1}, \gamma_{s_1}(f(t))\} & f \text{ decreasing on } [t_{n-1}, t_n]. \end{cases}$$

The recursion above depends on the choice of the left and right bounding functions $\gamma_{s_1}, \gamma_{s_2}$ that are depicted in Figure 4.1. These are given in terms of a single ridge function $\gamma : \mathbb{R} \to \mathbb{R}$ with

$$\begin{aligned} \gamma_{s_2}(\cdot) &:= \gamma(\cdot - s_2), \\ \gamma_{s_1}(\cdot) &:= \gamma(\cdot - s_1). \end{aligned} \tag{4.5}$$

As noted in [81], the definition of κ is extended for any $f \in C[0,T)$ by a continuity and density argument.

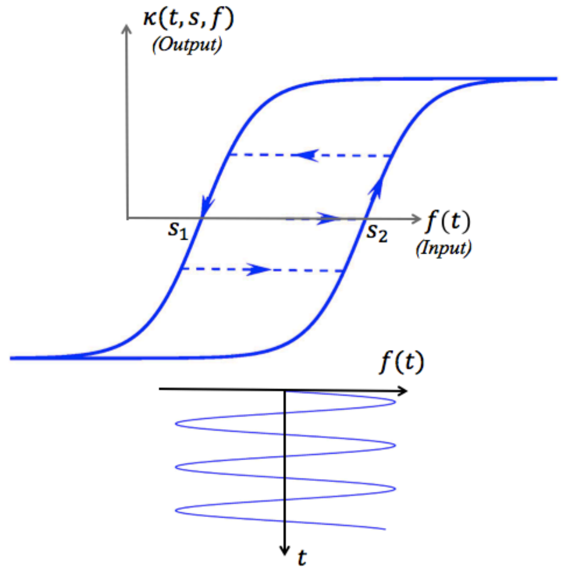


Figure 4.1: Elementary hysteresis kernel $t \to \kappa(s,t,f)$ for fixed $s = (s_1, s_2) \in \mathbb{R}^2$ and piecewise continuous $f : [0,t) \to \mathbb{R}$.

4.1.1 Approximation of History Dependent Operators

The integral operator introduced in Equation 4.2 allows for the representation of complex hysteretic response via the superposition or weighting of fundamental kernels κ_i. These fundamental kernels, each of which has simple input-output relationships, play the role of building blocks for modeling much more complex response characteristics. See [51] for studies of history dependent active materials, [15] for applications that represent nonlinear aerodynamic loading, or Section 4.5 to see an example of richness of this class of models. In this section we emphasize another important feature of this particular class of history dependent operators. We show that relatively simple approximation methods yields bounds on the error in approximation of the history dependent operator that are uniform in time and over the class of functions $\mu \in P$. We now can state our principle approximation result for the class of history dependent operators.

Theorem 5. *Suppose that the function γ that defines the history dependent kernel in Equation 4.5 is a bounded function in $C^{\alpha}(\mathbb{R})$, and define the approximation h_j associated with the grid level j of the history dependent operator h to be*

$$(h_j f)(t) \circ \mu := \iint_{\Delta} \left(\sum_{\ell \in \Gamma_j} 1_{\Delta_{j,\ell}}(s) \kappa(\xi_{j,\ell}, t, f) \right) \mu(s) ds.$$

Then there is a constant $C > 0$ such that

$$|(h_j f)(t) \circ \mu - (hf)(t) \circ \mu| \leq C 2^{-(\alpha+1)j} \tag{4.6}$$

for all $f \in C[0,T]$, $t \in [0,T]$, and $\mu \in P$. If in addition $\mu \in \mathcal{A}_2^{\alpha+1}$, there is a constant $\tilde{C} > 0$ such that

$$|(h_j f)(t) \circ \Pi_j \mu - (hf)(t) \circ \mu| \leq \tilde{C} 2^{-(\alpha+1)j} \tag{4.7}$$

for all $f \in C[0,T]$ and $t \in [0,T]$.

Proof. We first prove the inequality in Equation 4.6. By definition of the operator h, we can write

$$\begin{aligned}
|(h_j f)(t) \circ \mu - (hf)(t) \circ \mu| &\leq \iint_{\Delta} \left| \left(\sum_{k \in \Lambda_j} 1_{\Delta_{j,k}}(s) \kappa(\xi_{j,k}, t, f) - \kappa(s,t,f) \right) \mu(s) \right| ds \\
&\leq \iint_{\Delta} \left| \sum_{k \in \Lambda_j} 1_{\Delta_{j,k}}(s) \left(\kappa(\xi_{j,k}, t, f) - \kappa(s,t,f) \right) \right| |\mu(s)| \, ds
\end{aligned}$$

Since the ridge function γ is a bounded function in $C^{\alpha}(\mathbb{R})$, the output mapping $s \mapsto \kappa(s,t,f)$ is also a bounded function in $C^{\alpha}(\Delta)$ where the Lipschitz constant is independent of $t \in [0,T]$

and $f \in C[0,T]$. Using Proposition 2.5 of [81], we have

$$
\begin{aligned}
&|(h_j f)(t) \circ \mu - (hf)(t) \circ \mu| \\
&\leq \iint_{\Delta} \left| \sum_{k \in \Lambda_j} 1_{\Delta_{j,k}}(s) L \|\xi_{j,k} - s\|^{\alpha} \right| |\mu(s)| \, ds \\
&\leq L \sum_{k \in \Lambda_j} \left(m(\Delta_{j,k}) \left(\frac{\sqrt{2}(\overline{s} - \underline{s})}{2^j} \right)^{\alpha} \right) \iint_{\Delta_{j,k}} |\mu(s)| ds \\
&\leq L \sum_{k \in \Lambda_j} \left(m(\Delta_{j,k}) \left(\frac{\sqrt{2}(\overline{s} - \underline{s})}{2^j} \right)^{\alpha} \right) m^{1/2}(\Delta_{j,k}) \|\mu\|_P \\
&\leq L 2^{2j} \left(\frac{1}{2} \frac{(\overline{s} - \underline{s})^2}{2^{2j}} \left(\frac{\sqrt{2}(\overline{s} - \underline{s})}{2^j} \right)^{\alpha} \right) \left(\frac{1}{2} \frac{(\overline{s} - \underline{s})^2}{2^{2j}} \right)^{1/2} \|\mu\|_P \\
&= C 2^{-(\alpha+1)j} \|\mu\|_P
\end{aligned}
$$

Since we have

$$
|(h_j f)(t) \circ \Pi_j \mu - (hf)(t) \circ \mu| \leq |(h_j f)(t) \circ \Pi_j \mu - (h_j f)(t) \circ \mu| + |(h_j f)(t) \circ \mu - (hf)(t) \circ \mu| \, ,
$$

the second inequality in Equation 4.7 follows from the first Equation 4.6 provided we can show that

$$
|(h_j f)(t) \circ (\mu - \Pi_j \mu)| \leq C 2^{-(\alpha+1)j}
$$

for some constant C. But it is a standard feature of the approximation spaces that if $\mu \in \mathcal{A}_2^{\alpha+1}$, then $\|\mu - \Pi_j \mu\|_P \leq 2^{-(\alpha+1)j} \|\mu\|_{\mathcal{A}_2^{\alpha+1}}$. To see why this is so, suppose that $\mu \in \mathcal{A}_2^{\alpha+1}$.

We have

$$
\begin{aligned}
\|\mu - \Pi_j \mu\|_P^2 &= \sum_{k=j+1}^{\infty} \|(\Pi_k - \Pi_{k-1})\mu\|_P^2 \\
&\leq \sum_{k=j+1}^{\infty} 2^{-2(\alpha+1)k} 2^{2(\alpha+1)k} \|(\Pi_k - \Pi_{k-1})\mu\|_P^2 \\
&\leq 2^{-2(\alpha+1)j} \sum_{k=j+1}^{\infty} 2^{2(\alpha+1)k} \|(\Pi_k - \Pi_{k-1})\mu\|_P^2 \leq 2^{-2(\alpha+1)j} \|\mu\|_{\mathcal{A}_2^{\alpha+1}}^2 .
\end{aligned}
$$

When we apply this to our problem, the upper bound follows immediately

$$
\begin{aligned}
|(h_j f)(t) \circ (\mu - \Pi_j \mu)| &\leq \sup_{t \in [0,T]} \|(h_j f)(t)\|_{P^*} \|\mu - \Pi_j \mu\|_P \\
&\leq C 2^{-(\alpha+1)j},
\end{aligned}
$$

since the boundedness of the ridge function γ implies the uniform boundedness of the history dependent operators $(h_j f)(t)$ over $[0,T]$. □

Theorem 1 can now be used to establish error bounds for input-output maps that have the form in Equation 4.4.

Theorem 6. *Suppose that the hypotheses of Theorem 1 hold. Then we have*

$$
\|(\mathcal{H}X)(t) - (\mathcal{H}_j X)(t)\Pi_j\| \lesssim 2^{-(\alpha+1)j}.
$$

Where $\mathcal{H}$ is defined in Equations 4.1,4.4, and $\mathcal{H}_j$ is defined in Equations 4.8, 4.9 and 4.10 below.

Proof. Recall that for $i = 1 \dots q$ we had

$$
y_i(t) = \sum_{\ell=1\dots l} b_{i\ell}(X(t))(h_\ell(a(X))(t) \circ \mu_\ell.
$$

In matrix form this equation can be expressed as

$$\begin{bmatrix} y_1(t) \\ \vdots \\ y_q(t) \end{bmatrix} = \underbrace{\begin{bmatrix} b_{11}(X(t)) & \dots & b_{1\ell}(X(t)) \\ \vdots & \ddots & \vdots \\ b_{q1}(X(t)) & \dots & b_{q\ell}(X(t)) \end{bmatrix}}_{\mathbb{R}^{q\times l}} \underbrace{\begin{bmatrix} h_1(a(X))(t)\circ\mu_1 \\ \vdots \\ h_\ell(a(X))(t)\circ\mu_\ell \end{bmatrix}}_{(HX)(t)\circ\mu}.$$

It follows that,

$$y(t) = \underbrace{(\mathcal{H}X)(t)\circ\mu}_{\in\,\mathbb{R}^q} = \underbrace{b(X(t))}_{\in\mathbb{R}^{q\times\ell}} \underbrace{(HX)(t)\circ\mu}_{\mathbb{R}^\ell}.$$

By assumption $X \in C([0,T],\mathbb{R}^m)$. The construction of H and $\mathcal{H}$ guarantees that $(HX)(t) \in \mathcal{L}(P,\mathbb{R}^l)$,

$H : C([0,T],\mathbb{R}^m) \to C\left([0,T],\mathcal{L}(P,\mathbb{R}^l)\right)$, and $\mathcal{H} : C([0,T],\mathbb{R}^m) \to C([0,T],\mathbb{R}^q)$. In this proof we denote by $(\mathbb{R}^l, \|.\|_u)$ the norm vector space that endows $\mathbb{R}^l$ with the l^m norm $\|v\|_u := \left(\sum_{i=1}^l |v_i|^u\right)^{\frac{1}{u}}$ for $1 \le u \le \infty$. The normed vector space $(\mathbb{R}^{q\times l}, \|.\|_{s,u})$ denotes the induced operator norm on matrices that map $(\mathbb{R}^l, \|.\|_u)$ into $(\mathbb{R}^q, \|.\|_s)$. Now we define an approximation on the mesh level j of $\mathcal{H}$ to be

$$(\mathcal{H}_j X)(t) = b(X(t)((H_j X)(t)\Pi_j, \tag{4.8}$$

$$(H_j X)(t) = \begin{bmatrix} h_{1,j}(a(X))(t) & 0 & \cdots & 0 \\ 0 & h_{2,j}(a(X))(t) & 0 & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & h_{l,j}(a(X))(t) \end{bmatrix} \tag{4.9}$$

and

$$h_{i,j}(t)\circ\nu = \iint_\Delta \sum_{k\in\Lambda_j} 1_{\Delta_{k,j}}(s)\kappa(\xi_{j,k},t,f)\nu(s)\mathrm{d}s \tag{4.10}$$

for $i = 1 \dots \ell$ and $\nu_i \in P_i$. To simplify the derivation or an error bound for approximation of $\mathcal{H}X(t) \circ \mu$, let $(HX)(t) \circ \mu$ be denoted by $g(t)$. We have assumed that $X \to b(X)$ and $t \to X(t)$ are continuous mappings. Therefore $t \mapsto b(X(t))$ is continuous and on a compact set $[0,T]$, and $b(X(\cdot)) \in C([0,T], \mathbb{R}^{q\times l})$. We therefore by definition have

$$
\begin{aligned}
\|b(X(t))\|_{(\mathbb{R}^{q\times l},\|\cdot\|_{s,u})} &\leq \sup_{\tau\in[0,T]} \|b(X(\tau))\|_{(\mathbb{R}^{q\times l},\|\cdot\|_{s,u})}, \\
&= \|b(X(\cdot))\|_{C([0,T],(\mathbb{R}^{q\times l},\|\cdot\|_{s,u}))}, \\
\|b(X(t))g(t)\|_{(\mathbb{R}^{q},\|\cdot\|_{q})} &\leq \|b(X(\cdot))\|_{C([0,T],(\mathbb{R}^{q\times l},\|\cdot\|_{s,u}))} \|g(t)\|_{(\mathbb{R}^{l},\|\cdot\|_{u})}.
\end{aligned}
$$

with the norms explicitly denoted in the subscript. For $t \in [0,T]$, and applying these definitions,

$$
\begin{aligned}
&\|b(X(t))\left((HX)(t) - (H_jX)(t)\Pi_j\right)\mu\|_{(\mathbb{R}^{q},\|\cdot\|_{s})} \\
&\quad \leq \|b(X(t))\|_{(\mathbb{R}^{q\times l},\|\cdot\|_{s,u})} \|((HX)(t) - (H_jX)(t)\Pi_j) \circ \mu\|_{(\mathbb{R}^{l},\|\cdot\|_{u})} \\
&\quad \leq \|b(X(\cdot))\|_{(C([0,T],(\mathbb{R}^{q\times l},\|\cdot\|_{s,u}))} \|((HX)(t) - (H_jX)(t)\Pi_j) \circ \mu\|_{(\mathbb{R}^{l},\|\cdot\|_{u})}
\end{aligned}
$$

with

$$
\|((HX)(t) - (H_jX)(t)\Pi_j) \circ \mu\|_{(\mathbb{R}^{l},\|\cdot\|_{u})} =
$$

$$
\left\| \begin{bmatrix} h_{1,j}(a(X))(t) - h_{1,j}(a(X))(t)\Pi_j & & & 0 \\ & h_{2,j}(a(X))(t) - h_{2,j}(a(X))(t)\Pi_j & & \\ & & \ddots & \\ 0 & & & h_{l,j}(a(X))(t) - h_{l,j}(a(X))(t)\Pi_j \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_l \end{bmatrix} \right\|_{(\mathbb{R}^{l},\|\cdot\|_{u})}.
$$

