REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
28-03-2019	Final Report	1-Sep-2013 - 31-Aug-2017

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Distributed Consensus Learning and Approximation for Geometric and Abstract Surfaces	W911NF-13-1-0407
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Virginia Polytechnic Institute & State Unive North End Center, Suite 4200 300 Turner Street, NW Blacksburg, VA 24061 -0001	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 64080-MA.14

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Andrew Kurdila
UU	UU	UU	UU		19b. TELEPHONE NUMBER 540-231-8028

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 28-Mar-2019

Agency Code:

Proposal Number: 64080MA **Agreement Number: W911NF-13-1-0407**

INVESTIGATOR(S):

Name: Andrew J. Kurdila
Email: kurdila@vt.edu
Phone Number: 5402318028
Principal: Y

Organization: **Virginia Polytechnic Institute & State University**
Address: North End Center, Suite 4200, Blacksburg, VA 240610001
Country: USA
DUNS Number: 003137015 EIN: 546001805

Report Date: 30-Nov-2017 Date Received: 28-Mar-2019
Final Report for Period Beginning 01-Sep-2013 and Ending 31-Aug-2017
Title: Distributed Consensus Learning and Approximation for Geometric and Abstract Surfaces
Begin Performance Period: 01-Sep-2013 **End Performance Period:** 28-Feb-2018
Report Term: 0-Other

Submitted By: Andrew Kurdila Email: kurdila@vt.edu
Phone: (540) 231-8028

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 3 **STEM Participants:** 0

Major Goals: The ARO grant entitled “Distributed Consensus Learning for Geometric and Abstract Surfaces ” (ARO Grant #W911NF-13-1-0407) has extended the results of the recent theory and algorithms developed in [1] under the ARO MURI entitled “Model Classes, Approximation, Metrics for Dynamic Processing of Urban Terrain Data” (ARO Grant #W911NF-07-1-0185) to consider the distributed, sparse or scattered observations from a large numbers sensors in decentralized sensor vehicle networks. The overall goals of the research program can be organized in three major objectives. (1) One goal of this program of research has been to extend the semi-optimal rates of convergence that have been obtained for nonparametric estimation in statistics and approximation to the case of multiple learning agents. This approach is novel in that it provides a characterization of rates of convergence in space of a class of distribution free learning problems cast in appropriate infinite dimensional state smoothness spaces.
This result should be contrasted with the large collection of papers and approaches that derive rates of convergence in time for states that evolve in finite dimensional spaces. The convergence rates in [1] are characterized by upper bounds that hold in probability, as is commonly encountered in the statistical learning theory literature. (2) A second goal of the proposed program of research has been to extend adaptive estimation methodologies to treat problems where observations are collected by multiple distributed observers and where the observations take the form of scattered data. (3) Finally, a primary goal of the research program has be to develop further a high throughput vehicle network that can be used to verify and validate the rates of approximation and estimate convergence under typical operating conditions.

Accomplishments: The discussion here is a short summary of the numerous accomplishments of the award. See the attached pdf file for the complete description of the major accomplishments. Here, we summarize three key categories of major accomplishments in the areas of discrete consensus learning, reproducing kernel Hilbert (RKH) space embedding, and approximation of Koopman operators using approximation space. In the first topic, the error estimate for two stage learning is the first of its kind in the study of consensus estimation of functions by networks of agents. A method for constructing approximations where the mesh dimension depends on the number of samples yields an error rate that holds with overwhelming (geometrically increasing) probability as the number of samples increase. This error bound has no counterpart in the existing theory of consensus estimation.
The second major accomplishment of this award is the introduction of reproducing kernel Hilbert (RKH) space embedding for (continuous time) consensus estimation of unknown functions by agent teams. This method generalizes a host of approaches for conventional continuous time adaptive estimation in \$\mathbb{R}^d\$ to consensus estimates of functions in \$\mathbb{H}^N\$ with \$H\$ a RKH space. In particular, the definition of a new notion of persistency of excitation is introduced in RKH embedding that has the potential to be transformative and constitutes a third major result. It introduces insights regarding persistence in signals that take

values in function spaces that are not possible via the conventional definition for signals that take values in $\mathbb{R}^d$.
A final major accomplishment of this research program is the development of rates of convergence for Koopman operators and Frobenius-Perron operators associated with dependent processes governed by Markov chains. The analysis breaks down the study of such approximations depending on whether the problem data is known, partially unknown, or unknown, and numerous error bounds are derived. New error bounds for Koopman and Frobenius-Perron operators are derived, which induce cor- responding error bounds for consensus estimation of functions.

Training Opportunities: Over the course of this award, opportunities for training included acting as primary advisors to three graduate students working on different aspects of this research. These students included

1) Ms. Jessica Gregory, MS Thesis, ``A Rate of Convergence for Learning Theory with Consensus,'' Department of Mechanical Engineering, Virginia Tech, 2015.

2) Mr. Adam Shoemaker, MS Thesis, ``Nonholonomic Control Utilizing Kinematic Constraints of Differential and Ackermann Steering Based Platforms,'' Virginia Tech, 2015.

3) Dr. Parag Bobade, PhD Dissertation, ``Modeling, Approximation, and Control for a Class of Nonlinear Systems,'' Virginia Tech, 2017.

Ms. Jessica Gregory worked on consensus learning, which was studied mostly during the first year of the grant, while Adam Shoemaker worked on the development of the hardware testbed to be used in the award. Dr. Parag Bobade began the development of the RKH embedding method for adaptive estimation, and has played a critical role in the subsequent refinement of the method for consensus estimation.

Results Dissemination: Over the duration of this award, the results have been disseminated through a number of publications at conferences, through journal publications, online, theses, and a PhD dissertation. The primary dissemination of the results from the first year, as well as select subtopics from the previous two years, has been through the open defense at Virginia Tech and electronic publication of the MS Thesis entitled "A Rate of Convergence for Learning Theory with Consensus," Jessica Gray Gregory, Masters Thesis, Virginia Tech, December, 2014. This thesis considers a set of agents or nodes whose communication is designed in terms of a communication network. This network is a connected undirected graph. Communication between a node and its neighbors flows in both directions and that there are no isolated nodes or islands of nodes. During the second and third years, the research under the grant resulted in the open defense of an MS thesis at Virginia Tech, as well as the subsequent electronic publication entitled "Nonholonomic Control Utilizing Kinematic Constraints of Differential and Ackermann Steering Based Platforms" by Adam Shoemaker during July, 2016. This research effort summarizes many of the hardware developments carried out over the award. The dissertation entitled "Modeling, Approximation, and Control for a Class of Nonlinear Systems," by Parag Bobade was defended in 2017. This dissertation outlines the extension of classical adaptive estimation theory, in finite dimensional state spaces, to adaptive estimation in reproducing kernel Hilbert spaces (RKHS). Rates of approximation in abstract approximation spaces are presented and numerical experiments are summarized. The emphasis and motivation for this research has been to establish a theory that naturally accommodates distributed, scattered measurements. In addition, the following two publications make use of some of the theory summarized in the PhD dissertation above and are a by-product of the analysis. Shirin Dadashi, Parag Bobade, Andrew J. Kurdila, "Error estimates for multiwavelet approximations of a class of history dependent operators" IEEE Conference on Decision and Control (CDC), Las Vegas, 2016. Shirin Dadashi, Hunter G. McClelland, and Andrew Kurdila, "Learning Theory and Empirical Potentials for Modeling Discrete Mechanics," American Control Conference (ACC), 2017. The research carried out under the final year of the grant have been uploaded to the report website. These have been uploaded to the ARO reporting web site as products of the research.

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: There is nothing to report in technology transfer over the final reporting period.

PARTICIPANTS:

Participant Type: PD/PI
Participant: Andrew Kurdila 2318028
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI
Participant: Alex Leonessa
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Parag Bobade
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: 2016 IEEE 55th Conference on Decision and Control (CDC)
Date Received: 30-Oct-2017 Conference Date: 12-Dec-2016 Date Published:
Conference Location: Las Vegas, NV, USA
Paper Title: Error estimates for multiwavelet approximations of a class of history dependent operators
Authors: Shirin Dadashi, Parag Bobade, Andrew Kurdila
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: 2017 American Control Conference (ACC)
Date Received: 30-Oct-2017 Conference Date: 24-May-2017 Date Published:
Conference Location: Seattle, WA, USA
Paper Title: Adaptive estimation in reproducing kernel Hilbert spaces
Authors: Parag Bobade, Suprotim Majumdar, Savio Pereira, Andrew J. Kurdila, John Ferris
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 2-Awaiting Publicat
Conference Name: 2013 American Control Conference (ACC)
Date Received: 01-Nov-2017 Conference Date: 06-Jul-2013 Date Published: 01-Jul-2013
Conference Location: Washington, DC, USA
Paper Title: Adaptive Control via Embedding in Reproducing Kernel Hilbert Spaces
Authors: Andrew Kurdila and Yu Lei
Acknowledged Federal Support: **N**

Publication Type: Conference Paper or Presentation **Publication Status:** 2-Awaiting Publicat
Conference Name: 2012 American Control Conference
Date Received: 01-Nov-2017 Conference Date: 27-Jun-2012 Date Published: 27-Jun-2012
Conference Location: Montre?al, Canada
Paper Title: Learning Theory with Consensus in Reproducing Kernel Hilbert Spaces
Authors: Zhaoda Deng, Jessica Gregory, Andrew Kurdila
Acknowledged Federal Support: **N**

Publication Type: Conference Paper or Presentation **Publication Status:** 2-Awaiting Publicat
Conference Name: 2019 European Control Conference
Date Received: 27-Mar-2019 Conference Date: 25-Jun-2019 Date Published: 25-Jun-2019
Conference Location: Naples, Italy
Paper Title: Multi-agent adaptive estimation with consensus inreproducing kernel Hilbert spaces
Authors: Parag Bobade, Dmitra Panagou, and Andrew J. Kurdila
Acknowledged Federal Support: **Y**

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: Virginia Tech
Date Received: 30-Oct-2017 Completion Date: 9/29/17 10:37AM
Title: Modeling, Approximation, and Control for a Class of Nonlinear Systems
Authors: Parag Bobade
Acknowledged Federal Support: **N**

Publication Type: Thesis or Dissertation
Institution: Virginia Tech
Date Received: 30-Oct-2017 Completion Date: 10/30/14 9:53PM
Title: A Rate of Convergence for Learning Theory with Consensus
Authors: Jessica Gray Gregory
Acknowledged Federal Support: **N**

Publication Type: Thesis or Dissertation
Institution: Virginia Tech
Date Received: Completion Date: 10/30/14 5:53PM
Title: A Rate of Convergence for Learning Theory with Consensus
Authors: Jessica Gray Gregory
Acknowledged Federal Support: **N**

RPPR Final Report
as of 28-Mar-2019

Publication Type: Thesis or Dissertation
Institution: Virginia Tech
Date Received: 30-Oct-2017 Completion Date: 11/30/16 6:58PM
Title: Nonholonomic Control Utilizing Kinematic Constraints of Di♂erential and Ackermann Steering Based Platforms
Authors: Adam Shoemaker
Acknowledged Federal Support: **N**

Final Report

Distributed Consensus Learning for Geometric and Abstract Surfaces

ARO Grant #W911NF-13-1-0407)