Therefore we can now write

$$
\|((HX)(t) - (H_jX)(t)\Pi_j) \circ \mu\|_{\left(\mathbb{R}^{l},\|\cdot\|_{u}\right)} \leq \|((HX)(t) - (H_jX)(t)\Pi_j)\|_{(\mathcal{L}(P,(\mathbb{R}^{l},\|\cdot\|_{u})))} \|\mu\|_{P}.
$$

Hence, recalling Theorem 1 we can now derive the convergence rate

$$
\begin{aligned}
\|((HX)(t)-(H_jX)(t)\Pi_j)\|_{\left(\mathcal{L}\left(P,\left(\mathbb{R}^l,\|\cdot\|_u\right)\right)\right)} &= \sup_{\|\mu\|<1} \|\left((HX)\,(t)-(H_jX)\,(t)\Pi_j\right)\circ\mu\|_{(\mathbb{R}^l,\|\cdot\|_u)} \\
&\leq \sup_{\|\mu\|<1} |((h_{i,j}(a(X)))(t)-(h_{i,j}(a(X)))(t)\Pi_j)\circ\mu| \\
&\leq \hat{C}2^{-(\alpha+1)j}.
\end{aligned}
$$

Therefore we obtain the final bound

$$
\|(\mathcal{H}X)(t)-(\mathcal{H}_jX)(t)\Pi_j)\|_{(\mathbb{R}^l,\|\cdot\|_u)} \lesssim 2^{-(\alpha+1)j} \tag{4.11}
$$

for all $t\in[0,T]$. $\square$

4.2 Well-Posedness: Existence and Uniqueness

The history dependent governing equations studied here are a special case of the more general class of abstract Volterra equations or functional differential equations. A general treatise on abstract Volterra equations can be found in [12], while various generalizations of the theory for the existence and uniqueness of functional differential equations have been given in [28], [69], [39]. The general form of the governing equations we consider have the form

$$
\dot{X}(t) = AX(t) + B((\mathcal{H}X)(t)\circ\mu + u(t)) \tag{4.12}
$$

where the state vector $X(t)\in\mathbb{R}^m$, the control inputs $u(t)\in\mathbb{R}^q$, $A\in\mathbb{R}^{m\times m}=\mathbb{R}^{2n\times 2n}$ is a Hurwitz matrix, and $B\in\mathbb{R}^{m\times q}$ is the control input matrix. We make the following assumptions about the history dependent operators $\mathcal{H}$:

H1) $\mathcal{H} : C([0,\infty),\mathbb{R}^m) \mapsto C([0,\infty),P^*)$

H2) $\mathcal{H}$ is causal in the sense that for all $x, y \in C([0,\infty);\mathbb{R}^m)$,

$$x(\cdot) \equiv y(\cdot) \text{ on } [0,\tau] \implies (\mathcal{H}x)(t) = (\mathcal{H}y)(t) \quad \forall\, t \in [0,\tau].$$

H3) Define the closed set consisting of all continuous functions f that remain within radius r of the initial condition X_0 over the closed interval $[t, t+h]$,

$$\overline{\mathcal{B}}_{[t,t+h],r}(X_0) : \left\{ f \in C([0,h),\mathbb{R}^m) \middle| f(0) = X_0 \text{ and } \|f(s) - X_0\|_{\mathbb{R}^m} \leq r \text{ for } s \in [t, t+h] \right\}$$

for a fixed $X_0 \in \mathbb{R}^m$. For each $t \geq 0$, we assume that there exist $h, r, L > 0$ such that

$$\|(\mathcal{H}X)(s) - (\mathcal{H}Y)(s)\|_{P^*} \leq L\|X - Y\|_{[t,t+h]} \qquad s \in [t, t+h] \tag{4.13}$$

for all $X, Y \in \overline{\mathcal{B}}_{[t,t+h],r}(X_0)$.

Our first result guarantees the existence and uniqueness of a local solution to Equation 1.8, and also describes an important case when such local solutions can be extended to $[0,\infty)$. This theorem can be proven via the existence and uniqueness Theorem 2.3 in [39] for functional delay-differential equations. However, since we are not interested in delay differential equations in this paper, but rather on a highly structured class of integral hysteresis operators, the proof can be much simplified.

Theorem 7. *Suppose that the history dependent operator $\mathcal{H}$ satisfies the hypotheses (H1), (H2), and (H3). Then there is a $\delta > 0$ such that Equation 4.12 has a solution $X \in C([0,\delta),\mathbb{R}^m)$. Suppose the interval $[0,\delta)$ is extended to the maximal interval $[0,\omega) \subset [0,\delta)$ over which such a solution exists. If the solution is bounded, then $[0,\omega) = [0,\infty)$.*

Corollary 1. *Suppose that the history dependent operator $\mathcal{H}$ in Equation 4.12 is defined as in Equation 4.3 and 4.4 in terms of a globally Lipschitz, bounded continuous ridge function $\gamma : \mathbb{R} \to \mathbb{R}$ in Equation 4.5. Then Equation 4.12 has a unique solution $X \in C([0,\infty),\mathbb{R}^m)$ for each $\mu \in P$.*

Proof. For completeness, we outline a simplified version the proof of Theorem 7 for our class of history and parameter dependent equations. As a point of comparison, the reader is urged to compare the proof below to the conventional proof for systems of nonlinear ordinary differential equations, such as in [45]. If we integrate the equations of motion in time, we can define an operator $T : C([0,h),\mathbb{R}^m) \to C([0,h),\mathbb{R}^m)$ from

$$
\begin{aligned}
X(t) &= X_0 + \int_0^t AX(\tau) + B((\mathcal{H}X)(\tau) \circ \mu + u(\tau))\mathrm{d}\tau, \\
X(t) &= (TX)(t),
\end{aligned}
$$

for all $t \in [0,h]$. As introduced in hypothesis (H3), we select $h, r > 0$ and define

$$
\overline{\mathcal{B}}_{[0,h],r}(X_0) := \left\{ X \in C([0,h),\mathbb{R}^m) \middle| X(0) = X_0, \|X_0 - X\|_{[0,\delta]} \leq r \right\}.
$$

such that the local Lipschitz condition in Equation 4.13 holds. Now we consider restricting the equation to a subinterval $[0,\delta] \subseteq [0,h]$, and investigate conditions on T that enable the application of the contraction mapping theorem. We first study what conditions on $\delta > 0$

are sufficient to guarantee that $T : \overline{\mathcal{B}}_{[0,\delta],r}(X_0) \to \overline{\mathcal{B}}_{[0,\delta],r}(X_0)$. We have

$$
\begin{aligned}
\|TX(t) - X_0\|_{\mathbb{R}^m} &\le \int_0^t \|AX(s) + B((\mathcal{H}X(s) \circ \mu + u(s))\|_{\mathbb{R}^m} \mathrm{d}s \\
&\le \int_0^t \Big(\|A\| \|X(s) - X_0\|_{\mathbb{R}^m} + \underbrace{\|AX_0\|_{\mathbb{R}^m}}_{\le \|A\|\|X_0\| = M_A} + \|B\| \underbrace{(\|(\mathcal{H}X)(s) - (\mathcal{H}X_0)(s)\|_{P^*}}_{\le L\|X - X_0\|_{[0,\delta]}} \underbrace{\|\mu\|_P}_{M_\mu} \\
&\qquad + \underbrace{\|(\mathcal{H}X_0)(s)\|_{P^*}}_{\le M_H = \|\mathcal{H}X_0\|_{C([0,\delta],P^*)}} + \underbrace{\|u\|_{C([0,h),\mathbb{R}^p)}}_{\le M_u = \|u\|_{C([0,\delta],\mathbb{R}^q)}}) \Big) \mathrm{d}s \\
&\le ((\|A\| + \|B\| M_\mu L) r + M_A + \|B\| M_T) t \\
&\le ((\|A\| + \|B\| M_\mu L) r + M_A + \|B\| M_T) \delta
\end{aligned}
$$

where $M_T = M_H + M_u$. Now we restrict δ so that

$$
\Big((\|A\| + \|B\| M_\mu L) r + M_A + \|B\| M_T \Big) \delta \le r,
$$

which implies

$$
\delta < \frac{r}{(\|A\| + \|B\| M_\mu L) r + M_A + \|B\| M_T)}.
$$

We thereby conclude that

$$
\|TX(t) - X_0\|_{C([0,h),\mathbb{R}^p)} \le r \quad \text{for} \quad t \in [0, \delta],
$$

and it follows that $T : \overline{\mathcal{B}}_{[0,\delta],r} \to \overline{\mathcal{B}}_{[0,\delta],r}$. Next we study conditions on δ that guarantee that $T : \overline{\mathcal{B}}_{[0,\delta],r} \to \overline{\mathcal{B}}_{[0,\delta],r}$ is a contraction. We compute directly a bound on the difference of the output as

$$
\begin{aligned}
\|(TX)(t) - (TY)(t)\|_{\mathbb{R}^m} &\le \int_0^t \|AX(s) - AY(s) + B((\mathcal{H}X)(s) - (\mathcal{H}Y(s)) \circ \mu)\|_{\mathbb{R}^m} \mathrm{d}s \\
&\le (\|A\| + \|B\| M_\mu L_\mu) \|X - Y\|_{\mathbb{R}^m} \delta.
\end{aligned}
$$

If we choose

$$\delta < min\Big\{h, \frac{r}{(\|A\| + \|B\|M_\mu L)r + M_A + \|B\|M_T)}, \frac{1}{\|A\| + \|B\|M_\mu L}\Big\},$$

it is apparent that T is a contraction that maps the closed set $\overline{\mathcal{B}}_{[0,\delta],r}$ into itself. There is a unique solution in $\overline{\mathcal{B}}_{[0,\delta],r}$ on $[0, \delta]$. □

4.2.1 Numerical Integration

When we seek to control the error in the approximation of our history dependent equation of motion, two distinct types of errors can arise. First, since the history dependent contribution to the equations of motion often cannot be calculated in closed form, some approximation of the history dependent operators must be used. We refer to this as operator approximation error. Even if the history dependent terms are expressed exactly and without error, the resulting equations are a collection of functional differential equations. The usual collection of time stepping integration rules for ordinary differential equations are not directly applicable to such functional differential equations. We employ the strategy first proposed in [53] for numerical time integration of functional differential equations. These techniques assume that the functional differential equations have the form

$$\eta'(t) = F(\eta, t), \quad t \in [t_b, t_c]$$

$$\eta(t) = \eta_a(t), \quad t \in [t_a, t_b]$$

where $\eta_a \in C[t_a, t_b]$ is the initial condition and the functional $F : [a, c] \times [a, c] \to \mathbb{R}^n$. A linear multistep method for functional differential equations constructs the recursion for the solution η_h on a grid having time step h via the formula

$$\begin{aligned} & a_k \eta_h(t_{i+k-1} + rh) + a_{k-1}(r) y_k(t_{i+k-1}) + \cdots \\ & \quad + a_0(r)\eta_h(t_i) - h[b_k(r) F_h(y_h, t_{i+k}) + \cdots \\ & \quad + b_0(r) F_h(y_h, t_i)] = 0 \end{aligned} \tag{4.14}$$

where $r \in [0,1]$, $i = 0, 1, \cdots, N-k$,and $h = (b-a)/N_0$. Three observations should be noted about these approximation scheme:

- solution of this equation yields an extrapolation of the solution η_h on $[t_{i+k-1}, t_{i+k-1} + rh] = [t_{i+k-1}, t_{i+k}]$ since $r \in [0,1]$.
- The discrete equation depends on the history of the discrete approximation through the history dependent functionals $F_h(y_h, t_{i+k})$ and $F_h(y_h, t_i)$.
- The term $F_h(y_h, t_{i+k})$ gives rise to implicit methods in that y_h must be defined on $[t_{i+k-1}, t_{i+k}]$ which is not known at time t_i.

The last observation, in particular, means that the solution of Equation 4.14 involves in general an implicit nonlinear solver over the future history of the solution during the next time step. We generalize the strategy in [53] to employ a predictor-corrector structure. In the predictor phase, we choose the constant a_k and b_k so that $b_k(r) \equiv 0$. Hence the nonlinear dependence on the future solution does not appear. Subsequently, we choose a corresponding corrector in which $b_k(r) \neq 0$. In the correction step we calculate $F_h(y_h^p, t_{i+k})$ in terms of the predictor solution y_h^p. It is defined on the entire future interval. Numerical examples demonstrate that this approach is computationally efficient and accurate. The constants a_k and b_k are selected as described in section 4 of [53] based on conventional linear multistep predictor-corrector integrator schemes.