March 26, 2019

Contents

1 Goals **3**

2 Outline **3**

3 Major Accomplishments **5**

3.1 An Overview of the Principal Results 5

3.1.1 Discrete, Two-Stage Learning with Consensus 5

3.1.2 RKH Space Embedding for Consensus 6

3.1.3 Koopman Theory and Approximation Spaces 8

4 The Open Problem **9**

4.1 Motivation 9

4.2 State-of-the-Art in Consensus Estimation and Control 10

5 Discrete, Two Stage Learning for Consensus **12**

6 RKH Space Embedding and Consensus **16**

6.1 The Class of Systems 17

6.1.1 Unknown Nonlinear Agent Dynamics 17

6.1.2 Unknown Field Approximation by Agent Teams 18

6.1.3 The General Form of the Agent Equations 18

6.2 The Adaptive Consensus Estimation Framework 19

6.2.1 Conventional Adaptive Estimation 19

6.2.2 Adaptive Estimation in $\mathbb{R}^d \times \mathbb{H}$ 21

6.2.3 Consensus Estimation in $\mathbb{R}^d \times \mathbb{H}$ 22

6.3 Definitions of Consensus 22

6.3.1 Consensus in $\mathbb{R}^N$ 23

6.3.2 Consensus in $\mathbb{H} := H^N$ 24

6.4 Persistence of Excitation in the RKH Spaces 25

6.5 Primary Result for Consensus Estimation 27

7 Approximation Spaces and Koopman Theory **28**

8 Hardware Development **30**

9 Appendix **33**

9.1 Reproducing Kernel Hilbert Spaces 33

9.2 Persistence and Positive Limit Sets 35

1 Goals

The ARO grant entitled "Distributed Consensus Learning for Geometric and Abstract Surfaces " (ARO Grant #W911NF-13-1-0407) considers abstract surface approximation using distributed, sparse, or scattered observations obtained from a large class of sensors that are used by decentralized sensor vehicle networks. The overall goals of the research program can be organized in three major objectives. (1) One goal of this program of research has been to study and derive rates of convergence for (discrete time) consensus function estimates obtained from collectives of multiple learning agents. One topic under this objective is to study rates of convergence in appropriate infinite dimensional approximation or smoothness spaces. This goal should be contrasted with the large body of literature that derives rates of convergence *in time* for states that evolve in some fixed, low dimensional spaces $\mathbb{R}^d$. We find that the convergence results derived in this research program are characterized by upper bounds that hold in probability, as is commonly encountered in the approximation theory, distribution-free learning theory, and the theory of nonlinear regression. In one sense this grant has studied methods to endow consensus estimation with such an associated framework. (2) A second goal of the proposed program of research has been to extend adaptive (continuous time) estimation methodologies to treat problems where observations are collectedalong trajectories of multiple distributed observers and where the observations take the form of scattered data. This has culminated in the development of the method of reproducing Hilbert (RKH) space embedding. This technique subsumes the conventional theory of adaptive estimation that forms the grounding for numerous adaptive estimation and control strategies. (3) Finally, a goal of the research program has been to develop further a vehicle sensor network that can be used to study such estimation algorithms under typical operating conditions.

2 Outline

The number and breadth of new results, methods, and algorithms derived under this grant is substantial. We carefully present a narrative in this document that is reasonably self-contained and outlines the primary contributions of this award to the scientific community. The program of research has the potential to be transformative for those in the scientific community that use or will use sensor vehicle networks for consensus estimation of unknown abstract functions or fields.

This document begins in Section 3 with an executive summary of the major accomplishments of this research award. We provide a succinct technical description of the formulations, theory, and algorithms derived under the grant in Section 4 through Section 8. The discussion is necessarily brief, and the reader is referred in the discussion to the details in either [78, 39, 88, 89, 90], which have been uploaded to the report website, or the MS thesis [79] and Dissertation [78] also available online.

We begin with an introduction to the open problem in consensus estimation via sensor vehicle teams that is addressed in this report in Section 4. The challenges that are faced in solving this problem are discussed in detail. Following this introduction, we present our first formulation that yields an initial solution to the problem. We define a new, general statement of *learning theory with consensus* in Section 5. This section formulates the problem of estimation using sensor networks and multiagent teams as a foundational problem of learning theory. We derive a realizable algorithm for the joint estimation by a sensor vehicle team of an abstract surface or function in a reproducing kernel Hilbert (RKH) space using this approach. This algorithm is referred to as the two-stage learning dynamic for consensus estimation.

The error results obtained for the two-stage learning dynamic are unique among existing consensus estimation and control approaches. Also, this first approach, among many new conclusions regarding errors in consensus estimation derived in this award, illustrates the importance and the role of approximation spaces in estimation methods for general dynamical systems. This is a topic the investigators focus on toward the end of the award. We return to this topic shortly. In addition, one of the underlying assumptions that plays a critical role in this analysis is that each agent collects samples of the unknown field that are independent and identically distributed. This assumption is quite strong which allows for correspondingly strong conclusions about error rates. It is not generally applicable when the samples are collected by agents in the network along a particular vehicle path. See [90] for a detailed discussion of how independence and dependence can manifest for various system models and experimental setups.

With this latter limitation in mind, in the next Section 6 we discuss the contributions to consensus estimation using the techniques of *reproducing kernel Hilbert (RKH) space embedding.* This is an entirely new formulation of adaptive consensus estimation, one that subsumes many of the standard approaches to adaptive estimation in finite dimensions. Excellent references on such "canonical structure" of adaptive estimation in $\mathbb{R}^d$ include [55, 62, 61, 67, 63], among others. The embedding approach extends the hypotheses that underly Section 4 in several respects. We cast the consensus estimation problem from vehicle teams as an (infinite dimensional) distributed parameter estimation problem in a RKH space. We derive sufficient conditions for the solution of the consensus estimation problem in continuous time. We introduce a new notion of persistency of excitation that is expressed in terms of evaluation operators acting on a RKH space. This condition has no precedent in adaptive estimation from vehicle sensor networks and enables strong conclusions about the convergence of multiagent estimators. The results of the research have the potential to induce a fundamental change how modeling within the sensor network and estimation community is carried out for consensus estimation.

As noted above, one of the ingredients in the study of the two-stage learning dynamic is that the rates of convergence rely on membership in an approximation space. In the last part of the grant, this topic has been studied in great detail. The results of the analysis are mostly contained in the research manuscript [90], which has been uploaded to the report website. In Section 7 we introduce several new results that characterize rates of approximations for quite general dynamical systems, including the special case of joint estimation using vehicle teams.

Finally, we conclude with a description of the experimental testbed in Section 8.

3 Major Accomplishments

The ARO grant entitled "Distributed Consensus Learning for Geometric and Abstract Surfaces, "Grant #W911NF-13-1-0407 has been successful in developing new and general theoretical methods, as well as realizable algorithms, for posing and solving problems of consensus function estimation by vehicle teams. The overall strategy constitutes a broad method of problem formulation that implies several distinct and specific results that are unprecedented in sensor vehicle estimation research. These results should have a profound impact on those researchers and developers working on multiagent estimation problems for vehicle sensor networks and multiagent teams.

3.1 An Overview of the Principal Results

Since there are actually quite a few new results in this report, we briefly outline the principal results. We have restricted the discussion to those results that have the greatest potential to have a profound impact on future research and applications of vehicle sensor teams.

3.1.1 Discrete, Two-Stage Learning with Consensus

The general setup of a discrete, two stage learning dynamic for consensus estimation was initially studied in [44], and it was later refined in the Master's Thesis entitled *A Rate of Convergence for Learning Theory with Consensus* in [78] by Ms. Jessica Gregory in 2015. The initial year of the grant studied and refined this result as described in the annual reports.

> *The error estimate for two stage learning is the first of its kind in the study of consensus estimation of functions by networks of agents. A method for constructing approximations where the mesh dimension depends on the number of samples yields an error rate that holds with overwhelming (geometrically increasing) probability as the number of samples increase. This error bound has no counterpart in the existing theory of consensus estimation.*

It is derived by extending the foundations of statistical learning theory, which treats a single agent constructing an estimate of an unknown function from samples, to consensus estimation of a function by a team of agents. We refer to this general topic as "learning theory with consensus." The error rates for the consensus learning algorithms described in this section are based on strategies rooted in approximation spaces and distribution-free learning theory. [102, 98, 99, 105, 104, 103, 100, 101]

In this formulation, each agent i in a team of N agents makes estimates $\hat{f}_{i,j}$ at the end of epoch j of an unknown function $f^* : X \to Y$. An epoch is an organization of the measurements into blocks each of length m_j, and the total number of samples collected by each agent through epoch j is $m(j) := \sum_j m_j$. The construction of the estimates is carried out on a grid whose refinement level is determined by the number of samples collected. The new result states that, provided the hypotheses summarized in [44, 78] hold, it is with probability at least $\prod_j (1 - (Cm_j)^{-\beta})$ in the collective sample space $Z^{Nm(j)}$ that the error in the approximation of the unknown functions decreases geometrically with the number of epochs for some fixed constants $\beta, C > 0$. The most important feature of this result is that the unknown regressor function has a certain degree of smoothness s, the number of samples and the dimension of the approximant space increase in a coupled fashion, and the rate of their increase depends on this smoothness. This notion of smoothness is expressed in terms of membership in an approximation space of order s, a topic that is considered in much more depth in the following years of the award. See Section 7 and the references therein. This subsequent development, the role of approximation spaces in equations of evolution, is described in the context of Koopman theory in Section 7.

While the details of the derivation and its refinements in [78] far exceed the limits of this report, a summary of the general strategy is given in Section 5 below to make this report self-contained. The interested reader is referred [44, 78] for the details. These manuscripts have been uploaded to the report website also.

3.1.2 RKH Space Embedding for Consensus

Even though the two stage learning dynamic described above is the first result of its kind for consensus estimation in discrete time, it assumes that samples collected by each agent are independent and identically distributed. A standing assumption is that observations made by each agent over discrete time are independent and identically distributed according to some unknown probability measure. One important goal of the program has been to allow measurements that are taken along the path of a dynamic system. Samples along such a path are generally not independent. In addition, there is a large and well-documented collection of continuous time adaptive estimation methods that form the foundation of numerous estimation and control problems in $\mathbb{R}^d$ for systems governed by ordinary differential equations (ODEs).

> *The second major accomplishment of this award is the introduction of reproducing kernel Hilbert (RKH) space embedding for (continuous time) consensus estimation of unknown functions by agent teams. This method generalizes a host of approaches for conventional (continuous time) adaptive estimation in* $\mathbb{R}^d$ *to consensus estimates of*

functions in $\mathbb{H} := H^N$ with H an RKH space.

The development of the RKH embedding method for adaptive estimation began with a discussion of a single agent and was initially studied in [39]. Further discussion of the approach is documented in [79], the PhD dissertation entitled *Modeling, Approximation, and Control for a Class of Nonlinear Systems* by Dr. Parag Bobade who graduated in 2017. Ongoing work on the extension of this work to multiagent consensus estimation is documented in [88, 89].