Integration Error

As discussed in the previous section we use specialized integration schemes to approximate discrete solutions to Volterra Functional Differential Equations [53]. The Adam Bashforth (explicit) Predictor and Adam Moulton (implicit) Corrector is used to numerically solve the functional differential equation. The predictor step computes the approximate value and the corrector step refines the approximation to improve accuracy. The order of accuracy quantifies the rate of convergence for the approximation, The numerical solution is said to be pth order accurate if the error E is proportional to pth power of step size h, i.e $E = Ch^p$. The local truncation error given by $\tau_n = y_n - y_{n-1}$ is of the order $\mathcal{O}(h)^{p+1}$. The rate of convergence for the numerical solutions for two examples are presented below. Example 1 involves history dependent nonlinearity, while Example 2 is a standard integro-differential equation from used to benchmark to validate the numerical solution against the analytical solution.

Example 1 As a first example, we choose to modify the usual harmonic oscillator equation with an additional history dependent term,

$$\begin{bmatrix} \dot{y_1}(t) \\ \dot{y_2}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1+H & 0 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \tag{4.15}$$

where H is the history dependent term and its numerical value depends on the entire history of y_1. For simplicity we choose $H = \kappa(s,t,f)$ in this example. The kernel constructed using methods described in previous section, takes the history of y_1 until current time step as input and gives the value of H as output. We numerically integrate the above equation using predictor-corrector methods. Since we do not have means to derive a closed form solution, we usually rely on the smallest step size numerical solution as a best approximation of the true value. We compare this with solutions having larger step size. In practice h is atleast an order of magnitude greater than the finest step size. The order of accuracy of the numerical solution depends on the smoothness of H and hence the kernel that outputs H. The single step error τ shown in Figure 4.2 is plotted against the step size h on a log-log scale. The slope in the figure corresponds to the order of accuracy and is approximately equal to $p+1$ where p is the order of the integrator. The rate of convergence for the second order predictor corrector from plot is seen to be 2.974 and for the fourth order predictor corrector the rate is 4.9. We compare these results with the rate of convergence in Example 2.

Example 2 To illustrate the utility of predictor corrector algorithm, the following integro-differential equation was numerically solved and compared to its analytical solution.

$$u'(x) + 2u(x) + 5\int_0^x u(t)\mathrm{d}t = \begin{cases} 1 & \mathrm{x} \geq 0 \\ 0 & \mathrm{x} < 0 \end{cases}$$

where $u(0) = 0$. It can be verified that the closed form solution for the above problem is

$$u(x) = \frac{1}{2}e^{-x}\sin 2x.$$

The rate of convergence for the second order predictor-corrector is 3.045 and for the fourth order predictor-corrector is 5.119, which validates the expected rate of convergence $p+1$ for a given pth order numerical integrator.

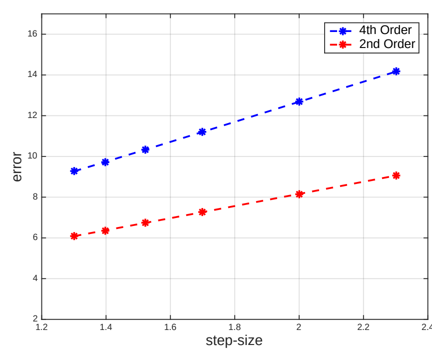


Figure 4.2: Single Step Error vs Step Size for Example 1

4.3 Online Identification

A substantial literature has emerged that treats online estimation problems for linear or nonlinear plants governed by systems of ordinary differential equations. Approaches for

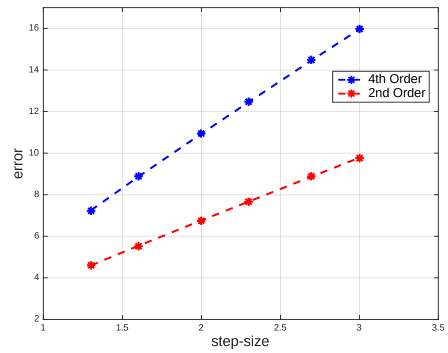


Figure 4.3: Single Step Error vs Step Size for Example 2

these finite dimensional systems that are based on variants of Lyapunov's direct method can be found in any of a number of good texts including, for instance, [58], [71], or [40]. The general strategies that have proven fruitful for such finite dimensional systems have often been extended to classes of systems whose dynamics evolve in an infinite dimensional space: distributed parameter systems. A discussion of the general considerations for identification of distributed parameter systems can be found in [5], for example, while studies that are specifically relevant include [18], [19], [20], and [6].

In this section we adapt the framework introduced in [6] to our class of history dependent, functional differential equations. The approach in [6] assumes that the state equations for the distributed parameter system have first order form, and they are cast in terms of a nonlinear, parametrically dependent bilinear form that is coercive. The resulting equations that govern the error in state and in distributed parameter estimates is a nonlinear function of the state trajectory of the plant. In contrast, a similar strategy yields error equations that depend nonlinearly on the history of the state trajectory.

The general online estimation problem discussed in this section assumes that we observe the value of the state $X(t) \in \mathbb{R}^m$ at each time $t \geq 0$ that depends on some unknown distributed parameter $\mu \in P$, and subsequently use the observed state to construct estimates $\hat{X}$ of the states and $\hat{\mu}$ of the distributed parameters. We construct online estimates that evolve on the state space $\mathbb{R}^m \times P$ according to the time-varying, distributed parameter system equations

$$\begin{aligned} \dot{\hat{X}}(t) &= A\hat{X}(t) + B\left(\left(\mathcal{H}X\right)(t) \circ \hat{\mu}(t) + u(t)\right), \\ \dot{\hat{\mu}}(t) &= -\left(B(\mathcal{H}X)(t)\right)^* \hat{X}(t), \end{aligned} \tag{4.16}$$

for $t \geq 0$ where the initial conditions are $\hat{X}_0 := X_0$, $\hat{\mu}(0) := \mu_0$. In these equations, we denote the adjoint operator L^* for any bounded linear operator L. These equations can be understood as incorporating a natural choice of a parameter update law. The learning law above can be interpreted as generalization of the conventional gradient update law that features prominently in approaches for finite dimensional systems [40] and that has been extended to distributed parameter systems in [6]. It is immediate that the error in estimation of the states $\tilde{X} := X - \hat{X}$ and in the distributed parameters $\tilde{\mu} := \mu - \hat{\mu}$ satisfy the homogeneous system of equations

$$\begin{Bmatrix} \dot{\tilde{X}}(t) \\ \dot{\tilde{\mu}}(t) \end{Bmatrix} = \begin{bmatrix} A & B(\mathcal{H}X)(t) \\ -\left(B(\mathcal{H}X)(t)\right)^* & 0 \end{bmatrix} \begin{Bmatrix} \tilde{X}(t) \\ \tilde{\mu}(t) \end{Bmatrix}.$$

4.3.1 Approximation of the Estimation Equations

The governing system in Equations 4.16 constitute a distributed parameter system since the functions $\hat{\mu}(t)$ evolve in the infinite dimensional space P. In practice these equations must be approximated by some finite dimensional system. We define $\tilde{X}_j = \hat{X} - \hat{X}_j$ and $\tilde{\mu}_j = \hat{\mu} - \hat{\mu}_j$ where $\tilde{X}_j$ and $\tilde{\mu}_j$ express approximation errors due to projection of solutions in $\mathbb{R}^m \times P$ to a finite dimensional approximation space. We construct a finite dimensional approximation of the the online estimation equations using the results of Section 4.1.1 and obtain

$$\dot{\hat{X}}_j(t) = A\hat{X}_j(t) + B\left((\mathcal{H}_j X)(t)\Pi_j \circ \hat{\mu}_j(t) + u(t)\right), \tag{4.17}$$

$$\dot{\hat{\mu}}_j(t) = -\left(B(\mathcal{H}_j X)(t)\Pi_j\right)^* X(t). \tag{4.18}$$

Theorem 8. *Suppose that the history dependent operator $\mathcal{H}$ in Equation 4.12 is defined as in Equations 4.3 and 4.4 in terms of a globally Lipschitz, bounded continuous ridge function $\gamma : \mathbb{R} \to \mathbb{R}$ in Equation 4.5. Then for any $T > 0$, we have*

$$\|\hat{X} - \hat{X}_j\|_{C([0,T],\mathbb{R}^m)} \to 0,$$

$$\|\hat{\mu} - \hat{\mu}_j\|_{C([0,T],P)} \to 0,$$

as $j \to \infty$.

Proof. Define the operators $G(t) : P \to \mathbb{R}^m$ and $G_j(t) : P \to \mathbb{R}^m$ for each $t \geq 0$ as

$$G(t) := B(\mathcal{H}X)(t),$$

$$G_j(t) := B(\mathcal{H}_j X)(t)\Pi_j.$$

The time derivative of the error in approximation can be expanded as follows:

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\Big((\tilde{X}_j,\tilde{X}_j)_{\mathbb{R}^m}+(\tilde{\mu}_j,\tilde{\mu}_j)_P\Big) &= (\dot{\tilde{X}}_j,\tilde{X}_j)_{\mathbb{R}^m}+(\dot{\tilde{\mu}}_j,\tilde{\mu}_j)_P \\
&= (A\tilde{X}_j+G\hat{\mu}-G_j\hat{\mu}_j,\tilde{X}_j)_{\mathbb{R}^m}+(-(G-G_j)^*X,\tilde{\mu}_j)_P \\
&= (A\tilde{X}_j,\tilde{X}_j)_{\mathbb{R}^m}+\Big((G-G_j)\hat{\mu},\tilde{X}_j\Big)_{\mathbb{R}^m}+\Big(G_j(\hat{\mu}-\hat{\mu}_j),\tilde{X}_j\Big)_{\mathbb{R}^m} \\
&\quad -((G-G_j)\tilde{\mu}_j,X)_{\mathbb{R}^m} \\
&\leq c(\tilde{X}_j,\tilde{X}_j)_{\mathbb{R}^m}+\|(G-G_j)\hat{\mu}\|_{\mathbb{R}^m}\|\tilde{X}_j\|_{\mathbb{R}^m} \\
&\quad +\|G_j\|_{\mathcal{L}(P,\mathbb{R}^m)}\|\tilde{\mu}_j\|_P\|\tilde{X}_j\|_{\mathbb{R}^m}+\|G-G_j\|_{\mathcal{L}(P,\mathbb{R}^m)}\|\tilde{\mu}_j\|_P\|X\|_{\mathbb{R}^m}.
\end{aligned}
$$

We will next use a common inequality that can be derived from two applications of the triangle inequality. We have

$$
\begin{aligned}
(a+b,a+b) &= (a,a)+2(a,b)+(b,b)\geq 0, \\
(a-b,a-b) &= (a,a)-2(a,b)+(b,b)\geq 0.
\end{aligned}
$$

We conclude from this pair of inequalities that

$$
|(a,b)|\leq\frac{1}{2}\left(\|a\|^2+\|b\|^2\right).
$$

The specific form that we apply this theorem is written as

$$
|(a,b)|=|(\sqrt{\epsilon}a,\frac{1}{\sqrt{\epsilon}}b)|\leq\epsilon\frac{\|a\|^2}{2}+\frac{1}{\epsilon}\frac{\|b\|^2}{2}. \tag{4.19}
$$

We apply the inequality in Equation 4.19 to each term in which $\tilde{\mu}_j$ and $\tilde{X}_j$ appear in a product.

$$\frac{1}{2}\frac{d}{dt}\left(\|\tilde{X}_j\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j\|^2_{P)}\right) \leq c\|\tilde{X}_j\|^2_{\mathbb{R}^m} + \frac{1}{2a}\|(G - G_j)\hat{\mu}\|^2_{\mathbb{R}^m} + \frac{a}{2}\|\tilde{X}_j\|^2_{\mathbb{R}^m}$$
$$+ \frac{1}{2b}\|G_j\tilde{\mu}_j\|^2_{\mathbb{R}^m} + \frac{b}{2}\|\tilde{X}_j\|^2_{\mathbb{R}^m} + \frac{1}{2c}\|\tilde{\mu}_j\|^2_P$$
$$+ \frac{c}{2}\|G - G_j\|^2_{\mathcal{L}(P,\mathbb{R}^m)}\|X\|^2_{\mathbb{R}^m}.$$

We will employ the integral form of Gronwall's Inequality to obtain our final convergence result. Many forms of Gronwall's Inequality exist, and we will use a particularly simple version. See Section 3.3.4 in [40]. If the piecewise continuous function f satisfies the inequality

$$f(t) \leq \alpha(t) + \int_0^t \beta(s)f(s)ds$$

with some piecewise continuous functions α, β where α is nondecreasing, then

$$f(t) \leq \alpha(t)e^{\int_0^t \beta(s)ds}.$$

Then

$$\frac{d}{dt}\left(\|\tilde{X}_j\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j\|^2_P\right) \leq c\|G - G_j\|^2_{\mathcal{L}(P,\mathbb{R}^m)}\|X\|^2_{\mathbb{R}^m} + (2c + a + b)\|\tilde{X}_j\|^2_{\mathbb{R}^m}$$
$$+ \left(\frac{1}{c} + \frac{1}{b}\|G_j^*G_j\|\right)\|\tilde{\mu}_j\|^2_P + \frac{1}{a}\|G - G_j\|^2_{\mathcal{L}(P,\mathbb{R}^m)}\|\hat{\mu}\|^2_P.$$

We integrate this inequality in time from 0 to t to obtain

$$\|\tilde{X}_j(t)\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j(t)\|^2_P \leq \|\tilde{X}_j(0)\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j(0)\|^2_P$$
$$+ \int_0^t c\|G(s) - G_j(s)\|^2_{\mathcal{L}(P,\mathbb{R}^m)}\|X(s)\|^2_{\mathbb{R}^m}ds$$
$$+ \int_0^t \left\{(2c + a + b)\|\tilde{X}(s)\|^2_{\mathbb{R}^m} + \left(\frac{1}{c} + \frac{1}{b}\|G_j^*(s)G_j(s)\|\right)\|\tilde{\mu}_j\|^2_P\right\}ds$$
$$+ \int_0^t \frac{1}{a}\|G(s) - G_j(s)\|^2_{\mathcal{L}(P,\mathbb{R}^m)}\|\hat{\mu}\|^2_P ds.$$

Choose $a, b > 0$ large enough so that $(2c + a + b) > 0$ and set $\gamma > 1$. If we define

$$\gamma := \max\left(2c + a + b, \frac{1}{c} + \frac{\eta}{b} \sup_{s\in[0,T]} \|G^*(s)G(s)\|, 1\right),$$

$$\lambda_j(t) := \|(I - \Pi_j)\hat{\mu}(0)\|_P + \int_0^t \|G(s) - G_j(s)\|^2_{\mathcal{L}(P,\mathbb{R}^m)} \left(c\|X\|^2_{\mathbb{R}^m} + \frac{1}{a}\|\hat{\mu}\|^2_P\right) ds,$$

then the inequality can be written as

$$\|\tilde{X}_j(t)\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j(t)\|^2_P \le \lambda_j(t) + \gamma \int_0^t \left(\|\tilde{X}_j(s)\|^2_{\mathbb{R}^m} + \|\tilde{\mu}_j(s)\|^2_P\right) ds.$$

Gronwall's Inequality now completes the proof of the theorem. $\square$

we further investigate $\lambda_j(t)$ to derive the convergence rate for the approximate states and parameters evolving associated with level j resolution. According to the convergence results obtained in Theorem 1 we have $\|G(s) - G_j(s)\|_{\mathcal{L}(P,\mathbb{R}^m)} = \|B(\mathcal{H}X)(t) - B(\mathcal{H}_jX)(t)\Pi_j\| \le C_2 2^{-(\alpha+1)j}$. Therefore, $\|G(s) - G_j(s)\|^2 \le C_2^2 2^{-(\alpha+1)2j}$. It then follows that

$$\begin{aligned}
\lambda_j(t) &= \|(I - \Pi_j)\hat{\mu}(0)\|_P + \int_0^t 2^{-(\alpha+1)2j} \left(c\|X\|^2_{\mathbb{R}^m} + \frac{1}{a}\|\hat{\mu}\|^2_P\right) \mathrm{d}s \\
&\le \|(I - \Pi_j)\|\|\hat{\mu}(0)\|_P + 2^{-(\alpha+1)2j} \left(c\|X\|^2_{\mathbb{R}^m} + \frac{1}{a}\|\hat{\mu}\|^2_P\right) t.
\end{aligned}$$

then we can conclude that $\lambda_j(t) < \mathcal{O}(2^{-(\alpha+1)j}) \quad \text{for} \quad t \in [0, C_3 2^{(\alpha+1)j}]$.

4.4 Adaptive Control Synthesis

In order to estimate the function μ that weighs the contribution of history dependent kernels to the equations of motion, we first map it to an n-dimensional subspace of square integrable functions using a projection operator $\Pi^n : P \mapsto P^n$. Let

$$\dot{X} = AX + B((\mathcal{H}X) \circ (\mu - \hat{\mu}) + v) \tag{4.20}$$

be the governing equation of a robotic system after applying a feedback linearization control signal as mentioned in Equation 1.7 with $u = v - (\mathcal{H}X) \circ \hat{\mu}$. We substitute $\mu = \Pi^n \mu + (I - \Pi^n)\mu$ and write

$$\dot{X} = AX + B((\mathcal{H}X) \circ (\Pi^n \mu - \hat{\mu}) + v) + B((\mathcal{H}X) \circ (I - \Pi^n)\mu). \tag{4.21}$$

Finally, by replacing $d = \{(\mathcal{H}X)(I - \Pi^n) \circ \mu\}$ we obtain

$$\dot{X} = AX + B((\mathcal{H}X) \circ (\Pi^n \tilde{\mu}) + v + d(t)), \tag{4.22}$$

where

$$\dot{\tilde{\mu}} = -((\mathcal{H}X)\Pi^n)^* B^T \mathcal{P} X. \tag{4.23}$$

Theorem 9. *Suppose the state equations have the form of Equation 1.7 and the matrix $\mathcal{P}$ is a symmetric positive definite solution of the Lyapunov equation $A^T\mathcal{P} + \mathcal{P}A = -Q$ where $Q > 0$. Then by employing the update law $\dot{\tilde{\mu}} = -((\mathcal{H}X)\Pi^n)^* B^T \mathcal{P} X$, the control signal*

$$v(t) = \begin{cases} -k\frac{B^T\mathcal{P}X}{\|B^T\mathcal{P}X\|}, & \text{if } \|B^T\mathcal{P}X\| \geq \epsilon \\ \\ -\frac{k}{\epsilon}B^T\mathcal{P}X, & \text{if } \|B^T\mathcal{P}X\| < \epsilon \end{cases} \tag{4.24}$$

with $k > \|d\|$ drives the tracking error dynamics of the closed loop system so that it is uniformly ultimately bounded and its norm is eventually $O(\epsilon)$.

Proof. We choose the Lyapunov function

$$V = \frac{1}{2}X^T\mathcal{P}X + \frac{1}{2}(\tilde{\mu}, \tilde{\mu})_P \tag{4.25}$$

where $\mathcal{P}$ is the solution of the Lyapunov equation $A^T\mathcal{P} + \mathcal{P}A = -Q$. The derivative of the Lyapunov function V along the closed loop system trajectory is

$$
\begin{aligned}
\dot{V} &= \frac{1}{2}(\dot{X}^T\mathcal{P}X + X^T\mathcal{P}\dot{X}) + \left(\dot{\tilde{\mu}}, \tilde{\mu}\right)_p \\
&= \frac{1}{2}\big(AX + B((\mathcal{H}X)\circ(\Pi^n\tilde{\mu}) + v + d)\big)^T\mathcal{P}X + X^T\mathcal{P}(AX + B((\mathcal{H}X)\circ(\Pi^n\tilde{\mu}) + v + d)\big) + \left(\dot{\tilde{\mu}}, \mu\right)_P \\
&= \frac{1}{2}X^T(A^T\mathcal{P} + \mathcal{P}A)X + X^TPB(v+d) + X^T\mathcal{P}B\big((\mathcal{H}X)\circ(\Pi^n\tilde{\mu})\big) + \left(\dot{\tilde{\mu}}, \mu\right)_P \\
&= -\frac{1}{2}X^TQX + X^T\mathcal{P}B(v+d) + \left(\dot{\tilde{\mu}} + ((\mathcal{H}X)\Pi^n)^*B^T\mathcal{P}X, \tilde{\mu}\right)_P \\
&= -\frac{1}{2}X^TQX + X^T\mathcal{P}B(v+d).
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
\dot{V} &\leq -\frac{1}{2}X^TQX + X^T\mathcal{P}B(v+d), \\
&\leq -\frac{1}{2}X^TQX + \begin{cases} X^T\mathcal{P}B\left(-k\frac{B^T\mathcal{P}X}{\|B^T\mathcal{P}X\|} + d\right) & \text{if } \|B^T\mathcal{P}X\| \geq \epsilon \\ X^T\mathcal{P}B\left(-\frac{k}{\epsilon}B^T\mathcal{P}X + d\right) & \text{if } \|B^T\mathcal{P}X\| \leq \epsilon \end{cases}, \\
&\leq -\frac{1}{2}X^TQX + \begin{cases} -\left(k - \|d\|\right)\|B^T\mathcal{P}X\| & \text{if } \|B^T\mathcal{P}X\| \geq \epsilon \\ \epsilon k & \text{if } \|B^T\mathcal{P}X\| \leq \epsilon \end{cases} \\
&\leq -\frac{1}{2}X^TQX + \epsilon k.
\end{aligned}
$$

By Theorem 4.18 in [45] we conclude that there is a $\bar{T} > 0$ and $t > 0$ such that $\|X(t)\| \leq \bar{C}\epsilon$ for all $t \geq \bar{T}$. □

4.5 Numerical Simulations

In this section, first the operator approximation error bound presented in Theorem 1 is validated. Then a single wing section with a leading and trailing edge flaps is modeled and the proposed sliding mode adaptive controller presented in Theorem 4 is implemented on this model. The stability of the closed loop system and convergence of the closed-loop system trajectories to the equilibrium point is illustrated.

4.5.1 Operator Approximation Error

In order to show that Equation 4.7 holds, start by choosing a function $\mu(s)$ over Δ and then calculate $(h_j f)(t) \circ \mu_j$ for different levels of refinement. Since the computation of $(hf)(t) \circ \mu$ exactly is numerically infeasible, $J \gg j$ is chosen as the finest level of refinement in our simulation. According to Theorem 1,

$$|(h_J f)(t) \circ \mu_J - (hf)(t) \circ \mu| \leq C_J 2^{-(\alpha+1)J},$$

and for $j \ll J$ it is seen that

$$|(h_j f)(t) \circ \mu_j - (hf)(t) \circ \mu| \leq C_j 2^{-(\alpha+1)j}.$$

Assuming $C = \max\{C_j, C_J\}$ and using the triangle inequality,

$$\begin{aligned} &|(h_J f)(t) \circ \mu_J - (h_j f)(t) \circ \mu_j| \\ &\qquad \leq C(2^{-(\alpha+1)J} + 2^{-(\alpha+1)j}) \end{aligned} \tag{4.26}$$

Therefore, given the weights μ_J for the finest level of refinement J, we can evaluate $\mu_j = \Pi_j \mu_J$ and numerically verify Equation 4.26. Figure 4.4 shows the simulation results for $J = 7$ and

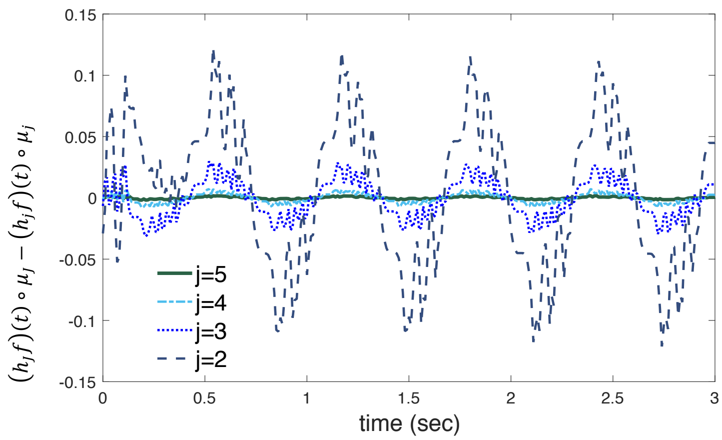


Figure 4.4: Error for different resolution simulations, $J = 7$

$j = 2, 3, 4, 5$. The error term attenuates with increasing j. In order to investigate the rate of attenuation, we evaluate constant C for different levels of refinements. As shown in figure 4.5, C is approximately constant with respect to j which agrees with the result from Equation 4.26.

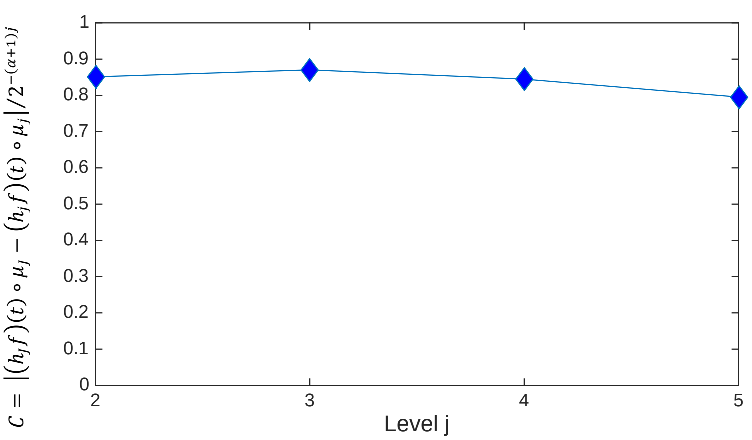


Figure 4.5: C for different level j refinement simulations

4.5.2 Online Identification for a Wing Model

The reformatted governing equations of the system take the form of Equation 1.6 where $Q_a(t,\mu)$ is the vector of generalized history dependent aerodynamic loads. The dynamic equation of the system can be written in the form of Equation 1.8, where the history dependent term $M^{-1}(q)Q_a(t,\mu)$ is rewritten in terms of a history dependent operator $(\mathcal{H}X)(t)$ acting on the distributed parameter function μ. The history dependent operator includes a family of fixed history dependent kernels and the distributed parameters μ act as a weighting vector that determines the contribution of a specific history dependent kernel to the overall history dependent operator. We perform an offline identification based on

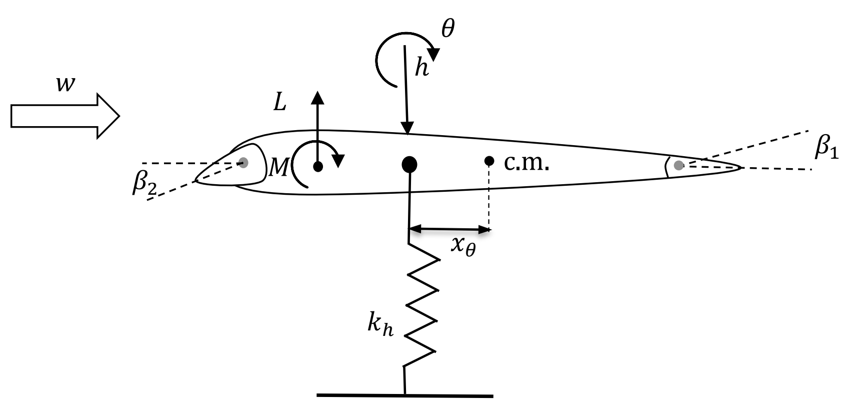


Figure 4.6: Prototypical model for a wing section

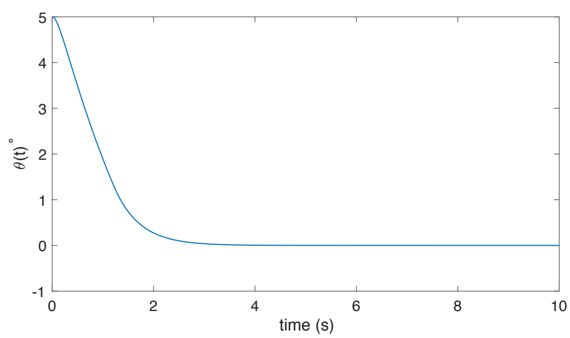
a set of experimental data collected from a wind tunnel experiments or CFD simulations. These define a nominal model for the history dependent aerodynamic loads that appear in the governing equations of the system. We can exploit the model in the numerical simulations to perform an online estimation for the history dependent aerodynamics and adaptive control of a simple wing model. The details of offline identification of history dependent aerodynamics follow the steps explained in [15].