In the derivation to the final result of consensus RKH embedding, there are again intermediate results that are noteworthy. In particular, the definition of

> *a new notion of persistency of excitation again has the potential to be transformative and constitutes a third major result: it introduces insights regarding persistence in signals that take values in function spaces that are not possible via the conventional definition for signals that take values in $\mathbb{R}^d$.*

As in the last section, the details of this formulation far exceed the limitations of this report. We give a survey of the general strategy in Section 5 below. From [89] the overview in Section 5 starts with "... the introduction of a new persistency of excitation condition in Definitions 2 and 3 for semidynamical systems that feature a state that evolves in a RKH space $\mathbb{H} := H^N$. In contrast to the persistency condition for conventional collections of ODEs, the new condition is cast in terms of the evaluation operator E_x on the RKH space H, and it adjoint E_x^*. Theorem 3 demonstrates that the new definition of persistency gives a succinct condition under which trajectories $t \mapsto \mathbb{G}(t) \in \mathbb{H}$ converge to zero in the norm on the vector-valued RKH space $\mathbb{H} := H^N$. This theorem requires that $\mathbb{G} \in L^\infty((0,\infty),\mathbb{H})$ and that each agent trajectory $t \mapsto x_i(t)$ satisfies $\mathfrak{K}(x_i(\cdot), x_i(\cdot)) \in L^1((0,\infty),\mathbb{R})$ with $\mathfrak{K}$ the kernel for H. This theorem shows that the new definition of persistency for the RKH space gives an intuitive generalization of persistency over a finite dimensional vector space. In fact, it is possible to interpret the new condition as a novel type of "partial persistence of excitation" either over an indexing set Ω or a restricted domain Ω.

The method of RKH embedding for consensus estimation constructs estimates as follows. At time t each agent i shares its local function estimate $\hat{f}_i \in H$ with all its neighbors $j \in \mathcal{N}(i)$, with $\mathcal{N}(i)$ equal to the set of neighbors of node i. Neighbors are defined in terms of an undirected, connected, time-invariant graph having a graph Laplacian $\mathcal{L}$. The graph Laplacian is used to lift the notion of conventional agreement $\mathcal{H}_a$ and disagreement spaces $\mathcal{H}_d$ with $\mathbb{R}^N \equiv \mathcal{H}_a \oplus \mathcal{H}_d$ to corresponding spaces of functions, $\mathbb{H}_a$ and $\mathbb{H}_d$, that induce an orthogonal decomposition of $\mathbb{H} := H^N := \mathbb{H}_a \oplus \mathbb{H}_d$ of *vector-valued functions*. The estimator in Equation 12 and the learning law in Equation 13 is used so that each agent constructs at time t its own local estimate $\hat{x}_i(t)$ of its state and a local function approximation $\hat{f}_i(t)$ of the unknown function f from its own local state $x_i(t)$ and the shared knowledge of the estimates of its neighbors. Theorem 4 shows that the state error $\tilde{x}(t) := x_i(t) - \hat{x}_i(t)$ approaches zero as $t \to \infty$, but also that the the vector of function estimates $\hat{\mathbb{F}}(t)$ at time t approaches the agreement subspace of functions $\mathbb{H}_a \subseteq \mathbb{H}$ as $t \to \infty$. Finally, it is proven that if the system is persistently excited in the sense of either Definition 2 or 3, then the total error $\tilde{\mathbb{F}}_j(t) := \mathbb{F} - \hat{\mathbb{F}}_j(t)$ approaches zero as $t \to \infty$. "

Again, while the details of the approach exceed the scope of this report, a summary of the overall strategy is given in Section 5. The discussion in the section extracts and summarizes many of the details in [39, 79, 88, 89]. References [39, 88, 89] have been uploaded to the report website and the MS thesis [79] is also available online.

3.1.3 Koopman Theory and Approximation Spaces

The final major theoretical accomplishment of this report has been to construct a theory for the estimation of the dynamics of a family of discrete, but generally nonlinear, systems in terms of linear approximation spaces. This line of research addresses the primary objective of investigating the role of smoothness or approximation spaces in defining priors for adaptive estimation and in particular consensus estimation. In Section 5 it has been noted that one of the underlying assumptions of consensus estimation using the two-stage learning dynamic is that the observations $\{(x_{i,k}, y_{i,k})\}_{k \geq 0}$ collected by agent i are independently and identically distributed according to an unknown probability measure ρ_i defined over the sample space $Z := X \times Y$. While this assumption facilitates intuitive and precise derivation of rates of convergence for the consensus learning theory methods in Section 5, this assumption does not hold, strictly speaking, for most detailed models of vehicle team dynamics. Rather, discrete and nonlinear models of the motion of vehicle teams are most frequently understood as generating dependent observations along the sample path of the agents.

The major accomplishments in this portion of the research seek to lay the foundation for a more general theory of consensus approximation from sensor vehicle networks, and to do so with a careful study of rates of approximation in terms of approximation spaces.

> *A final major accomplishment of this research program is the development as outlined in the research manuscript [90] of rates of convergence for Koopman operators and Frobenius-Perron operators associated with dependent processes governed by Markov chains. The analysis breaks down the study of such approximations depending on whether the problem data is known, partially unknown, or unknown, and numerous error bounds are derived. New error bounds for Koopman and Frobenius-Perron operators are derived, which induce corresponding error bounds for consensus estimation of functions.*

The research monograph [90] includes 1) estimates of error bounds for Koopman and Perron-Frobenius operators that establish a bias versus variance tradeoff for consensus estimation, 2) improvements of existing convergence results for the well-known and popular extended dynamic mode decomposition (EDMD) method, and 3) new error bounds for dynamical systems that collect samples along the sample path. These results are summarized in Section 7 in this report and the details can be found in [90]. Reference [90] has been uploaded to the report web site and is also available online.

4 The Open Problem

Consensus in multiagent networks has been studied carefully over last few decades. References [41, 42, 54] provide a good survey of the state of the art in this sub-discipline of dynamical systems theory in $\mathbb{R}^d$, as does the recent book [108]. The report reviews the overall recent trends in this research in Section 4.2. In this report we are specifically interested in the problem of achieving consensus and convergence of estimates of an unknown function that are generated by multiple agents of a sensor network. The unknown function f is assumed to be contained in a Banach space $\mathbb{B}$. In discussions of the two-stage learning dynamic of Section 4 the function is assumed to lie in the approximation space $A^{r,2}$. In the discussion of the RKH space embedding method discussed in Section 6, the unknown function is assumed to be an element of the RKH space H. In the description of the approximation techniques based on Koopman theory in Section 7, the unknown function is assumed to be an element of either a spectral approximation space $A^{r,q}_{\lambda}$ or more generally the approximation space $A^{r,q}$.

4.1 Motivation

The motivation for studying the consensus estimation problem problem stems from explosion of growth in networked sensor vehicle applications that enable distributed sensing. The systems in this report include some canonical examples. The methods here can be used to study cases when the agents seek to estimate some unknown function that determines their uncertain or unknown governing equations of motion, or when they seek to estimate jointly some unknown field defined over their domain of operation.

As noted in [88, 89], "... in the case of applications in wind estimation [46, 47], the mobile sensors could be unmanned autonomous air vehicles (UAVs). In such a case we desire to estimate the unknown or uncertain wind velocity using a team of such UAVs soaring over particular region in the atmosphere. In marine applications teams of surface or underwater vehicles can be used to measure chemical plumes in an ocean region. Often, due to the size of the region, it is computationally inefficient or just infeasible to visit a densely scattered collection of points throughout the domain of interest to collect data and build estimates. In a related problem, it may be that the actual equations of motion governing the vehicle dynamics of the multiagent team are uncertain or unknown. Then it can be of interest for the team to estimate jointly the unknown function that determines their motion.

In a general sense, the goal of the finite collection of agents $i = 1, \dots, N$ in this report can be defined easily. Each agent i must use measurements of their *local* full state $x_i(t) \in \mathbb{R}^s$ available at time $t \geq 0$ and certain shared information *about the family of function estimates in* $\hat{\mathbb{F}}(t) := \{\hat{f}_1(t), \dots, \hat{f}_N(t)\}$ to build a local estimate $\hat{f}_i(t)$ of the unknown, uncertain function f at time t. If the construction of the estimates is carried out at discrete times $\{t_k\}_{k\geq 0}$ we write $\hat{f}_{i,k} := \hat{f}_i(t_k)$. Of course, consensus estimation can include the case when all agents act in complete independence, but most frequently some information sharing is assumed in consensus estimation. The problem is most interesting of course when the shared information available to agent i refers only to estimates from its neighbors $\mathcal{N}(i)$. In some sense the notion of neighbors gives the function update the character of local approximations. Consensus is achieved if $\hat{f}_i(t) \to f$ as $t \to \infty$ in some suitable topology on functions for each i. When samples are restricted to some subdomain, it is of interest to understand in what sense the resulting collection of estimates converge to the unknown function. "

4.2 State-of-the-Art in Consensus Estimation and Control

This research grant has addressed a formidable and emerging problem in adaptive estimation using sensor networks, one that has received very little attention in the recent literature on consensus estimation from sensor estimation over the past 15 years. Since the earliest, well-cited research on consensus dynamics of agent teams in references such as [1, 2], the extensive body of literature on consensus estimation has primarily focused on

- models of the dynamics of agents networks in $\mathbb{R}^d$,
- sufficient conditions to guarantee convergence to consensus in $\mathbb{R}^d$,
- different structural assumptions about the evolution equations in $\mathbb{R}^d$, and
- rates of convergence *in time* of agent networks to consensus in $\mathbb{R}^d$.

for a fixed, relatively small, dimension $d > 0$. These topics are discussed in the text [108] and many of the recent publications on the topic in the last two decades. For instance, models of the dynamics are mostly linear systems that evolve in $\mathbb{R}^d$ as in References [1, 2, 3, 4, 5, 7, 8, 12, 19]. Some nonlinear evolution equations in $\mathbb{R}^d$ are also considered, but these appear much less frequently, as in [6, 33]. Examples of how the evolution equations can vary structurally include references [2, 9, 34] that study discrete, switched, and hybrid systems; [30] that study event-triggered dynamics; [17] that accounts for delay dynamics; [25] that investigates variable connectivity; [8, 24] that accounts for types of disturbances; [33, 35] that examines fault tolerance.

Despite the maturity of the consensus estimation field, there are emergent problems that face the consensus estimation community, problems that are anticipated to dominate questions in this area of research for years to come. As clearly shown by the above overview, research has typically been concerned with vehicle guidance and navigation equations in $\mathbb{R}^d$ with $d > 0$ a relatively small fixed number. Agents usually exchange information about the state of the vehicle, such as inertial locations, orientations, relative heading information, angular rates of rotation, velocities, angular velocities, accelerations, or other typical kinematic information. To be sure, some information regarding terrain, geography, or geometry has been shared in some aspects of consensus estimation. References [72, 73] specifically discuss general measurement exchange policies for the estimation of unknown functions. These methods and algorithms are based on exchange of raw measurement data and cast the problem of estimation in a reproducing kernel Hilbert (RKH) space.

In addition, the large body of literature on simultaneous localization and mapping (SLAM), and its variants, contains numerous references on map building. A good overview of this general approach can be found in [107] and the references therein. Still, such map building is ordinarily viewed from the context of a single agent or vehicle, the method largely limits agent estimates of terrain to piecewise constant estimates of probability fields, and method does not account for or analyze rates of approximation *in the spatial domain* of the unknown field. This latter topic is a central and defining feature in modern approaches to nonparametric regression, approximation theory, and variants of statistical learning theory. [105, 98, 99]

This research grant has studied consensus estimation of unknown, abstract *fields or functions* that are increasingly the target of interest of sensor vehicle teams. A typical schematic of a sensor vehicle team is illustrated in Figure 1. This network consists of several potentially autonomous agents, each of which is equipped with a collection of high resolution, high bandwidth sensors. The sensors can include visual spectrum CCD cameras, infrared CCD cameras, acoustic arrays, active cameras that feature arrays of lasers or acoustic sensors, and various laser imaging systems such as scanning or flash LIDAR. Such systems provide high resolution "point cloud" measurements at extremely high rates, often measured in megabytes per second or minute. A schematic of such a system is depicted in Figure 1. One common three dimensional scanner studied at Virginia Tech generates highly dense point cloud measurements over time as illustrated in the Figure 2. For the camera systems the point cloud measurements are obtained by processing the raw CCD images via simple triangulation, epipolar geometry, stereo imaging, structure from motion, or simultaneous localization and mapping (SLAM) algorithms. Detailed mapping exercises using these systems can and often do require archival of gigabytes of data for hours, or even just a few minutes, of operation.

In summary, and in view of the above review of the current state-of-the-art in consensus estimation from vehicle teams, this report summarizes the overall progress made during the grant to derive

- a general formulation of consensus estimation for abstract functions and surfaces,
- sufficient conditions that guarantee convergence of agent estimates of abstract functions and surfaces,
- associated realizable algorithms that generate consensus estimates, and
- convergence rates of the consensus estimates *in the spatial domain.*

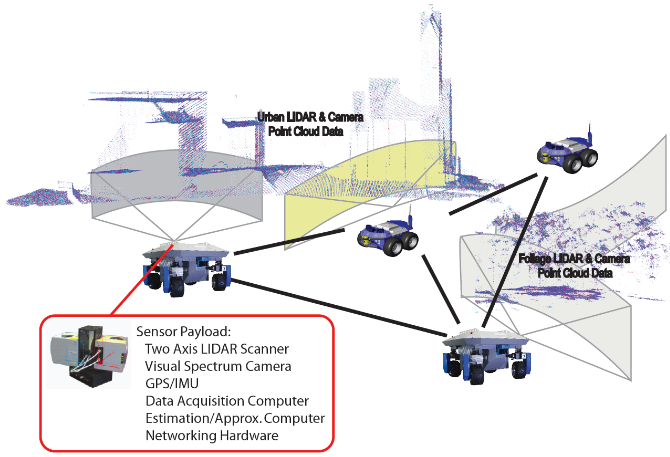
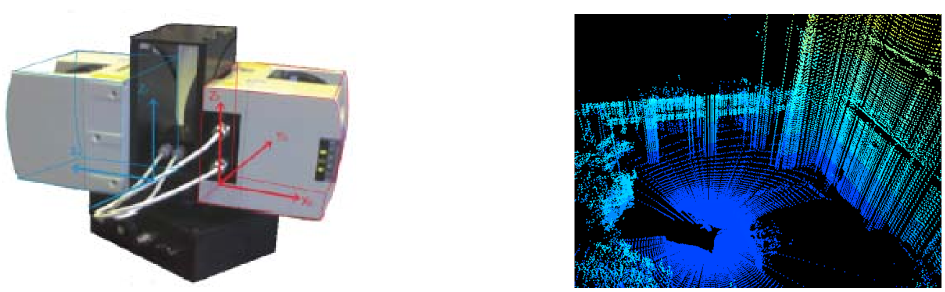


Figure 1: Typical Ground Vehicle Sensor Network: Proposal Document



(a) VT Designed Laser Scanning Imager

(b) Sample Urban Geometry

Figure 2: High Resolution, 3D LIDAR Imaging Sensor

These goals should be carefully contrasted with the general trends noted in the review of the state-of-the-art in consensus estimation and control in $\mathbb{R}^d$ above. The program of research described in this report seeks to synthesize dynamical systems theory for agent teams, approximation theory, and aspects of statistical learning theory.

5 Discrete, Two Stage Learning for Consensus

This initial phase of research in this grant is built on a general two-stage learning dynamic first introduced by the investigators in [44], which was refined and developed in its final form in [78] in 2015. Here we summarize the general strategy used to derive this result, with emphasis on the limitations that were addressed in subsequent years. We denote the number of agents by N, and the dynamics of agent i is governed by a state vector x_i that is assumed to evolve in the compact set $X \subset \mathbb{R}^s$. The motion of the collective is given by the vector $X := \{x_1, \ldots, x_N\} \in \mathbb{R}^d := \mathbb{R}^{Ns}$. The approach in this section alternates between periods of local estimation and information exchange. Each agent makes an observation of its own state $x_{i,k}$ and a measurement of the unknown, scalar-valued unknown $y_{i,k} \in \mathbb{R}$, the latter of which is a sample of some unknown function $f : X \to Y$, at each discrete time step k. So each pair of samples $(x_{i,k}, y_{i,k}) \in Z := X \times Y$, with Z the observation space of the agents. The overall goal of the sensor team is to construct an estimate $\hat{f}_{i,k} : X \to Y$ at time k that approximates the unknown function $f : X \to Y$ from its local observations, previous function estimates, and whatever information is shared among the agents at that step.

The goal is for the agents to reach agreement, or consensus, in their estimates of the unknown function. At any time step k we denote $\mathbb{F}_k := (f_{1,k}, \ldots, f_{N_k}) \in \mathbb{H} := H^N$, and we say that consensus is reached provided

$$\mathbb{F} \in \mathbb{H}_a := \{\mathbb{F} \in \mathbb{H} : \mathbb{F} = (f, \ldots, f) \text{ for } f \in H\}.$$

The set $\mathbb{H}_a$ is known as the agreement space. The agreement space $\mathbb{H}_a$ and the disagreement space $\mathbb{H}_d$ are defined as in [89] by lifting the analysis for consensus in $\mathbb{R}^d$ to a RKH space $\mathbb{H} := \mathbb{H}_a \oplus \mathbb{H}_d$. We enforce the condition that the agents reach consensus via the introduction of a constraint operator $\boldsymbol{B} \in \mathcal{L}(\mathbb{H}, \mathbb{H})$ that acts on functions and that is designed to have the property that $\mathbb{F} \in \mathbb{H}_a \leftrightarrow \boldsymbol{B}\mathbb{F} = 0$. As we discuss more fully in Section 3.1.2, or in reference [89], the operator $\boldsymbol{B}$ can be realized as the operator on $\mathbb{H}$ induced by multiplication by the Laplacian of an undirected, connected, time-invariant communication graph.

The general approach to come up with a learning law is to define a particular quadratic cost or performance functional for the collection of estimates and seek the best approximation from the agreement space. Let ρ_i be a probability measure on $Z = X \times Y$ that is the unknown distribution of the independent and identically distributed measurements $\{z_{i,k}\}$ collected by agent i. We define a local cost functional $J_i : H \to \mathbb{R}$ for each agent that depends on the measure ρ_i for agent i with

$$J_i(f) := \frac{1}{2} \int_Z (f(x) - y))^2 \rho^i(dz) + \frac{1}{2} \lambda^i \|f\|_H^2.$$

for any $f \in H$. Here λ_i is a regularization parameter. The reader may see [78, 89, 80] for the discussion of how this choice relates to the conventional one used in classical learning theory and approximation. The aggregate performance in estimation for all the agents in the team is written as the sum of the local cost functionals, and we have $\boldsymbol{J}(\mathbb{F}) := \sum_{i=1}^n J_i(f_i)$ with $\boldsymbol{J} : \mathbb{H} \to \mathbb{R}$ for $\mathbb{F} = (f^1, \ldots, f^n) \in \mathbb{H}$. We construct the learning dynamic so that it converges to the minimizer $\mathbb{F}^*$ of the constrained optimization problem

$$\mathbb{F}^* = \underset{\mathbb{F} \in \mathbb{H}, \boldsymbol{B}\mathbb{F}=0}{\operatorname{argmin}} \boldsymbol{J}(\mathbb{F}) := \underset{\mathbb{F} \in \mathbb{H}_a}{\operatorname{argmin}} \boldsymbol{J}(\mathbb{F}). \tag{1}$$

It is well-known that the solution of this constrained optimization problem is closely related to the solution of the saddle point problem expressed in terms of the Lagrangian function $\mathcal{L}$ that maps $\mathcal{L} : \mathbb{H} \times \mathbb{Q}$ defined as $\mathcal{L}(\mathbb{F}, \boldsymbol{p}) := \boldsymbol{J}(\mathbb{F}) + (\boldsymbol{p}, \boldsymbol{B}\mathbb{F})_{\mathbb{Q}}$ with $\mathbb{Q} := \text{range}(\boldsymbol{B})$. It is relatively simple to show that the saddle point of the functional $\mathcal{L}$ is characterized by the solution of the operator equation

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^* \\ \boldsymbol{B} & 0 \end{bmatrix} \begin{Bmatrix} \mathbb{F} \\ \boldsymbol{p} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{b} \\ 0 \end{Bmatrix} \tag{2}$$

where $\boldsymbol{b} = (b^1, \ldots, b^n) \in \mathbb{H}$, the operator $\boldsymbol{A} : \mathbb{H} \to \mathbb{H}$ with $\boldsymbol{A} := \text{diag}(A_i)$, and each entry $A_i : H \to H$ and $b^i \in H$ has the form derived in [78],

$$A_i f := \int_X K_x f(x) \rho_{i,X}(dx) + \lambda^i f, \tag{3}$$

$$b_i := \int K_x f_{\rho_i}(x) \rho_{i,X}(dx). \tag{4}$$

Here $K(\cdot,\cdot)$ is the kernel of the RKH space H, the basis function $K_x(\cdot) := K(x,\cdot)$, λ_i is a regularization parameter for agent i, $\rho_{i,X}$ is the marginal distribution over X of the probability measure ρ_i, and $f_{\rho_i}(x)$ is the regressor function for agent i. Since the measures ρ_i are unknown, the above expression cannot be used in the direct construction of estimates.

To construct the specific estimates, it it more convenient to group observations: samples are collected over some time period and then the update of the function estimate is carried out. Each stage of data collection between the construction of local estimates is referred to as the epoch j. Define m_j to be the number of samples collected during epoch $j > 0$. For notational simplicity it is assumed that the number of samples m_j is the same for all agents. In this scenario observations by each agent i from time step k from m to n are collected in $\boldsymbol{z}^i_{m:n} := \left\{(x^i_k, y^i_k)\right\}_{k=m,n}$, and we denote this block of measurements collected by agent i during epoch j as

$$\boldsymbol{z}^i_{m(j):n(j)} \subset Z^{m_j}.$$

The time step index is $k(j) := m(j) + k$ for $k = 0, 1, \ldots, n(j) - m(j)$, $k(j) \in [m(j), n(j)]$, and $n(j) = m(j) + m_j$.

The local approximation update for agent i that occurs on completion of j^{th} epoch of observations is denoted by the mapping $T_{i,j} := T^i_{k(j)}$. This update operator takes as inputs the collection of observations $\boldsymbol{z}^i_{m(j):n(j)}$ from epoch j, the previous local estimate $f_{i,j-1}$ at the completion of epoch $j-1$ created by agent i, and certain information $I_{j-1,i}$ shared with agent i that is defined at the end of epoch j. The next iteration then sets

$$f_{i,j} = T^i_j(\boldsymbol{z}^i_{m(j):n(j)}, f_{i,j-1}, I_{i,j-1})$$

In this equation j is the number of the epoch, and the function estimate $f_{i,j} := f_{i,k(j)}$ with $k(j)$ the time step between $m(j)$ and $n(j)$. In summary then, learning dynamic proceeds as summarized in Figure 3, which is extracted from [78].