The model developed in Figure 4.6 is chosen to validate our proposed adaptive sliding mode controller where w is the velocity of wind, k_h is spring constant in plunge, k_θ is a spring constant in pitch, θ is the pitch angle, h is the plunge displacement, c_θ and c_h are viscous damping coefficients, m and I_θ are the mass and moment of inertia and, x_θ is the non-dimensionalized distance between center of mass and the elastic axis. Finally, L and M are lift and moment generated by the leading and trailing edge flaps. The angles β_1 and β_2 define the rotation of the trailing edge and leading edge flaps respectively. The dynamic equations of the wing model is derived in the Appendix C as

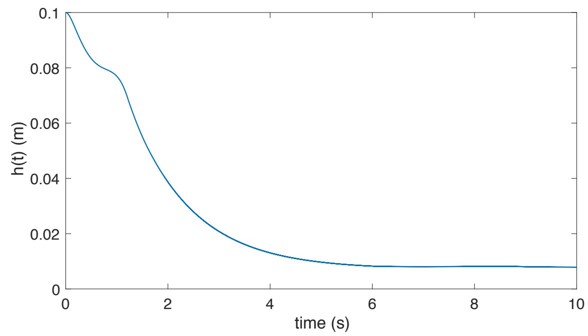
$$\begin{bmatrix} m & mx_\theta \\ mx_\theta & mx_\theta^2 + I_\theta \end{bmatrix} \begin{Bmatrix} \ddot{h} \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} c_h & 0 \\ 0 & c_\theta \end{bmatrix} \begin{Bmatrix} \dot{h} \\ \dot{\theta} \end{Bmatrix} + \begin{bmatrix} k_h & 0 \\ 0 & k_\theta \end{bmatrix} \begin{Bmatrix} h \\ \theta \end{Bmatrix} = \begin{Bmatrix} L \\ 0 \end{Bmatrix} + \begin{Bmatrix} f_1(\beta_1, \beta_2) \\ f_2(\beta_1, \beta_2) \end{Bmatrix}. \tag{4.27}$$

It is assumed the aerodynamic moment M to be zero and the distance x_a between the aerodynamic center A and support hinge point to be negligible to simplify the simulation. The unsteady aerodynamic lift is $L = Q_a(t, \mu)$ where $Q_a(t, \mu) = (\mathcal{H}X) \circ \mu$ reflects the history dependent nature of aerodynamic loads. We rewrite Equation 4.27 to achieve the standard form presented in Equation 1.6.

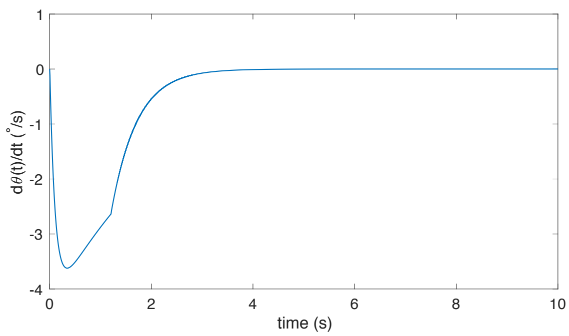
The adaptive controller presented in Theorem 4 is composed of two parts. The first part compensates for the flutter generated by the history dependent aerodynamic forces through online identification of the aerodynamics. The second part employs an sliding mode controller to compensate for modeling errors. It is important to note that numerical time integration of the evolution equations must accommodate history dependent terms. Since the dynamics of such systems are given via functional differential equations, the ordinary integration rules are not directly applicable. Figure 4.7 shows the simulation results for the case where $\epsilon = 0.01$ and $t_h = 0.001$. The system response eventually enters in a ϵ neighborhood of the sliding manifold. However, as depicted in the figure, a chattering behavior occurs in the control signal and system trajectories. We trace this behavior back to the integration error induced by the size of time step. When we increase ϵ or reduce the integration time step, the control signal and system trajectories become smooth. The simulation results for $\epsilon = 0.01$ and $t_h = 0.0005$ are depicted in Figure 4.8. The system trajectories converge to a $O(\epsilon)$neighborhood of zero when the control signals are relatively smooth. Also, Figure 4.9 shows the case when $\epsilon = 0.1$ and $t_h = 0.001$. The convergence rate of the signals to an $O(\epsilon)$ neighborhood of zero is slower but the results do not show any chattering. Therefore, the proposed smooth sliding mode adaptive controller proves to be effective to identify and compensate for the unknown history dependent aerodynamic forces.



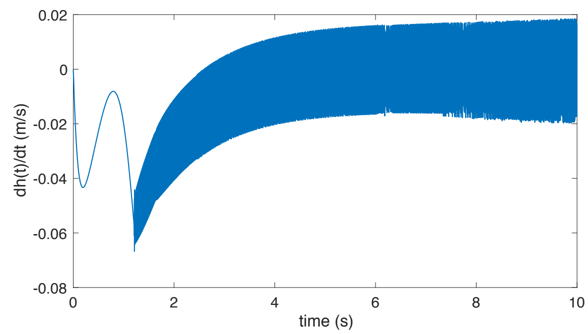
(a) State trajectory, $\theta(t)$



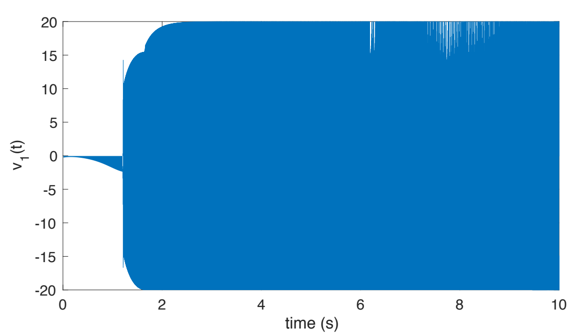
(b) State trajectory, $h(t)$



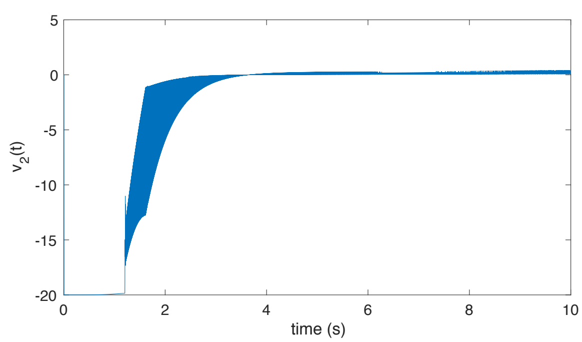
(c) State trajectory, $\dot{\theta}(t)$



(d) State trajectory, $\dot{h}(t)$

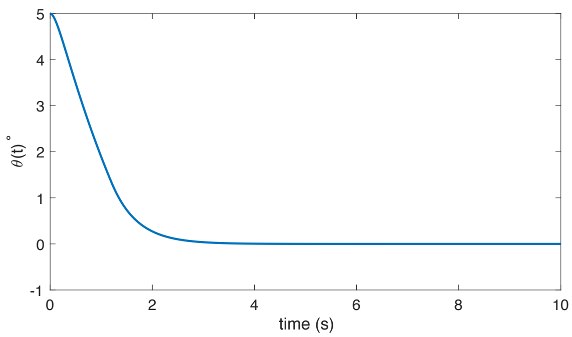


(e) Sliding mode control input signal, $v_1(t)$

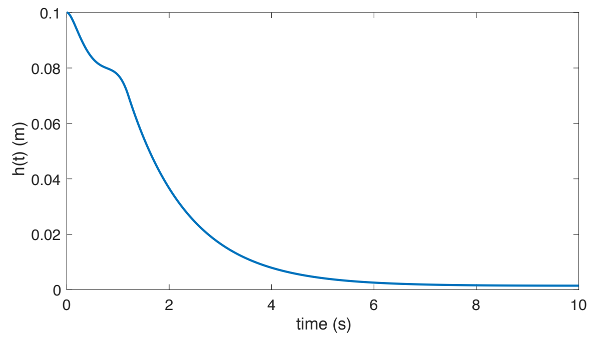


(f) Sliding mode control input signal, $v_2(t)$

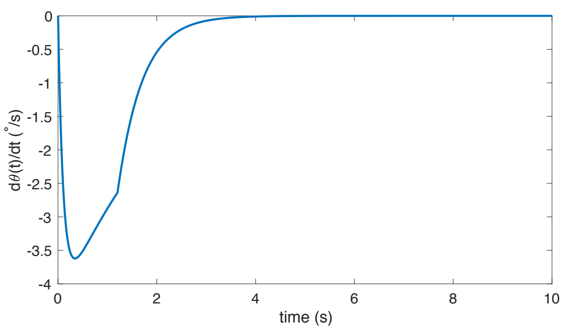
Figure 4.7: Time histories of the states and input signals for $\epsilon = 0.01$, $t_h = 0.001$(sec) and, $k = 20$



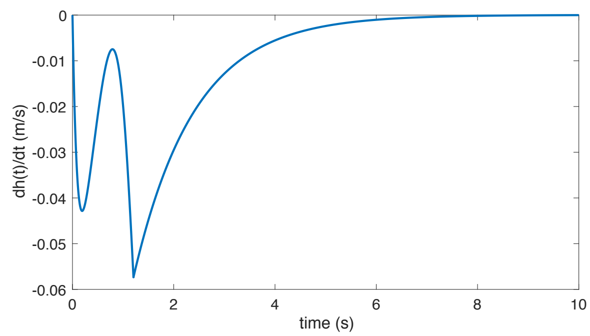
(a) State trajectory, $\theta(t)$



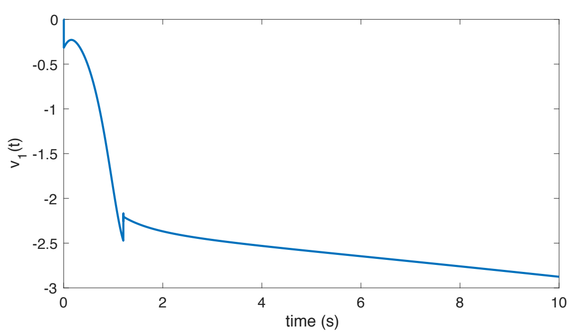
(b) State trajectory, $h(t)$



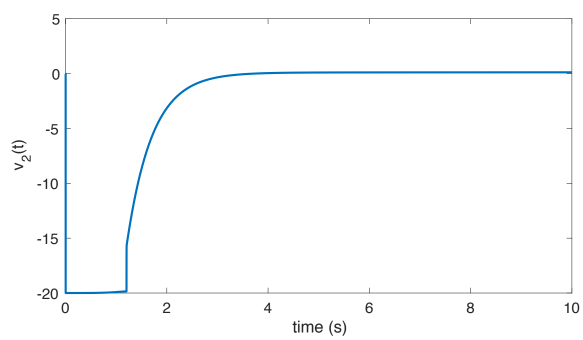
(c) State trajectory, $\dot{\theta}(t)$



(d) State trajectory, $\dot{h}(t)$

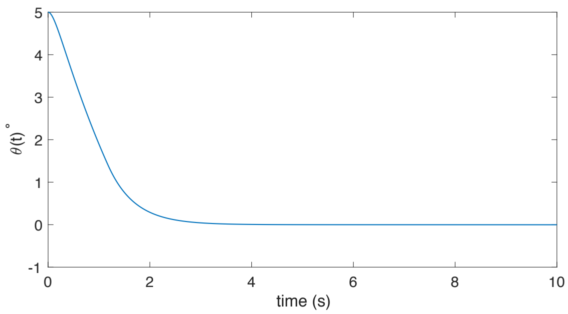


(e) Sliding mode control input signal, $v_1(t)$

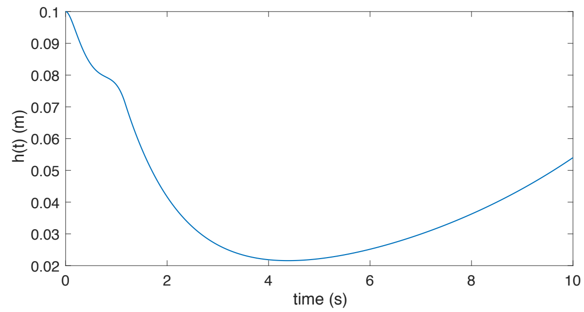


(f) Sliding mode control input signal, $v_2(t)$

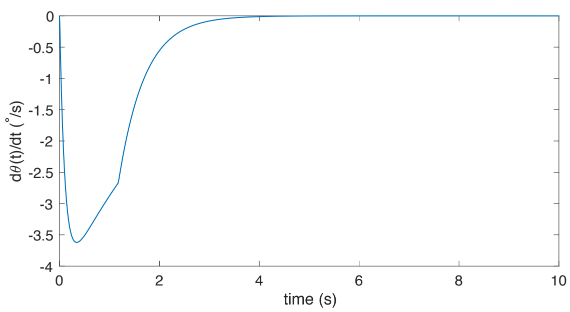
Figure 4.8: Time histories of the states and input signals for $\epsilon = 0.01$, $t_h = 0.0005$ (sec) and, $k = 20$



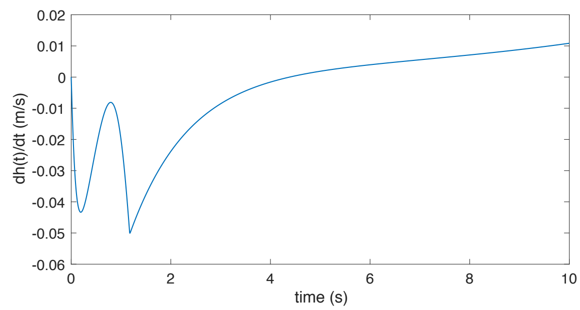
(a) State trajectory, $\theta(t)$



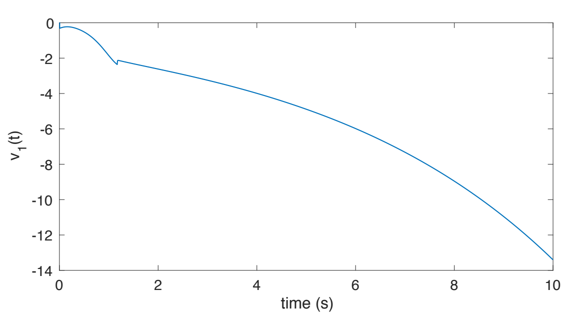
(b) State trajectory, $h(t)$



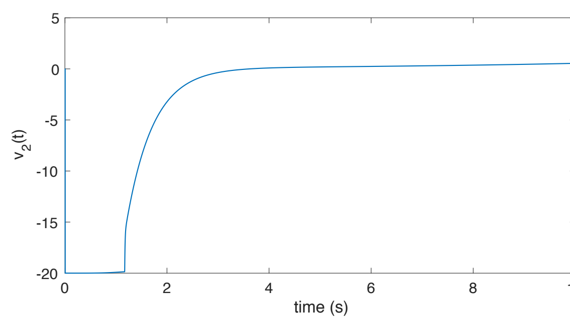
(c) State trajectory, $\dot{\theta}(t)$



(d) State trajectory, $\dot{h}(t)$



(e) Sliding mode control input signal, $v_1(t)$



(f) Sliding mode control input signal, $v_2(t)$

Figure 4.9: Time histories of the states and input signals for $\epsilon = 0.1$, $t_h = 0.001$ (sec) and, $k = 20$

Chapter 5

On Persistency of Excitation

A plethora of approaches exist for the construction of adaptive estimation and control strategies for nonlinear, uncertain systems. See [34, 36, 40, 58, 60, 71] for treatises on popular techniques for uncertain, nonlinear, ordinary differential equations (ODEs). Also of particular pertinence to this chapter are [6, 8, 18, 19, 20, 43], that are representative of methods applicable to classes of uncertain, distributed parameter systems (DPS). The notion of persistently exciting input signals has been extensively studied for wide variety of problems in [72, 83, 50, 9] and references therein. Variants of sufficient conditions for adaptive identification of distributed parameter systems are studied in [6, 8, 19, 20]. The key intuition behind the conventional PE condition is that the input signal needs to have sufficient "energy or richness" in each channel of a finite dimensional system to drive the adaptive estimates of parameters to their true value. A vector $u : \mathbb{R}^+ \to \mathbb{R}^n$ is said to persistently exciting if there exists γ_1, γ_2, $\Delta > 0$ such that

$$\gamma_2 \|v\|^2_{\mathbb{R}^n} \geq \int_t^{t+\Delta} v^T u(\tau) u^T(\tau) v d\tau \geq \gamma_1 \|v\|^2_{\mathbb{R}^n} \tag{5.1}$$

for all $v \in \mathbb{R}^n$ and for any t sufficiently large. The PE condition requires that the integral of the matrix $u(\tau)u^T(\tau)$ is uniformly positive definite over the interval $t, t+\Delta$. Various definitions for PE condition can be found in the texts [40, 71, 58].

5.1 Persistency of excitation in RKHS

Definition 4. *In this section we introduce the definition of persistency of excitation in a RKHS that suffices to prove parameter convergence in this setting. We say that the plant in the RKHS Equation 3.8 is* strongly persistently exciting *if there exist constants* $\Delta, \gamma > 0$, *and* T *such that for* $f \in H$ *with* $\|f\|_H = 1$ *and* $t > T$ *sufficiently large,*

$$\int_t^{t+\Delta} \left(E^*_{x(\tau)} E_{x(\tau)} f, f \right)_H d\tau \gtrsim \gamma.$$

As in the consideration of ODE systems, persistency of excitation above in a RKHS is the key ingredient to prove convergence of the function parameter estimates to the true function.

Theorem 10. *Suppose that the plant in Equation 3.8 is strongly persistently exciting and that either (i) the function* $k(x(.), x(.)) \in L^1((0,\infty); \mathbb{R})$, *or (ii) the matrix* $-A$ *is coercive in the sense that* $(-Av, v) \geq c\|v\|^2 \ \forall \ v \in \mathbb{R}^d$ *and* $\Gamma = P = I_d$. *Then the parameter function error* $\tilde{f}$ *converges strongly to zero,*

$$\lim_{t \to \infty} \|f - \hat{f}(t)\|_H = 0.$$

Proof. We begin by assuming (i) holds. In the proof of Theorem 3 it is shown that V is bounded below and non-increasing, and therefore approaches a limit

$$\lim_{t\to\infty} V(t) = V_\infty < \infty.$$

Since $\tilde{x}(t) \to 0$ as $t \to \infty$, we can conclude that the limit

$$\lim_{t\to\infty} \|\tilde{f}(t)\|_H \lesssim V_\infty.$$

Suppose that $V_\infty \neq 0$. Then there exists a positive, increasing sequence of times $\{t_k\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} t_k = \infty$ and some constant $\delta > 0$ such that

$$\|\tilde{f}(t_k)\|_H^2 \geq \delta$$

for all $k \in \mathbb{N}$. Since the RKHS is persistently exciting, we can write

$$\int_{t_k}^{t_k+\Delta} \left(E_{x(\tau)}^* E_{x(\tau)} \tilde{f}(t_k), \tilde{f}(t_k)\right)_H d\tau \gtrsim \gamma \|\tilde{f}(t_k)\|_H^2 \geq \gamma\delta$$

for each $k \in \mathbb{N}$. By the reproducing property of the RKHS, we can then see that

$$\begin{aligned}
0 < \gamma\delta \leq \gamma\|\tilde{f}(t_k)\|_H^2 \lesssim & \int_{t_k}^{t_k+\Delta} \left(k_{x(\tau)}, \tilde{f}(t_k)\right)_H^2 d\tau \\
& \leq \|\tilde{f}(t_k)\|_H^2 \int_{t_k}^{t_k+\Delta} \|k_{x(\tau)}\|_H^2 d\tau \\
& = \|\tilde{f}(t_k)\|_H^2 \int_{t_k}^{t_k+\Delta} \left(k_{x(\tau)}, \kappa_{x(\tau)}\right)_H d\tau \\
& = \|\tilde{f}(t_k)\|_H^2 \int_{t_k}^{t_k+\Delta} k(x(\tau), x(\tau)) d\tau.
\end{aligned}$$

Since $k(x(.), x(.)) \in L^1((0,\infty);\mathbb{R})$ by assumption, when we take the limit as $k \to \infty$, we obtain the contradiction $0 < \gamma\delta \leq 0$. We conclude therefore that $V_\infty = 0$ and $\lim_{t\to\infty} \|\tilde{f}(t)\|_H = 0$.

We outline the proof when (ii) holds, which is based on slight modifications of arguments that appear in [18, 6, 19, 20, 8, 43] that treat a different class of infinite dimensional nonlinear systems whose state space is cast in terms of a Gelfand triple. Perhaps the simplest analysis follows from [6] for this case. Our hypothesis that $\Gamma = P = I_d$ reduces Equations 3.10 to the form of Equations 2.20 in [6]. The assumption that $-A$ is coercive in our theorem implies the coercivity assumption (A4) in [6] holds. If we define $\mathbb{X} = \mathbb{Y} := \mathbb{R}^n \times H$, then it is clear that the imbeddings $\mathbb{Y} \to \mathbb{X} \to \mathbb{Y}$ are continuous and dense, so that they define a Gelfand triple. Because of the trivial form of the Gelfand triple in this case, it is immediate that the Garding inequality holds in Equation 2.17 in [6]. We identify $BE_{x(t)}$ as the control influence operator $\mathcal{B}^*(\overline{u}(t))$ in [6]. Under these conditions, Theorem 10 follows from Theorem 3.4 in [6] as a special case. $\square$

5.2 PE condition for History dependent operators

We proceed with our analysis in similar style as described in previous section.

Definition 5. *The Equation 4.18 is strongly PE if there exists constants $\Delta, \gamma > 0$ and T such that for $\mu \in P$ with $\|\mu\|_P = 1$ and $t > T$ sufficiently large, we have*

$$\int_t^{t+\Delta} (G^*(\tau)G(\tau)\mu, \mu)_P \, \mathrm{d}\tau \geq \gamma. \tag{5.2}$$

We recall the infinite dimensional governing equations can be written in the form

$$\begin{bmatrix} \dot{\tilde{x}}(t) \\ \\ \dot{\tilde{\mu}}(t) \end{bmatrix} = \begin{bmatrix} A & G(t) \\ \\ -G^*(t) & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}(t) \\ \\ \tilde{\mu}(t) \end{bmatrix} \tag{5.3}$$

with $G(t) = B(\mathcal{H}X)(t)$ and $G(t) : P \to \mathbb{R}^m$

Theorem 11. *Suppose the Equation 5.3 is strongly persistently exciting and either (i) the mapping $\tau \mapsto G^*(\tau)G(\tau)$ is in $L^1([0,\infty);\mathcal{L}(P))$ or (ii) matrix $-A$ is coercive in the sense that $(-Av, v) \geq c\|v\|^2 \ \forall \ v \in \mathbb{R}^d$ and $\Gamma = P = I_d$. Then the parameter function error $\tilde{\mu}$ converges strongly to zero,*

$$\lim_{t\to\infty} \|\mu - \hat{\mu}(t)\|_P = 0.$$

Proof. We begin by assuming (i) holds, it seen that V is bounded below and non-increasing, and therefore approaches a limit

$$\lim_{t\to\infty} V(t) = V_\infty < \infty.$$

Since $\tilde{x}(t) \to 0$ as $t \to \infty$, we can conclude that the limit

$$\lim_{t\to\infty} \|\tilde{\mu}(t)\|_H \lesssim V_\infty.$$

Suppose that $V_\infty \neq 0$ then there exists a positive, increasing sequence of times $\{t_k\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} t_k = \infty$ and some constant $\delta > 0$ such that

$$\|\tilde{\mu}(t_k)\|_H^2 \geq \delta$$

for all $k \in \mathbb{N}$.

$$\begin{aligned}\int_{t_k}^{t_k+\Delta} (G^*(\tau)G(\tau)\tilde{\mu}(t_k), \tilde{\mu}(t_k))_P \,\mathrm{d}\tau &\leq \int_{t_k}^{t_k+\Delta} \|(G^*(\tau)G(\tau)\|_{\mathcal{L}(P)}\|\tilde{\mu}(t_k)\|_P^2 \mathrm{d}\tau \\ &\leq \|\tilde{\mu}(t_k)\|_P^2 \int_{t_k}^{t_k+\Delta} \|(G^*(\tau)G(\tau)\|_{\mathcal{L}(P)} \mathrm{d}\tau\end{aligned}$$

Since the equations are assumed to be strongly persistently exciting for all $k \in \mathbb{N}$,

$$\int_{t_k}^{t_k+\Delta} (G^*(\tau)G(\tau)\tilde{\mu}(t_k), \tilde{\mu}(t_k))_P \,\mathrm{d}\tau \gtrsim \gamma\|\tilde{\mu}(t_k)\|_P^2 \geq \gamma\delta$$

Since we assumed that the map $\tau \mapsto G^*(\tau)G(\tau)$ is in $L^1([0,\infty);\mathcal{L}(P))$ then taking the limit as $k \to \infty$, we have $\int_{t_k}^{t_k+\Delta} (G^*(\tau)G(\tau)\tilde{\mu}(t_k), \tilde{\mu}(t_k))_P \,\mathrm{d}\tau \to 0$. This is a contradiction since we get $0 < \gamma\delta \leq 0$. Therefore we can conclude that $V_\infty = 0$ and $\lim_{t\to\infty} \|\tilde{\mu}(t)\|_P = 0$. $\square$

Recall that already the μ was assumed to be in space of square integrable functions P. Therefore we can infer that the assumption in Definiton 5 implicitly constrains the the properties of the history dependent operator $\mathcal{H}$. This difficult problem may be source of future research.

Chapter 6

Conclusion

In the first part, a novel framework is introduced based on the use of RKHS embedding to study online adaptive estimation problems. The applicability of this framework to solve estimation problems that involve high dimensional scattered data approximation provides much of the motivation for the theory and algorithms is described for online and adaptive estimation. A quick overview of the background theory on RKHS enables rigorous derivation of the results in Sections 3.2 and 3.3. We derive (1) sufficient conditions for the existence and uniqueness of solutions to the RKHS embedding problem, (2) the stability and convergence of the state estimation error, and (3) the convergence of the finite dimensional approximate solutions to the solution of the infinite dimensional state space equations. To illustrate the utility of this approach, a simplified numerical example of adaptive estimation of a road profile is studied and the results are critically analyzed.

In the second part, online adaptive estimation and control methods for class of history dependent differential equation are presented. The class of history dependent operators in this work are constructed using elementary hysteresis kernels, particularly the generalized play operator. Further, we derive an explicit bound for the error of multiwavelet approximation for these operators that are used in construction of FDE's arising in robotics application. The corresponding rate of convergence for finite dimensional approximation of the infinite dimensional solution is determined as a function of resolution level and smoothness parameter of the ridge functions used to construct the kernels. We then proceed to establish the well-posedness for this formulation and further develop an adaptive control strategy that uses sliding mode to identify and compensate the unknown history dependent dynamics. This approach is simulated numerically for a simple pitch-plunge model of wing in a wind tunnel.