Initialize

Set initial estimates $\mathbb{F}_0 := (f_{1,0}, \ldots, f_{N,0})$
Set initial information $\mathbb{I}_0 := (I_1(\mathbb{F}_0), I_2(\mathbb{F}_0), \ldots, I_N(\mathbb{F}_0))$

For $j \in \mathbb{N}$

Step (j.1) Update the local estimate $f_{i,j}$ using local measurements

$$f_{i,j} = T_j^i(\boldsymbol{z}^i_{m(j):n(j)}, f_{i,j-1}, I_i)$$

Step (j.2) Exchange information

Calculate $I_i(\mathbb{F}_j) = I_i(f_{1,j}, f_{2,j}, \ldots, f_{N,j})$
Transmit $I_i(\mathbb{F}_j) \longrightarrow i$ to node i

Figure 3: Summary of learning process

The specific form of the learning dynamic above is quite general and depends on the nonlinear (local) approximation operators T_j^i as well as the shared information I_i. The iteration is constructed starting with the estimates and information collected during epoch $j-1$,

$$\mathbb{F}_{j-1} := (f_{1,j-1}, f_{2,j-1}, \ldots, f_{N,j-1}),$$

$$\mathbb{I}_{j-1} := (I_{1,j-1}, \ldots, I_{N,j-1}),$$

as well as m_j observations collected just before information exchange is to occur. From the discussion of the consensus learning theory framework in [78], I_j is precisely the estimate of the Lagrange multiplier ı in Equation 2. See [78] for the details. Again let $K(\cdot,\cdot) : X \times X \to Y$ be the kernel of the (RKH) space H and define $K_x := K(x, \cdot)$ to be the associated basis function centered at $x \in X$. The local learning dynamic for agent i is defined by selecting an initial function g_0^i and performing the recursion

$$g_{i,k} = g_{i,k-1} - \gamma_{k-1}\left\{\left(g_{i,k-1}(x_{i,k(j)}) - y_{i,k(j)}\right) K_{x_{k(j)}} + \lambda^i g_{i,k-1}\right. \quad (5)$$

$$\left. + I_{i,j-1} + f_{i,j-1}(x_{k(j)})K_{x_{k(j)}} + \lambda^i f_{i,j-1}\right\} \quad (6)$$

for $k(j) := m(j) + k,\ k = 0, \ldots, n(j) - m(j)$. The the local estimates f_j^i for learning stage j are then the last iteration of this iterative procedure, and local update operators $T_{i,j}$ are defined in the equation

$$f_{i,j} := g_{i,m_j} := T_j^i(\boldsymbol{z}^i_{m(j):n(j)}, f_{i,j-1}, I_{i,j-1})$$

The initial result studied in this grant is the following theorem that characterizes the rate of convergence

Theorem 1 *Let the hypotheses leading to Theorem 7.2 of [78] hold. Fix some constant* $\hat{\kappa} < 1$ *and choose the number of learning steps* m_j *in epoch* j *large enough so that*

$$\frac{log(m(j))}{m(j)} \leq \left(\frac{\hat{\kappa}}{\sqrt{N}\tilde{c}} \frac{2 - \gamma_\lambda M}{2 + \gamma_\lambda M} \|\mathbb{F}_j^{*i}\|_V \right)^{\frac{2s+1}{2}}$$

In this equation M *is the upper bound on the Schur complement operator defined in [78],* $\tilde{c}$ *is the approximation constant in [100], and* s *is the smoothness index of the regressor function in [100]. Then there is a constant* $\eta < 1$ *such that with probability at least* $\prod_j (1 - (Cm_j)^{-\beta}))$ *in* $Z^{Nm(j)}$ *the error in the estimates* $\|\boldsymbol{E}_j^{\boldsymbol{f}}\|_{\mathcal{V}}$ *and the error in the multipliers* $\|\boldsymbol{E}_j^{\boldsymbol{p}}\|_{\S}$ *satisfies the inequality*

$$\kappa \|\boldsymbol{E}_j^{\boldsymbol{f}}\|_{\mathcal{V}} + \|\boldsymbol{E}_j^{\boldsymbol{p}}\|_{\boldsymbol{S}}^2 \leq \eta^{2j} \left(\hat{\kappa} \|\boldsymbol{E}_0^{\boldsymbol{f}}\|_{\mathcal{V}}^2 + \|\boldsymbol{E}_0^{\boldsymbol{p}}\|_{\boldsymbol{S}}^2 \right)$$

for a fixed constant $\kappa > 0$.

The errors $\boldsymbol{E}_j^{\boldsymbol{f}}$ abd $\boldsymbol{E}_j^{\boldsymbol{p}}$ in the above theorem are defined in terms of the operators appearing in the Equation 2 and the theory of approximation of saddle point equations. For the precise definition of the errors $\boldsymbol{E}_j^{\boldsymbol{f}}$ abd $\boldsymbol{E}_j^{\boldsymbol{p}}$, see [78]. The Lagrange multipliers $\boldsymbol{p}$ are used to define the information $\mathbb{I}$ that is exchange among the agents.

The results in the theorem above guarantee that there is a constant $\eta < 1$ such that with overwhelming probability, that is, probability that increases geometrically in terms of the number of samples, the errors $\boldsymbol{E}_j^{\boldsymbol{f}}$ and $\boldsymbol{E}_j^{\boldsymbol{p}}$ decrease geometrically in the number of epochs. As mentioned in the summary of the major accomplishments of the program of research, the error analysis above for the two-stage learning dynamic 1) holds when the samples collected by the agents are IID, and 2) requires that the grid on refinement level j is selected based on the number of samples.

The above concise summary of the rate of convergence is given in detail in [78]. This document has been uploaded to the reporting website and is available online.

6 RKH Space Embedding and Consensus

In view of the standing hypotheses of the two-stage learning dynamic, years two and three largely focussed on the method of RKH space embedding 1) to account for a larger class of sensor network dynamics including generally nonlinear plant models, and 2) to treat (dependent) samples collected along the sensor vehicle network trajectories. The study of the two-stage learning dynamic in Section 5 and the approximation of Koopman operators discussed in the next Section 7 each took roughly one year. Since a longer amount of time was dedicated to this development, its discussion below is somewhat longer.

The remainder of this section is largely an executive summary of the results in [39, 88, 89]. These manuscripts have been uploaded to the reporting website.

6.1 The Class of Systems

The breadth of systems represented in texts that study adaptive estimation of linear and nonlinear ODEs is vast. [55, 62, 61, 67, 63]. In this section we review the class of systems to which RKH embedding for consensus is applied in this report. Both types of systems described below are associated with consensus estimation by nonlinear multiagent systems.

6.1.1 Unknown Nonlinear Agent Dynamics

In the simplest example of this case, we assume that we are given a multiagent family where the dynamics of each agent $i = 1, \ldots, N$ is governed by the nonlinear ODE

$$\dot{x}_i(t) = Ax_i(t) + Bf(x_i(t)) \tag{7}$$

with $x_i(t) \in \mathbb{R}^s$, $A \in \mathbb{R}^{s\times s}$ a Hurwitz matrix, $B \in \mathbb{R}^s$, and $f : \mathbb{R}^s \to \mathbb{R}$. Such a system can arise when the nonlinear dynamics of each agent is the same, but unknown. A collection of identical sensor vehicles will have dynamics governed by equations of this type. We make the following assumptions about our estimation problem:

(A1) Each agent makes a full measurement of its own local state $x_i(t) \in \mathbb{R}^s$ at each time $t \in [0, T]$,

(A2) the matrices A, B are known,

(A3) the function f is unknown, and

(A4) information can be shared only among certain neighbors of each agent.

As discussed in Section 6.3, the information that is shared is described by a communication graph that describes agent connectivity. We are interested in defining rules by which agent i makes a local estimate $\hat{x}_i(t)$ in $\mathbb{R}^s$ of its local state $x_i(t)$ at time t and also constructs a *function estimate* $\hat{f}_i(t) : \Omega \to \mathbb{R}$ with $\hat{f}_i(t) \in H$ of the unknown function f based on its local measurements and specific types of information shared at time $t \in [0, T]$ among the agents. The authors feel that it is the formulation decribed in this report for the construction of the function estimates $\hat{f}_i$ that constitutes the primary innovation in the report, not the manner in estimating the state $\hat{x}_i$. All the function estimates constructed by the agents by the agents are collected in $\hat{\mathbb{F}} := \{\hat{f}_1, \ldots, \hat{f}_N\}^T$, which is an element of $\mathbb{H} := H^N$.

There are many alternatives of Equation 7 that also may be treated by the approach outlined in this report with straightforward modifications. One example is systems of the form

$$\dot{x}_i(t) = f_0(x_i(t)) + f(x_i(t))$$

with a known function $f_0 : \mathbb{R}^s \to \mathbb{R}^s$ and the unknown function $f : \mathbb{R}^s \to \mathbb{R}^s$. These are equally well-suited to our approach. This only requires that we redefine the unknown function as f' with $f'(x) := f_0(x) + f(x) - Ax$, of course. Estimation of f' is tantamount to estimation of f since f_0 and A are known. We find the explicit inclusion of the linear term leads to error equations for the RKHS method that emphasize its close resemblance to the form of equations that arise in the study of persistency of excitation for ODEs. [56, 66, 67]

6.1.2 Unknown Field Approximation by Agent Teams

Suppose now that the dynamics of the vehicle team has a known form, but the goal of the team is to approximate some unknown external field. For purposes of illustration, suppose to start that each agent has the linear dynamics

$$\dot{x}_i(t) = Ax_i(t)$$

for $i = 1, \ldots, N$ and $x_i \in \mathbb{R}^s$. The goal of the agent team is to approximate some scalar, generally nonlinear field variable $g : \mathbb{R}^s \to \mathbb{R}$. There are various ways to express the dynamics of the multiagent system in this case so that the equations have the form similar to that in Equation 8. We only mention one simple case here. We set $y_i(t) = g(x_i(t))$, define $f(x, y) = (\partial g/\partial x)(x) \cdot Ax - \tilde{A}y$ for a known negative constant $\tilde{A}$, and subsequently define the input-output behavior of the i^{th} agent to be

$$\frac{d}{dt}\begin{Bmatrix} x_i(t) \\ y_i(t) \end{Bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \tilde{A} \end{bmatrix}\begin{Bmatrix} x_i(t) \\ y_i(t) \end{Bmatrix} + \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} f(x_i(t), y_i(t)).$$

Estimation of f is tantamount to estimation of $\partial g/\partial x$ since $\tilde{A}$ is known. When convergence of estimates to $\partial g/\partial x$ are established, it can be possible to conclude convergence of estimates of g using compactness results associated with the RKH space. (As an example, if the RKHS space H is equivalent to a Sobolev space like $H^{1,2}$, L^2-boundedness of g and $\partial g/\partial x$ is sufficient to conclude convergence of estimates in the weak topology of H. Such an analysis depends on the specific RKH space H in which $\partial g/\partial x$ lies, and the details exceed the scope of the current report.) Of course, this equation has the form of Equation 7 for the augmented state $x_i' := \{x_i^T, y_i\}^T \in \mathbb{R}^{s+1} := \mathbb{R}^{s'}$. The same assumptions regarding measurements described in (A1) through (A4) in the last section are made in this case also.