While the adaptive control and estimation strategies guarantee convergence of state estimates, an additional condition is needed to guarantee convergence of parameters to the unknown function. Given the possibility of potential application for the framework, it is important to have the convergence in parameters as well. Therefore, a persistence of excitation condition in the RKHS setting is investigated and the understanding of its implications are analyzed. Similarly, an analogous approach is carried out for the history dependent functional differential equations in this dissertation. Conditions on the class of history dependent operators to guarantee persistently exciting input signals is much harder to verify analytically in this case. The satisfaction of these conditions are coupled to the choice of Volterra kernels used to formulate the functional differential equation. Nonetheless, this study stimulates further investigation in adaptive estimation techniques that are suited to governing equations modeled using history dependent functional differential equations.

A common theme throughout this dissertation is to be able to leverage ideas from approximation theory to extend the conventional adaptive estimation and control frameworks. The governing equations for both areas studied are described by evolution in product of finite dimensional Euclidean space and an infinite dimensional Hilbert space. The infinite dimensional states must be approximated. Therefore it has proven to be important to choose approximation methods that are tailored to the estimation problem. The wisdom here is to meticulously examine the applications sought for the mathematical model and then wisely choose your approximation method.

Before we conclude the discussion in this final chapter, we will briefly summarize possible future directions of research based on methods presented in the dissertation. The last section discusses scope of the proposed future work with a two pronged approach: 1) we would like to further probe the theoretical understanding 2) and then identify some challenging applications.

6.1 Scope for future work

The first part of the dissertation extended the adaptive estimation framework in a RKHS and was motivated from an important structural assumption. The ansatz assumed that the unknown function describing the uncertainty belonged to a compact set of a particular Hilbert space of functions. The choice of the hypotheses space and the corresponding *priors* is ordinarily based on assumptions about the smoothness, compactness, approximation properties of the function. Naturally the question arises: under the additional assumptions about the smoothness properties in the hypotheses space, is it possible to extend and establish the well-posedness of the same framework corresponding to the generated stronger topology? or, can we guarantee the strong convergence of results in this subspace? Moreover, is possible to leverage the smoothness of the unknown function to generate faster convergence rates of the estimates in other spaces? The answers to all the above questions is of interest and has practical implications. Accordingly, to achieve the desired convergence rates for functions in a scale of Sobolev spaces, having additional assumptions on continuity and differentiability, we might need to dive deeper in approximation theory. Of the many generalizations in classical wavelet theory, the multiwavelets seem to be a likely candidate to explore to attain better properties in these estimation algorithms. Typically, multiwavelets have short compact support and arbitrary smoothness order. These features could potentially be exploited to deduce estimators according to prior smoothness assumptions of the unknown function. It would likewise be important to extend this framework to adaptively generate bases over the state space, an topic that has been studied in the literature, but not analyzed in the context of the RKHS embedding technique.

6.1.1 Wind Estimation using Micro Aerial Vehicles

The application of the new adaptive estimation framework in terrain mapping is illustrated in Chapter 3. The key idea has been to estimate the terrain by sensing only the vehicle dynamics and without explicitly measuring the road profile. This approach could be potentially extended to wind estimation problems in various applications. The applications could span areas like wind energy harvesting, wake estimation for ships, determining coherent structures in flows and wind patterns in atmosphere. The underlying fundamentals remain the same, wherein we only measure the dynamics of the aerial vehicle to infer the wind speed distribution over a geographical area. The added spatial dimensions of the unknown wind-force field interacting with the aerial vehicle poses new challenges in extending this new adaptive estimation framework. Likewise the unknown function representing the spatial distribution could have different compactness and smoothness properties in each spatial variable. These challenges open up new avenues towards developing suitable estimation algorithms. These novel wind estimation techniques may also have decisive impact on various industries including wind energy, ship design, or even help advance the prediction and tracking of natural calamities like hurricanes, tornadoes, and tsunamis.

6.1.2 Model order reduction for Functional Differential Equations

It is evident throughout the dissertation that the adaptive estimation of governing equations can be achieved using distributed parameter systems. The approximation of these distributed parameter systems can have large number of degrees of freedom. Their have massive dimensionality can make them computationally expensive to solve. The computational cost is proportional to the number of first order ODEs generated. These are numerically solved simultaneously to approximate the DPS. Moreover, for the history dependent FDE problem formulated in Chapter 4, to achieve the convergence rates in approximation, it is seen that the entire history of the states is required by the time-stepping numerical method. This also can result in further computational costs and inefficiencies. The real-time implementations for the applications that motivated the dissertation faster algorithms and efficient computational methods. Hence, the dimensionality reduction for the approximated evolution equations prompts several intriguing research directions. In the past, several techniques have been extensively studied to generate reduced order models for partial differential equations, but not much has been explored in this domain for functional differential equations. To get the ball rolling in the context of this dissertation we need to start by asking, 1) what does it mean to have a low dimensional approximation of an infinite dimensional history dependent equation? 2) can we extract some kind coherent structure from the solutions of infinite dimensional history dependent equations? 3) If we do, then how do we use the finite dimensional approximations of a FDE to span the solution manifold? In particular, the hysteretic behavior of aerodynamics loads changes with variation in boundary conditions and operating parameters such as wind speed, flapping frequency, angle of attack. Is it then possible to capture this transition in behavior in the corresponding reduced order models? Of late, some data-driven methods have been added to the repertoire of model reduction techniques. Likewise again, it would be very promising to probe the understanding of FDE's through the lens of data-driven methods. All of these are intriguing directions that can certainly deepen the understanding of dynamical systems modeled using functional differential equations.

Bibliography

[1] A.J.Kurdila, F.J. Narowich, and Joseph Ward. Persistency of excitation in identification using radial basis function approximants. *SIAM Journal of Control and Optimisation*, 1995.

[2] A.P.Morgan and K.S Narendra. On stability of nonautonomous differential equations $\dot{x} = [a + b(t)]x$, with skew symmetric $b(t)$. *SIAM Journal of Control and Optimisation*, 1977.

[3] Joseph W. Bahlman, Sharon M. Swartz1, and Kenneth S. Breuer. Design and Characterization of a Multi-Articulated Robotic Bat Wing. *Bioinspiration & Biomimetics*, 2013.

[4] T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics Automation Magazine*, 2006.

[5] H.T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems.* Birkhauser, 1989.

[6] J. Baumeister, W. Scondo, M.A. Demetriou, and I.G. Rosen. On-Line Parameter Estimation for Infinite Dimensional Dynamical Systems. *SIAM Journal of Control and Optimization*, 35(2):678–713, 1997.

[7] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 2004.

[8] M. Bohm, M.A. Demetriou, S. Reich, and I.G. Rosen. Model Reference Adaptive Control of Distributed Parameter Systems. *SIAM Journal of Control and Optimization*, 36(1):33–81, 1998.

[9] Stephen Boyd and Shankar Sastry. On parameter convergence in adaptive control. *System and Control Letters*, 3, Dec 1983.

[10] Martin Brokate and Jurgen Sprekels. *Hysteresis and Phase Transitions*. Springer-Verlag, 1996.

[11] Albert Cohen, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Maximal Spaces with Given Rate of Convergence for Thresholding Algorithms. *Applied and Computational Harmonic Analysis*, 11(2):167–191, 2001.

[12] C. Corduneaunu. *Integral Equations and Applications*. Cambridge University Press, 2008.

[13] Felipe Cucker and Ding-Xuan Zhou. *Learning Theory: an Approximation Theory Viewpoint*. Cambridge Monographs on Applied and Computational Mathematics, 2007.

[14] S. Dadashi, P. Bobade, and A.J. Kurdila. Error Estimates for Multiwavelet Approximations of a Class of History Dependent Operators. In *Proceedings of 55th IEEE Conference on Decision and Control CDC 2016*.

[15] S. Dadashi, J. Feaster, J. Bayandor, F. Battaglia, and A.J. Kurdila. Identification and adaptive control of history dependent unsteady aerodynamics for a flapping insect wing. *Nonlinear Dynamics*, 2016.

[16] Wolfgang Dahmen. Stability of Multiscale Transformations. *The Journal of Fourier analysis and applications*, 2:341–362, 1995.

[17] Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[18] M.A. Demetriou. *Adaptive Parameter Estimation of Abstract Parabolic and Hyperbolic Distributed Parameter Systems*. PhD thesis, University of Southern California, 1993.

[19] M.A. Demetriou and I.G. Rosen. Adaptive Identification of Second Order Distributed Parameter Systems. *Inverse Problems*, 10:261–294, 1994.

[20] M.A. Demetriou and I.G. Rosen. On the Persistence of Excitation in the Adaptive Identification of Distributed Parameter Systems. *IEEE Transactions of Automatic Control*, 39:1117–1123, 1994.

[21] Ronald DeVore. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Acta Numerica*, 1998.

[22] Ronald DeVore, Gerard Kerkyacharian, Dominque Picard, and Vladimir Temlyakov. Approximation methods for supervised learning. *Foundations of Computational Mathematics*, 6(1):358, 2006.

[23] Ronald DeVore and George Lorentz. *Constructive Approximation*. Springer, 1993.

[24] G. Dissanayake, H. Durrant-Whyte, and T. Bailey. A computationally efficient solution to the simultaneous localisation and map building problem. In *Proceedings 2000 ICRA. Millennium Conference*.

[25] G. Dissanayake, S. Huang, Z. Wang, and R. Ranasinghe. A review of recent developments in Simultaneous Localization and Mapping. *6th International Conference on Industrial and Information Systems*, pages 477–482, 2011.

[26] W. Dong, Y. Zhao, Y. Chen, and J.A. Farrell. Tracking Control for Nonaffine Systems: A Self-Organizing Approximation Approach. *IEEE Transactions on Neural Networks and Learning Systems*, 23(2), 2012.

[27] David L. Donoho and Ian M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 1995.

[28] Rodney D. Driver. Existence and Stability of Solutions of a Delay-Differential System. *Archive for Rational Mechanics and Analysis*, 10(1):401–426, 1962.

[29] T.E. Duncan, B. Maslowski, and B. Pasik-Duncan. Adaptive Boundary and Point Control of Linear Stochastic Distributed Parameter Systems. *SIAM Journal of Control and Optimization*, 1997.

[30] T.E. Duncan and B. Pasik-Duncan. Adaptive Control of Linear Delay Time Systems. *Stochastic Reports*, 24:45–74, 1988.

[31] T.E. Duncan, B. Pasik-Duncan, and B. Goldys. Adaptive Control of Linear Stochastic Evolution Systems. *Stochastic Reports*, 36:71–90, 1991.

[32] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: Part I. *IEEE Robotics Automation Magazine*, 2006.

[33] Balint Farkas and Sven ake Wegner. Variations on Barbalat's Lemma. *arxiv:1411.1611v3*, 2016.

[34] Jay A. Farrell and Marios M. Polycarpou. *Adaptive approximation based control: unifying neural, fuzzy and traditional adaptive approximation approaches*. Wiley, 2006.

[35] P. Gopalakrishnan and D. Tafti. Effect of Wing Flexibility on Lift and Thrust Production in Flapping Flight. *AIAA Journal*, pages 865–877, May 2010.

[36] N. Hovakimyan and C. Cao. $\mathcal{L}^1$ *Adaptive Control Theory*. SIAM, 2010.

[37] I.D.Mayergoyz. *Mathematical Models of Hysteresis and their Applications*. Elsevier, 2003.

[38] A. Ilchmann, H. Logemann, and E.P. Ryan. Tracking with Prescribed Transient Performance for Hysteretic Systems. *SIAM Journal of Control and Optimisation*, 48, 2010.

[39] Achim Ilchmann, E.P. Ryan, and C.J. Sangwin. Systems of Controlled Functional Differential Equations and Adaptive Tracking. *SIAM Journal of Control and Optimisation*, 40(6):1746–1764, 2002.

[40] Petros Ioannou and Jing Sun. *Robust Adaptive Control*. Dover, 2012.

[41] K. Jeonghwan, A.J. Kurdila, and Thomas W. Strganac. Nonlinear Control of a Prototypical Wing Section with Torsional Nonlinearity. *Journal of Guidance, Control, and Dynamics*, 20(6):1181–1189, 1997.

[42] S. J. Julier and J. K. Uhlmann. A counter example to the theory of simultaneous localization and map building. In *Proceedings 2001 IEEE International Conference on Robotics and Automation*.

[43] Joseph Kazimir and I.G. Rosen. Adaptive Estimation of Nonlinear Distributed Parameter Systems. *International Series of Numerical Mathematics, Birkhauser Verlag*, 118:231–251, 1994.

[44] Fritz Keinert. *Wavelets and Multiwavelets*. Chapman & Hall, 2005.

[45] H.K. Khalil. *Nonlinear Systems*. Prentice-Hall, 2003.

[46] S.V. Konyagin and V.N. Temlyakov. The Entropy in Learning Theory. Error Estimates. *Constructive Approximation*, 25(1):1–27, 2007.

[47] M.A. Krasnoselskii and A.V. Pokrovskii. *Systems with Hysteresis*. Springer, 1989.

[48] N.N. Krasovskii. On the Application of the Second Method of A.M. Lyapunov to Equations with Time Delays. *Prikl. Mat. i Mekh.*, 20, 1956.

[49] N.N. Krasovskii. On the Asymptotic Stability of Systems with After-Effect. *Prikl. Mat. i Mekh.*, pages 513–518, 1956.

[50] K.S.Narendra and A.M.Annaswamy. Persistent excitation in adaptive systems. *International Journal of Control*, 45, Dec 1987.

[51] A.J. Kurdila, J. Li, T.W. Strganac, and G. Webb. Nonlinear Control Methodologies for Hysteresis in PZT Actuated On-Blade Elevons. *Journal of Aerospace Engineering*, pages 167–176, October 2003.

[52] Frank L. Lewis, Darren M. Dawson, and Chaouki T. Abdallah. *Robot Manipulator Control: Theory and Practice*. Marcel Dekker, 2004.

[53] L.Tavernini. Linear multistep methods for the numerical solution of volterra functional differential equations. *Applicable Analysis*, pages 169–185, 1973.

[54] Stephane Mallat. *A wavelet tour of signal processing*. Academic Press, 1999.

[55] Yves Meyer. *Wavelets and operators*. Cambridge University Press, 1992.