6.1.3 The General Form of the Agent Equations

In summary then, the system of ODEs for the collection of agents in this report is taken to be a particular example of the evolution law

$$\dot{X}(t) = \mathbb{A}X(t) + \mathbb{B}\mathbb{F}(X(t)), \qquad X(0) = X_0 \tag{8}$$

with state $X(t) : \{x_1^T(t), \ldots, x_N^T(t)\}^T \in \mathbb{R}^d$, the known Hurwitz system matrix $\mathbb{A} \in \mathbb{R}^{d \times d}$, the known matrix $\mathbb{B} \in \mathbb{R}^{d \times N}$, and the unknown vector-valued function $\mathbb{F} : \mathbb{R}^d \to \mathbb{R}^N$ with $d = Ns$. In the examples above we have

$$\begin{aligned}
\mathbb{A} &:= \text{diag}\{A_i := A \in \mathbb{R}^{s\times s} \mid 1 \le i \le N\} \in \mathbb{R}^{d\times d} \\
\mathbb{B} &:= \text{diag}\{B_i := B \in \mathbb{R}^{s\times 1} \mid 1 \le i \le N\} \in \mathbb{R}^{d\times N} \\
\mathbb{F}(X) &:= \{f_1(x_1), \ldots, f_N(x_N)\} \in \mathbb{R}^N.
\end{aligned}$$

By a change of variables and a redefinition of the function $\mathbb{F}$ as described above, it can be easily shown that other common or interesting equations such as

$$\dot{X}(t) = \mathbb{F}_0(X(t)) + \mathbb{F}(X(t))$$

with $d = N$, $s = 1$, the known function $\mathbb{F}_0 : \mathbb{R}^d \to \mathbb{R}^d$, and the unknown function $\mathbb{F} : \mathbb{R}^d \to \mathbb{R}^d$ can be written as in Equation 8. The model in Equation 8 is selected to be typical, not to be the most general to which the methods of this section can be applied. It has been selected to illustrate clearly the nature of the problems that must be addressed in transitioning from an analysis state consensus in $\mathbb{R}^d$ to the consensus estimation of functions in $\mathbb{H} := H^N$.

6.2 The Adaptive Consensus Estimation Framework

Adaptive estimation and control for systems of both linear and nonlinear ODEs is a mature field. The general theory has given rise to numerous variants of pragmatic algorithms that are catalogued in popular references like [55, 62, 61, 67, 63]. These books and the references therein, as well as a number of other texts, contain excellent discussions of the wide range of classical and modern techniques developed in this area.

To understand how our consensus estimation problem is cast, we begin by summarizing how the RKH embedding strategy works in a general sense. In Section 6.2.1 the "conventional" adaptive estimation problem for nonlinear ODEs is presented, and it is contrasted with RKH adaptive estimation in Section 6.2.2. The introductions in Sections 6.2.1 and 6.2.2 makes clear how the finite dimensional scenario is lifted to an infinite dimensional setting. Finally, adaptive consensus RKH embedding is introduced in Section 6.2.3.

6.2.1 Conventional Adaptive Estimation

We begin with a short summary that places the problem at hand in context with regards to techniques that have become more or less standard for adaptive estimation of systems of nonlinear ODEs. As is well-known, many adaptive estimation algorithms for systems of nonlinear ODEs, such as in Equations 8 construct estimators by making a canonical *a priori* finite dimensional approximation assumption. This assumption gives rise to the usual "linear-in-parameters" types of adaptive estimation schemes. This assumption is so ubiquitous it is now a part of adaptive estimation folklore: it is often taken as the starting point of many analyses or method formulations. Suppose that we must carry out adaptive estimation for the system governed by Equations 8. Generically, it is assumed that estimators are constructed in the finite dimensional subspace $\mathbb{H}_K = \text{span}\{\Phi_k \mid 1 \leq k \leq K\}$. The approximation assumption states that there is a compact subset Ω containing the orbit of the governing equations and that for some unknown constants $\alpha^* := \{\alpha_k^*\}_{k=1,\ldots,K} \subset \mathbb{R}^K$ we have

$$\|\mathbb{F}(X) - \mathbb{F}_K^*(X)\|_{C(\Omega)} := \sup_{X\in\Omega} \left\| \mathbb{F}(X) - \sum_{1<k\leq N} \alpha_k^* \Phi_k(X) \right\|_{\mathbb{R}^K} \leq \epsilon$$

with ϵ a maximum allowable error that is small. With this assumption in place, one of the simplest estimator equations is taken as

$$\dot{\hat{X}}(t) = \mathbb{A}\hat{X}(t) + \mathbb{B}\sum_{i=k}^{K} \hat{\alpha}_k(t)\Phi_k(X(t)). \tag{9}$$

where $\hat{X}(t)$ is an estimate of the state $X(t)$ and $\hat{\mathbb{F}}_K(\cdot) := \sum_{1\leq k\leq K} \hat{\alpha}(t)\Phi_k(\cdot)$ is a time-varying estimate of the unknown function $\mathbb{F}$.

The full specification of an adaptive estimation algorithm then consists of an estimator such as in Equation 9 and a learning law for the time-varying parameters $\{\hat{\alpha}_k(t)\}_{k\leq K} \in \mathbb{R}^K$. One common choice of learning law, among many possibilities documented in classical texts, is the gradient learning law. We consider this learning law as iconic since it is so widespread and is often the first method introduced in introductory texts. It has a simple extension to the *scalar* RKH space setting in [39, 40, 77] and to the problems in this report. To assess the convergence of the associated adaptive estimation problem, it is standard to define the state error $\tilde{X}(t) := X(t) - \hat{X}(t)$ and parameter estimate error $\tilde{\alpha}(t) := \alpha^* - \hat{\alpha}(t)$. Suppose for the moment that $\mathbb{F} \in \mathbb{H}_K$. With the choice of the gradient learning law, the equations governing the error in the estimates can be written in the form

$$\frac{d}{dt}\begin{Bmatrix} \tilde{X}(t) \\ \tilde{\alpha}(t) \end{Bmatrix} = \begin{bmatrix} \mathbb{A} & \mathbb{B}\Phi(X(t)) \\ -\Gamma_d \Phi^T(X(t))\mathbb{B}^T\mathbb{P} & 0 \end{bmatrix} \begin{bmatrix} \tilde{X}(t) \\ \tilde{\alpha}(t) \end{bmatrix} \tag{10}$$

with $\Phi := [\Phi_1, \Phi_2, \ldots, \Phi_K] \in \mathbb{R}^{N\times K}$ and $\mathbb{P} \in \mathbb{R}^{d\times d}$ is the solution of an associated Lyapunov equation discussed following Equation **??**. These equations define an evolution in $\mathbb{R}^{d\times K}$. It is widely known that study of such equations is an important ingredient in understanding many adaptive estimation problems in the Euclidean space $\mathbb{R}^{d+K}$. Conditions that guarantee the stability of solutions and convergence of the state error $\tilde{X}(t) \to 0$ are found in a number of places, including the texts [61, 67, 66]. It is also well-known that the convergence of the parameter error $\tilde{\alpha}(t) \to 0 \in \mathbb{R}^K$ is typically more difficult to establish, and conditions of persistence of excitation as in Definition 1 below have been developed that guarantee such parameter convergence [66, 67, 56, 68].

Generally speaking, when as in the above discussion the unknown function $\mathbb{F}$ is known to be contained in the subspace $\mathbb{H}_K$, then the classical persistency of excitation condition implies that the zero solution of the error equations is asymptotically stable, and in particular $\tilde{\alpha}(t) \to 0$. If it is only known that the approximation condition holds for some $\epsilon > 0$, with $\mathbb{F} \notin \mathbb{H}_K$, then an ultimate boundedness condition holds for the parameter errors. The ultimate bound is typically shown to be $O(\epsilon)$. A good overview of these results can be found in [55].

It should be noted that once a basis is chosen and the approximation assumption is made in this strategy, in some sense any information about the topology on the space of functions in which the unknown function $\mathbb{F}$ resides is lost. Since any norm on the finite dimensional space $\mathbb{H}_K$ is equivalent, $\tilde{\alpha}(t) \to 0$ implies that $\|\tilde{\mathbb{F}}_K(t)\|_{\mathbb{H}_K} \to 0$ where $\tilde{\mathbb{F}}_K := \mathbb{F}_K^* - \hat{\mathbb{F}}_K$. For general $\mathbb{F} \notin \mathbb{H}_K$, the question as to whether $\lim_{K\to\infty} \hat{\mathbb{F}}_K \to \mathbb{F}$ in some appropriate or common function space norm $\mathbb{H}$ often, or perhaps usually, remains unanswered. In particular, it is not known that $\|\mathbb{F}_K - \mathbb{F}\|_{C(\Omega)} \to 0$ as $K \to \infty$.

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

To understand how our consensus estimation problem is formulated, we begin by summarizing how the RKH embedding strategy replaces the finite dimensional, approximate governing Equations 10 by an infinite dimensional DPS. The method of RKH space embedding interprets the unknown function f as an element of the RKH space H, without any *a priori* selection of the particular finite dimensional subspace used for estimation of the unknown function. The corresponding equation of the the plant, estimator, and learning laws are selected to be

$$\dot{X}(t) = \mathbb{A}X(t) + \mathbb{B}\mathbb{E}_{X(t)}\mathbb{F}, \tag{11}$$

$$\dot{\hat{X}}(t) = \mathbb{A}\hat{X}(t) + \mathbb{B}\mathbb{E}_{X(t)}\hat{\mathbb{F}}(t), \tag{12}$$

$$\dot{\hat{\mathbb{F}}}(t) = \Gamma^{-1}(\mathbb{B}\mathbb{E}_{X(t)})^*\mathbb{P}(X(t) - \hat{X}(t)), \tag{13}$$

where as before $X, \hat{X} \in \mathbb{R}^d$. But here $\mathbb{F}$ and $\hat{\mathbb{F}}(t) \in \mathbb{H} := H^N$, $E_x : H \to \mathbb{R}^s$ is the evaluation operator on H with $E_x f := f(x) \in \mathbb{R}^s$ for each $f \in H$ and $x \in \mathbb{R}^s$, and $\mathbb{E}_X : \mathbb{H} \to \mathbb{R}^d$ is the matrix of evaluation operators $\mathbb{E}_X := \text{diag}\{E_{x_i} \mid 1 \leq i \leq N\}$ for any $X := \{x_1, \ldots, x_N\} \in \mathbb{R}^d$. In addition $\mathbb{P} \in \mathbb{R}^{d\times d}$ is a symmetric, positive definite, real matrix solution of an associated matrix Lyapunov equation

$$\mathbb{P}\mathbb{A} + \mathbb{A}^T\mathbb{P} = -\mathbb{Q}$$

for some prescribed symmetric, positive definite, real matrix $\mathbb{Q}$. In our model problem, the matrices $\mathbb{P}$ and $\mathbb{Q}$ are assumed to have the same general block structure as $\mathbb{A}$. The operator $\Gamma := \text{diag}\{\Gamma_i : 1 \leq i \leq N\} \in \mathcal{L}(\mathbb{H}, \mathbb{H})$ is a block diagonal, self-adjoint, positive definite, linear operator. The error equation is then given by

$$\frac{d}{dt}\begin{Bmatrix}\tilde{X}(t)\\ \tilde{\mathbb{F}}(t)\end{Bmatrix} = \begin{bmatrix}\mathbb{A} & \mathbb{B}\mathbb{E}_{X(t)}\\ -\Gamma^{-1}(\mathbb{B}\mathbb{E}_{X(t)})^*\mathbb{P} & 0\end{bmatrix}\begin{Bmatrix}\tilde{X}(t)\\ \tilde{\mathbb{F}}(t)\end{Bmatrix}, \tag{14}$$

which defines a nonautonomous evolution on $\mathbb{R}^d \times \mathbb{H}$. These equations are a generalization to vector-valued unknown functions $\mathbb{F} \in \mathbb{H}$ of the scalar-valued unknown function that is studied in [39, 79] . We will find that it is informative and useful to study this case as a special case of the more general scenario of consensus adaptive estimation in $\mathbb{R}^d \times \mathbb{H}$ discussed next.