[56] A.D. Myshkis. General theory of differential equations with a retarded argument. *Uspekhi Matematicheskikh Nauk*, 4(5):99–141, 1949.

[57] Jun Nakanishia, Jay A. Farrell, and Stefan Schaala. Composite Adaptive Control with Locally Weighted Statistical Learning. *Neural Networks*, 2005.

[58] K.S Narendra and A.M.Annaswamy. *Stable Adaptive Systems*. Prentice Hall, 1989.

[59] K.S. Narendra and K.Parthasarthy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, 1:4–27, 1990.

[60] K.S. Narendra and P. Kudva. Stable Adaptive Schemes for System Identification and Control - Part II. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-4(6):552–560, 1974.

[61] Roland Opfer. Multiscale Kernels. *Advances in Computational Mathematics*, 25:357–380, 2006.

[62] Roland Opfer. Tight Frame Expansions of Multiscale Reproducing Kernels in Sobolev Spaces. *Applied Computational Harmonic Analysis*, 20:357–374, 2006.

[63] B. Pasik-Duncan. On the Consistency of a Least Squares Identification Procedure in Linear Evolution Systems. *Stochastic Reports*, 39:83–94, 1992.

[64] Vern I. Paulsen and Mrinal Raghupathi. *Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge, 2016.

[65] Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, 2011.

[66] Albrecht Pietsch. Approximation Spaces. *Journal of Approximation Theory*, 32:115–134, 1981.

[67] R.A.Adams and J.Fournier. *Sobolev spaces*. Elsevier, 2003.

[68] V.P. Rudakov. On necessary and sufficient conditions for the extendability of solutions of functional-differential equations of the retarded type. *Ukrainian Mathematical Journal*, 1974.

[69] V.P. Rudakov. Qualitative Theory in a Banach Space, Lyapunov-Krasovskii Functionals, and Generalization of Certain Problems. *Ukrainian Mathematical Journal*, pages 130–133, Feb 1978.

[70] E.P. Ryan and C.J. Sangwin. Controlled Functional Differential Equations and Adaptive Tracking. *Systems and Control Letters*, 2002.

[71] Shankar Sastry and Marc Bodson. *Adaptive Control: Stability, Convergence and Robustness*. Dover, 2011.

[72] Nahum Shimkin and Arie Feuer. Persistency of excitation in continuous-time systems. *System and Control Letters*, 9, Sept 1987.

[73] H Shoudong and G.Dissanayake. Convergence and Consistency Analysis for Extended Kalman Filter Based SLAM. *Transaction on Robotics*, 23(5), 2007.

[74] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo. *Robotics: Modeling, Planning, and Control*. Springer-Verlag, 2010.

[75] S. Smale and X. Zhou. Learning Theory Estimates via Integral Operators and Their Approximations. *Constructive Approximation*, 26:153–172, 2007.

[76] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. *Robot Modeling and Control*. Wiley, 2005.

[77] V.N. Temlyakov. Approximation in Learning Theory. *Constructive Approximation*, 2008.

[78] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 2005.

[79] M. Tobak and L.B. Schiff. On the formulation of the aerodynamic characteristics in aircraft dynamics. *National Aeronautics and Space Administration*, 1976.

[80] K. Vishwanath and D. Tafti. Effect of Stroke Deviation on Forward Flapping Flight. *AIAA Journal*, pages 477–482, Jan 2013.

[81] Augusto Visintin. *Differential Models of Hysteresis*. Springer, 1994.

[82] Holger Wendland. *Scattered data approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005.

[83] Jan Willems, Paolo Rapisarda, Ivan Markovsky, and B.L.M DeMoor. A note on persistency of excitation. *System and Control Letters*, 54, Apr 2005.

[84] Yuanyuan Zhao and Jay A. Farrell. Localized adaptive bounds for approximation-based backstepping. *Automatica*, 44:26072613, 2008.

Appendix A

Multiresolution Analysis over the Triangular Domain

We define the multiscaling functions

$$\phi_{j,k}(x) = 1_{\Delta_{i_1,i_2,\ldots,i_j}}(x)/\sqrt{m(\Delta_{i_1,i_2,\ldots,i_j})}$$

in which

$$1_{\Delta_s}(x) = \begin{cases} 1 & x \in \Delta_s \\ 0 & \text{otherwise} \end{cases}$$

and $m(\Delta_{i_1,i_2,\ldots,i_j})$ is the area of a triangle in the level j refinement. We have defined $(hf)(t) \circ \mu = \iint_\Delta \kappa(s,t,f)\mu(s)ds$. The approximation $(h_j f)(t) \circ \mu$ of this operator is given by

$$(h_j f)(t) \circ \mu = \iint_\Delta \sum_{l \in \Gamma_j} 1_{\Delta_{j,l}}(s)\kappa(\xi_{j,l}, t, f)\mu(s)ds,$$

where $\xi_{j,l}$ is the quadrature point of number l triangle of grid level j. We approximate $\mu(s) \approx \sum_{m\in\Gamma_j} \mu_{j,m}\phi_{j,m}(s)$. Therefore,

$$
\begin{aligned}
&(h_j f)(t) \circ \mu_j \\
&= \iint_S \left(\sum_{l\in\Gamma_j} 1_{\Delta_{j,l}}(s)\kappa(\xi_{j,l}, t, f) \sum_{m\in\Gamma_j} \mu_{j,m}\phi_{j,m}(s) \right) ds \\
&= \sum_{l\in\Gamma_j} \sum_{m\in\Gamma_j} \kappa(\xi_{j,l}, t, f) \left(\iint_S 1_{\Delta_{j,l}}(s)\phi_{j,m}(s) ds \right) \mu_{j,m} \\
&= \sum_{l\in\Gamma_j} \kappa(\xi_{j,l}, t, f)\sqrt{m(\Delta_{j,l})}\mu_{j,l}.
\end{aligned}
$$

For an orthonormal basis $\{\phi_k\}_{k=1}^{\infty}$ of the separable Hilbert space P, we define the finite dimensional spaces for constructing approximations as $P_n := \text{span}\,\{\phi_k\}_{k=1}^{n}$. The approximation error E_n of P_n is given by

$$E_n(f) := \inf_{g\in P_n} \|f - g\|_P.$$

The approximation space $\mathcal{A}_2^\alpha$ of order α is defined as the collection of functions in P such that

$$\mathcal{A}_2^\alpha := \left\{ f \in P \middle| |f|_{\mathcal{A}_2^\alpha} := \left\{ \sum_{n=1}^{\infty} (n^\alpha E_n(f))^2 \frac{1}{n} \right\}^{1/2} < \infty \right\}.$$

For our purposes, the approximation spaces are easy to characterize: they consist of all functions $f \in P$ whose generalized Fourier coefficients decay sufficiently fast. That is, $f \in \mathcal{A}_2^\alpha$ if and only if

$$\sum_{k=1}^{\infty} k^{2\alpha} |(f, \phi_k)|^2 \le C$$

for some constant C.

Appendix B

The Projection Operator $\Phi_{J\to j}$

The orthogonal projection operator $\Phi_{J\to j} : V_J \to V_j$ maps a distributed parameter μ_J to μ_j i.e. $\Phi_{J\to j} : \mu_J \mapsto \mu_j$. By exploiting the orthogonality property of the operator we have

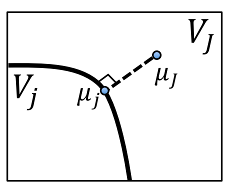


Figure B.1: Projection Operator $\Phi_{J\to j} : V_J \to V_j$

$$\iint_{\Delta} \left(\sum_{m\in\Gamma_j} \mu_{j,m}\phi_{j,m}(s) - \sum_{l\in\Gamma_J} \mu_{J,l}\phi_{J,l}(s) \right) \phi_{j,n}(s)ds = 0.$$

Therefore, we can write

$$\sum_{m\in\Gamma_j} \left(\iint_{\Delta} \phi_{j,m}(s)\phi_{j,n}(s)ds \right) \mu_{j,m} = \sum_{l\in\Gamma_J} \left(\iint_{\Delta} \phi_{J,l}(s)\phi_{j,n}(s)ds \right) \mu_{J,l}.$$

Since orthogonality implies $\iint_\Delta \phi_{j,m}(s)\phi_{j,n}(s)ds = \delta_{m,n}$, we conclude that

$$\mu_{j,n} = \sum_{l\in\Gamma_j}\left(\iint_\Delta \phi_{j,n}(s)\phi_{J,l}(s)ds\right)\mu_{J,l}.$$

From Theorem 1 we have

$$|(h_j f)(t)\circ \Pi_j\mu - (hf)(t)\circ\mu| \leq \tilde{C}2^{-\alpha j},$$

with

$$(hf)(t)\circ\mu = \iint_\Delta k(s,t,f)\mu(s)\mathrm{d}s,$$
$$(h_j f)(t)\circ\mu = \iint_\Delta \sum 1_{\Delta_j,l}(s)k(\zeta_{j,l},t,f)\mu(s)\mathrm{d}s,$$

where $\mu \in P = L^2(\Delta)$ and we approximate $\mu(s) \approx \sum_{l\in\Gamma_J}\mu_{J,l}\phi_{J,l}(s) \in V_J$. To implement this for the finest grid J, we compute

$$\begin{aligned}
(h_J f)(t)\circ\mu_J &= (h_j f)(t)\circ\Pi_J\mu_J,\\
&= \iint\left(\sum 1_{\Delta_J,l}(s)k(\zeta_{J,l},t,f)\sum_{m\in\Gamma_J}\mu_{J,m}\phi_{J,m}(s)\right)\mathrm{d}s,\\
&= \sum_{l\in\Gamma_J}\sum_{m\in\Gamma_J}k(\zeta_{J,l},t,f)\left(\iint 1_{\Delta_J,l}(s)\phi_{J,m}(s)\mathrm{d}s\right)\mu_{J,m},\\
&= \sum_{l\in\Gamma_J}\frac{k(\zeta_{J,l},t,f)\mu_{J,l}}{\left(\sqrt{m(\Delta_{J,l})}\right)},
\end{aligned}$$

when $\sqrt{m(\Delta_{J,l})}$ is the area of the corresponding triangle $\Delta_{J,l}$ in the grid having resolution level J.

Appendix C

Gronwall's Inequality

We employ the integral form of Gronwall's Inequality to obtain our final convergence result. Many forms of Gronwall's Inequality exist, and we will use a particularly simple version. See Section 3.3.4 in [40]. If the piecewise continuous function f satisfies the inequality

$$f(t) \leq \alpha(t) + \int_0^t \beta(s) f(s) ds$$

with some piecewise continuous functions α, β where α is nondecreasing, then

$$f(t) \leq \alpha(t) e^{\int_0^t \beta(s) ds}.$$

Appendix D

Modeling of a Prototypical Wing Section

Figure 4.6 shows a simplified model of the wing. In the figure we denote the center of mass by $c.m.$, A is the aerodynamic center, and O is the elastic axis of the wing. The constants K_h and K_θ are the linear and torsional stiffness, and h is the distance from origin to point O in the fixed reference frame. We denote by x_θ the distance between point O and center of mass, whereas x_a is the distance between O and A. Point O is the origin for the body fixed reference frame.

We employ the Euler-Lagrange technique to derive the equation of motion for the depicted wing model. The function $L(\theta, \dot{\theta})$ is the history dependent lift force acting at the aerodynamic center, and $M(\theta, \dot{\theta})$ is the history dependent aerodynamic moment about point A. The variables L_{β_1} and L_{β_2} are the actuating forces acting at point D, and β_1, β_2 are the angles between the mid chord of the wing and the trailing edge and leading edge flaps, respectively. The position vector of the mass center is given as

$$\mathbf{r}_{c.m.} = h\hat{n}_1 - x_\theta \hat{b}_2,$$

and therefore the corresponding velocity of point C is

$$\dot{\mathbf{r}}_{c.m.} = \dot{h}\hat{n}_1 + x_\theta \dot{\theta}\hat{b}_1.$$

The rotation matrix for transformation between inertial frame of reference to body fixed frame of reference is

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \hat{n}_1 \\ \hat{n}_2 \end{bmatrix}.$$

The kinetic energy is computed to be

$$T = \frac{1}{2}m(\mathbf{r}_{c.m.}.\mathbf{r}_{c.m.}) + \frac{1}{2}I_\theta\dot{\theta}^2,$$

$$T = \frac{1}{2}m(\dot{h}^2 + x_\theta^2\dot{\theta}^2 + 2x_\theta\dot{h}\dot{\theta}\cos\theta) + \frac{1}{2}I_\theta\dot{\theta}^2,$$

and the corresponding potential energy is

$$V = \frac{1}{2}K_h h^2 + \frac{1}{2}K_\theta\theta^2.$$

therefore we can write Lagrangian as $L = T - V$. We apply Euler-Lagrange equations to write the equation of motion as follows

$$\begin{bmatrix} m & mx_\theta \cos\theta \\ mx_\theta \cos\theta & mx_\theta^2 + J \end{bmatrix} \begin{bmatrix} \ddot{h} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} 0 & -mx_\theta \dot{\theta} \sin\theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{h} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} K_h & 0 \\ 0 & K_\theta \end{bmatrix} \begin{bmatrix} h \\ \theta \end{bmatrix}$$

$$= \begin{bmatrix} L(\theta, \dot{\theta}) \cos\theta \\ M(\theta, \dot{\theta}) + x_a L(\theta, \dot{\theta}) \end{bmatrix} + \begin{bmatrix} -L_{\beta_1} \cos(\theta + \beta_1) - L_{\beta_2} \cos(\theta + \beta_2) \\ -L_{\beta_1}(e_1 + d_1 \cos\beta_1) + L_{\beta_2}(e_2 + d_2 \cos\beta_2). \end{bmatrix} \quad \text{(D.1)}$$

The above equation is written in the form of a standard robotic equations of motion $M(q(t))\ddot{q}(t) + C(q(t), \dot{q}(t))\dot{q}(t) + K(q(t)) = Q_a(t) + \tau(t)$, where $q = [h\ \theta]^T$.