6.2.3 Consensus Estimation in $\mathbb{R}^d \times \mathbb{H}$

If we make the standing assumptions (A1) through (A4) summarized above, and the block diagonal assumptions described for $\mathbb{A}, \mathbb{B}, \mathbb{P}, \Gamma$ above, it turns out that we can view the Equations 12 as governing the errors in a *completely decentralized* consensus estimation problem. That is, each agent i uses observations of only its own state $x_i(t)$ at time $t > 0$ to construct estimates $\hat{f}_i(t)$ of the unknown function f. If we can show that $\tilde{\mathbb{F}}(t) \to 0$ as $t \to \infty$, each agent reaches the consensus estimate $\hat{f}_i(t) \to f(t)$ as $t \to \infty$.

Of course, this is not the usual sense in which consensus estimation is studied. Rather, it is characteristic of consensus estimation that some exchange of information is used to construct estimates, controllers, or system evolution laws. In this report we study the case when information regarding the estimates of the unknown function are shared among neighbors in the agent network. The learning law in Equation 13 is replaced with

$$\dot{\hat{\mathbb{F}}}(t) = \Gamma^{-1}\mathbb{E}^*_{X(t)}\mathbb{B}^T\mathbb{P}(X(t) - \tilde{X}(t)) - \mathcal{L}\hat{\mathbb{F}}(t)$$

with $\mathcal{L} : \mathbb{H} \to \mathbb{H}$ the induced operator on vector-valued functions generated by matrix multiplication by $\mathcal{L} \in \mathbb{R}^N$. Here the real, symmetric, positive semidefinite matrix $\mathcal{L}$ is the Laplacian matrix of an undirected, connected graph that prescribes communication and neighbors. Because of the assumed block structure of the matrices $\mathbb{B}, \mathbb{P}$, and $\mathbb{E}_X$, the update in the learning law for agent i of the function $\hat{\mathbb{F}}(t)$ depends only on its own state $\hat{x}_i$ and and the estimates $\hat{\mathbb{F}}_j$ generated by the neighbors $j \in \mathcal{N}(i)$ of agent i defined via the connectivity graph. The construction and basic properties of the graph Laplacian, as well as its role in multiagent consensus in $\mathbb{R}^N$ is discussed in Section 6.3. With the new learning law for consensus estimation, the error equations take the form

$$\frac{d}{dt}\begin{Bmatrix} \tilde{X}(t) \\ \tilde{\mathbb{F}}(t) \end{Bmatrix} = \begin{bmatrix} \mathbb{A} & \mathbb{B}\mathbb{E}_{X(t)} \\ -\Gamma^{-1}(\mathbb{B}\mathbb{E}_{X(t)})^*\mathbb{P} & -\mathcal{L} \end{bmatrix} \begin{Bmatrix} \tilde{X}(t) \\ \tilde{\mathbb{F}}(t) \end{Bmatrix}. \tag{15}$$

6.3 Definitions of Consensus

This report studies estimation with consensus in vector-valued RKH spaces, which requires a subtle generalization of some well-known results on consensus estimation in the Euclidean space $\mathbb{R}^d$. The background theory on consensus estimation in $\mathbb{R}^d$ can be found in many places including the recent book [108] or the highly cited [85]. Just over the past two years, efforts such as [84, 86] illustrate the continued refinement of the basic theory in $\mathbb{R}^d$. With respect to this report, the development of Lyapunov-based analyses of consensus estimates in $\mathbb{R}^d$ by agent teams as in [83, 82, 43] are similar in spirit to our analysis in H^d.

6.3.1 Consensus in $\mathbb{R}^N$

In this section we summarize some of the well-known, graph-theoretic constructions of communication topology in multiagent systems. We make no effort to address the most general analyses: these can include consideration of time-varying network connectivity, time-varying channel capacity, or delays, among other refinements. Instead, we focus on the most transparent case of the graph-theoretic analysis so as to emphasize how the theory on $\mathbb{R}^N$ must be extended to study multiagent consensus estimation in the function space $\mathbb{H} := H^N$. It is, of course, of interest to further extend the consensus analysis on $\mathbb{H}^N$ to the situation where connectivity changes with time, but we leave this task for future study.

A (finite) graph G is defined in terms of a finite set of nodes or vertices $\mathcal{V} := \mathcal{V}(G) = \{v_1, \ldots, v_n\}$ and edges $\mathcal{E} := \mathcal{E}(G) = \{e_1, \ldots, e_m\} \subset V \times V$. In this report we denote the vertex $v_i \sim i$ and the edge $e_k \sim (i, j)$. Intuitively, in a *directed graph*, each edge $e_k = (i, j)$ is viewed as an arrow from node i to j that denotes that node i can send information to node j. A node i is a neighbor of node j whenever $(i, j) \in \mathcal{E}$. In an *undirected graph*, $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. That is, every connection between a pair of nodes is bidirectional.

There are various techniques to give a precise description of the connectivity of a graph. One common description defines the adjacency matrix $\mathcal{A} := [\mathcal{A}_{ij}]$ as

$$\mathcal{A}_{ij} = \begin{cases} 1 & \text{if node } i \text{ is a neighbor of node } j, \\ 0 & \text{otherwise.} \end{cases}$$

Alternatively, it is possible to describe connectivity in terms of the edge incidence matrix $\mathcal{B} := [\mathcal{B}_{ij}]$ defined as

$$\mathcal{B}_{ij} = \begin{cases} +1 & \text{if node } i \text{ is a neighbor of node } j, \\ -1 & \text{if node } j \text{ is a neighbor of node } i, \\ 0 & \text{otherwise.} \end{cases}$$

For an undirected graph we always have

$$\mathcal{L} = \mathcal{B}^T \mathcal{B} = \mathcal{D} - \mathcal{A}$$

with $\mathcal{L}$ defined to be the Laplacian of the graph and $\mathcal{D} := \text{diag}(d_i)$ the degree matrix. The degree matrix $\mathcal{D} = [d_{ij}]$ is a diagonal matrix with the degree $d_i = d_{ii}$ of vertex i, which is the number of neighbors of node i.
In the general case, in which connectivity can be time-varying, the matrices $\mathcal{L}, \mathcal{A}, \mathcal{D}, \mathcal{B}$ can be time-varying. In this report we only consider the case when these matrices are constant.

A graph G is connected if there is a path of edges that connects every pair of its vertices. The spectral properties of the Laplacian matrix $\mathcal{L}$ provides much useful information for understanding the connectivity of the graph. For an undirected graph, the matrix $\mathcal{L} := \mathcal{B}^T \mathcal{B}$ is real and symmetric, so its eigenvalues are real and nonegative. Order the eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$ from minimum to maximum modulus. It is always that case that the vector $\mathbf{1}_N := \{1, 1, \dots, 1\}^T$ is an eigenvector of the Laplacian matrix $\mathcal{L}$ associated to the eigenvalue $\lambda_1 = 0$. It is known that the graph G is connected if and only if $\lambda_2 > 0$. Define the matrix

$$\mathcal{P} := \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T,$$

which is the orthonormal projection onto the eigenspace spanned by the eigenvector $\mathbf{1}_N$. Following convention, we define the agreement (or consensus) space $\mathcal{H}_a \subset \mathbb{R}^N$ and the disagreement space $\mathcal{H}_d \subset \mathbb{R}^N$ as

$$\mathcal{H}_a := \text{ran}(P) = \ker(\mathcal{L}),$$
$$\mathcal{H}_d := \text{ran}(\mathcal{L}) = \ker(P),$$

respectively.

Suppose that $x(t) \in \mathbb{R}^N$ is a semiflow defined for all $t \in [0, \infty)$. We say that the state converges to consensus as $t \to \infty$ when $d(x(t), \mathcal{H}_a) \to 0$. In a conventional and iconic application of these ideas, consider the linear ordinary differential equation

$$\dot{x}(t) = -\mathcal{L}x(t)$$

subject to the intial condition $x(0) = x_0$. If $\mathcal{L}$ is the Laplacian matrix generated from an undirected, connected graph, it is now a classical result that the state converges to the agreement space $\mathcal{H}_a \subset \mathbb{R}^N$ as $t \to \infty$.

We also note that it is not difficult to show using a spectral decomposition that that we can define a norm $\|\cdot\|_{\mathbb{R}^N_G}$ as

$$\|v\|^2_{\mathbb{R}^N_G} := ((\mathcal{P} + \mathcal{L})v, v)_{\mathbb{R}^N} \approx \|v\|^2_{\mathbb{R}^N}$$

that is equivalent to the usual Euclidean norm on $\mathbb{R}^N$. We will use an analogous construction on the RKH product space $\mathbb{H} := H^N$.

6.3.2 Consensus in $\mathbb{H} := H^N$

As opposed to the last section, the consensus analysis in this report treats the study of trajectories $t \mapsto \mathbb{F}(t) \in H^N := \mathbb{H}$ *of functions* for $t \geq 0$. Essentially, this requires that we "lift" the analysis of the last section over a finite dimensional space $\mathbb{R}^N$ to the infinite dimensional case. Much of this process is intuitive and easily carried out, but some nuances require a bit of care in the infinite dimensional case. The process is based on the construction operators on $\mathbb{H}$ from the matrices such as $\mathcal{L}, \mathcal{D}, \mathcal{B}, \mathcal{A}$.

In this section let $\mathcal{P}$ and $\mathcal{L}$ be the matrices associated with a finite, time-invariant, undirected, connected graph G. We denote $N = \#(V(G))$ and $M := \#(E(G))$. Using the same notation, we define the operators $\mathcal{P} : \mathbb{H} \to \mathbb{H}$, $\mathcal{L} : \mathbb{H} \to \mathbb{H}$, and $\mathcal{B} : H^M \to \mathbb{H}$ via

$$(\mathcal{L}f)(x) := \mathcal{L} \cdot f(x)$$
$$(\mathcal{P}f)(x) := \mathcal{P} \cdot f(x)$$
$$(\mathcal{B}f)(x) := \mathcal{B} \cdot g(x)$$

for all $x \in \Omega$, $f \in \mathbb{H} := H^N$ and $g \in H^M$. We likewise define the agreement and disagreement spaces $\mathbb{H}_a$ and $\mathbb{H}_d$ of functions, respectively, generalizing the consensus space $\mathcal{H}_a \subset \mathbb{R}^N$ and disagreement space $\mathcal{H}_d \subset \mathbb{R}^N$, from the identities

$$\mathbb{H}_a := \overline{\text{ran}(\mathcal{P})} = \overline{\{f \in \mathbb{H} \mid f = \mathcal{P}g, g \in \mathbb{H}\}}$$
$$\mathbb{H}_d := \overline{\text{ran}(\mathcal{L})} = \overline{\{f \in \mathbb{H} \mid f = \mathcal{L}g, g \in \mathbb{H}\}}$$

The closure $\overline{(\cdot)}$ in the above equations is required: in the finite dimensional situation it is automatically true that $\text{ran}(\mathcal{P}) = \overline{\text{ran}(\mathcal{P})}$, but this is not generally true in the infinite dimensional case. The definitions above for $\mathbb{H}_a, \mathbb{H}_d \subseteq H^N$ should be carefully compared to those of the finite dimensional counterparts $\mathcal{H}_a, \mathcal{H}_d \subseteq \mathbb{R}^N$, respectively. We note several properties of these operators and spaces that will be essential to the analysis that follows.

Theorem 2 *Let G be a finite, time-invariant, undirected, connected graph. The following hold :*

1. *The operators $\mathcal{L}, \mathcal{P}, \mathcal{B}$ are bounded, self-adjoint operators.*
2. *The inner product given by*

$$(f, g)_{\mathbb{H}_G} := (\mathcal{P}f, f)_{\mathbb{H}} + (\mathcal{L}f, f)_{\mathbb{H}}$$

   is equivalent to the usual inner product on $\mathbb{H}$.
3. *We have*

$$\mathbb{H}_a := \overline{ran(\mathcal{P})} = ker(\mathcal{L}),$$
$$\mathbb{H}_d := \overline{ran(\mathcal{L})} = ker(\mathcal{P}).$$

4. *The vector-valued RKH space $\mathbb{H}$ admits the orthogonal decomposition*

$$\mathbb{H} = \mathbb{H}_a \oplus \mathbb{H}_d.$$

See [89] for the proof of this theorem. It follows, for the most part, by lifting the definitions for conventional consensus to the case of vector-valued functions. Some of the results require the use of limits and closures in infinite dimensional spaces, but otherwise the result is intuitive and expected.

6.4 Persistence of Excitation in the RKH Spaces

In this section we introduce a novel form of the PE condition that is appropriate for the study of the multiagent system. This section has implications beyond our target application problem and illustrates some general relationships among PE systems and positive limit sets.

Recall that [53] a continuous semiflow or semidynamical system can be defined in terms of a continuous semigroup $\{S(\tau)\}_{\tau \geq 0}$ that acts on the complete metric space $(\mathcal{X}, d_{\mathcal{X}})$. In the problem at hand it is assumed that true solutions of the original governing ODEs in Equation 8 generate a semigroup in the usual way via $X(t) = S(t)X_0$. The positive orbit starting at X_0 of the true system is defined to be the set

$$\Gamma^+(X_0) := \cup_{\tau \geq 0} S(\tau)X_0 \subseteq \mathcal{X}.$$

The positive limit set $\omega^+(X_0)$ associated with the initial condition X_0 is defined to be

$$\omega^+(X_0) := \cap_{t \geq 0} \overline{\cup_{\tau \geq t} S(\tau)X_0},$$

which is equivalently expressed as

$$\omega^+(X_0) = \left\{ y \in \mathcal{X} \mid \exists t_k \to \infty \text{ such that } \lim_{k\to\infty} S(t_k)X_0 \to y. \right\}$$

Notions of Persistence of Excitation

We will use these definitions to gain insight about PE conditions and note how general the notions of PE in this definition are. The following is one of several equivalent, alternate forms of the conventional persistency of excitation conditions for the error in the conventional ODEs in Equations 10.

Definition 1 *The trajectory $X_0 \mapsto S(t)X_0 := X(t)$ is persistently exciting over $\mathbb{R}^d \times \mathbb{R}^K$ if there are positive constants $\gamma_1, \gamma_2, T, \Delta > 0$ such that*

$$\gamma_1 \|\alpha\|^2_{\mathbb{R}^K} \leq \int_t^{t+\Delta} \left(\Phi^T(X(\tau))\Phi(X(\tau))\alpha, \alpha\right)_{\mathbb{R}^K} d\tau \leq \gamma_2 \|\alpha\|^2_{\mathbb{R}^K}$$

for all $t \geq T$ and $\alpha \in \mathbb{R}^K$.

The generalization in this report introduces the definition of persistent excitation relative to an *index set* $\Omega \subseteq \mathcal{X}$ and an associated subspace $H_\Omega \subseteq H_{\mathcal{X}}$ of functions.

Definition 2 *The trajectory $X_0 \mapsto S(t)X_0 := X(t)$ is persistently exciting with respect to the index set $\Omega \subseteq \mathcal{X}$ and function space H_Ω if there are positive constants $\gamma_{\Omega,1}, \gamma_{\Omega,2}, T_\Omega, \Delta_\Omega > 0$ such that*

$$\gamma_{\Omega,1} \|\mathbb{F}\|^2_{\mathbb{H}_\Omega} \leq \int_t^{t+\Delta_\Omega} (\mathbb{E}^*_{X(\tau)} \mathbb{E}_{X(\tau)} \mathbb{F}, \mathbb{F})_{\mathbb{H}_\Omega} d\tau \leq \gamma_{\Omega,2} \|\mathbb{F}\|^2_{\mathbb{H}_\Omega}$$

for all $t \geq T_\Omega$ and $f \in \mathbb{H}_\Omega$.

The newly introduced definition of persistency, one that is suitable for the study of consensus estimation of functions, is the following.

Definition 3 *The trajectory $X_0 \mapsto S(t)X_0 := X(t)$ is persistently exciting with respect to the domain $\Omega \subseteq \mathcal{X}$ and function space $R_\Omega(H_\mathcal{X})$ if there are positive constants $\gamma_{\Omega,1}, \gamma_{\Omega,2}, T_\Omega, \Delta_\Omega > 0$ such that*

$$\gamma_{\Omega,1}\|\mathbb{F}\|^2_{R_\Omega(\mathbb{H}_\mathcal{X})} \leq \int_t^{t+\Delta_\Omega} (\mathbb{E}^*_{X(\tau)}\mathbb{E}_{X(\tau)}\mathbb{F}, \mathbb{F})_{R_\Omega(\mathbb{H}_\mathcal{X})} d\tau \leq \gamma_{\Omega,2}\|\mathbb{F}\|^2_{R_\Omega(\mathbb{H}_\mathcal{X})}$$

for all $t \geq T_\Omega$ and $\mathbb{F} \in R_\Omega(\mathbb{H}_\mathcal{X})$.

We end this section with a result that can serve as the kernel of the argument made to establish function (parameter) convergence for PE system in a RKH space.

Theorem 3 *Let $H_\mathcal{X}$ be a RKH space of real-valued functions over the set $\mathcal{X}$ generated by a kernel $\mathfrak{K}$, $\mathbb{H}_\mathcal{X} := H^N_\mathcal{X}$, and the function $G \in L^\infty((0,\infty), \mathbb{H}_\mathcal{X})$. If the trajectory $t \mapsto S(t)X_0 := X(t)$ is persistently exciting in the sense of Definition 3 and $\mathfrak{K}(x_i(\cdot), x_i(\cdot)) \in L^1((0,\infty), \mathbb{R})$ for each $i = 1, \dots, N$, then we have*

$$\lim_{t\to\infty} \|\mathbb{G}(t)\|_{\mathbb{H}_\Omega} = 0.$$

This theorem is proven in [89], which has been uploaded to the report website.

6.5 Primary Result for Consensus Estimation

With the foregoing background, the primary convergence result for the RKH embedding approach to consensus estimation is given in the following.

Theorem 4 *Let H be $H_\mathcal{X}, H_\Omega$, or $R_\Omega(H_\mathcal{X})$ and $\mathbb{H} := H^N$. Suppose that there is a unique solution $X \in C([0,\infty), \mathbb{R}^d)$ of the original Equation 11, and a unique mild solution $(\tilde{X}, \tilde{\mathbb{F}}) \in C([0,\infty), \mathbb{R}^d \times \mathbb{H})$ of the consensus error Equations 15. Then we have*

$$\lim_{t\to\infty} \tilde{X}(t) = 0,$$
$$\lim_{t\to\infty} d(\tilde{\mathbb{F}}(t), \mathbb{H}_a) = 0.$$

If the trajectory $X_0 \mapsto S(t)X_0 := X(t)$ is persistently exciting in the sense of Definition 2 or 3, and $\mathfrak{K}(x_i(\cdot), x_i(\cdot)) \in L^1((0,\infty), H)$, then

$$\lim_{t\to\infty} \tilde{\mathbb{F}}(t) = 0.$$

As briefly touched on in the overview of major accomplishments, this theorem establishes convergence of solutions of a DPS, rather than simply the convergence of states in $\mathbb{R}^d$. The approach provides a common and unifying foundation for many of the classical or conventional approaches to adaptive estimation for ODEs that evolve in $\mathbb{R}^d$, many of which appear in references such as [55, 62, 61, 67, 63]. The idea of persistency as an idea about positive definiteness in a RKH space over a domain Ω is also novel and extends many of the existing approaches to proving convergence of parameter estimates. The details and proofs of the discussion in this section are lengthy, and the reader is again referred to [88, 89, 39, 79] for the details of the analysis. References [88, 89, 39] have been uploaded to the report website and [79] is also available online.

7 Approximation Spaces and Koopman Theory

The two-state learning dynamic in Section 5 made use of priors defined as membership in approximation spaces to guarantee rates of convergence of estimates constructed from IID samples. In Section 6, a general approach to the analysis of consensus estimates in a RKH space $\mathbb{H}$ has been discussed. This latter strategy enables more general and realistic models of the agent dynamics and employs samples collected along the trajectory of the sensor vehicle team. Still, the RKH method as summarized in Section 6 does not employ approximation spaces, at least in its current form, to assess rates of convergence.

In contrast to the situation in Section 5, where the states are assumed to be IID, we hypothesize a discrete and dependent state process in this section. This section likewise establishes a foundation that will be used in future research to incorporate approximation space priors in the RKH embedding technique for consensus estimation. One common and well accepted model, among many possibilities outlined in [90], holds that the state $X := \{x_1, \ldots, x_N\}$ is given by the stochastic recursion

$$X_{n+1} = \mathbb{F}(X_n, \lambda_n)$$

for some unknown, generally nonlinear function $\mathbb{F} \in \mathbb{B}$ with $\mathbb{B}$ a Banach space and $\{\lambda_n\}_{n\geq 0}$ a driving stochastic process. Note that with this model the states $\{X_n\}_{n\geq 0}$ are not generally independent. In fact, the nonlinear recursion is assumed to define a Markov chain with transition probability kernel $\mathbb{P}(A, X)$ that is the probability that the next step in the chain is in the measurable set A given that the current state is X. This model can drive the approximation of another unknown, nonlinear function $\mathbb{G}$ via noisy observations from a measurement defined as $Y_n = \mathbb{G}(X_n) + w_n$ with $\{w_n\}_{n\geq 0}$ another random process. As above, many variants of the definition of the governing dynamics and measurements are possible in addition to the above motivational example. These are discussed in detail in the introduction to references [89] and [90].

Koopman theory is an operator theoretic approach to the study of dynamical systems. This approach studies the asymptotics of a dynamical system by studying the Koopman operator $\mathcal{U}$ and Perron-Frobenius operator $\mathcal{P}$ that are defined as

$$(\mathcal{P})\nu(dy) := \int_\Omega \mathbb{P}(dy, x)\nu(dx),$$
$$(\mathcal{U}f)dy := \int_\Omega \mathbb{P}(dy, x)f(y)$$

for a measure ν on the domain Ω and a function $f : \Omega \to \mathbb{R}$. To see how Koopman theory plays a role in our problem of estimation, suppose for simplicity that $\mathbb{P}(A, x) := \delta_{\mathbb{F}(X)}(A)$, which is the transition probability kernel for the deterministic recursion $X_{n+1} = \mathbb{F}(X_n)$. Then we have

$$(\mathcal{U}\mathbb{G})(X) := (\mathbb{G} \circ \mathbb{F})(X) = \mathbb{G}(\mathbb{F}(X)).$$

We see that approximations of the Koopman operator $\mathcal{U}$ can be used to induce approximations of the function $\mathbb{G} \circ \mathbb{F}$. But this is our original goal of approximating an unknown function from samples. If define measurements as $Y := \mathbb{G}(\mathbb{F}(X))$, the estimation process can use the family of observations $\{(X_n, Y_n)\}_{n \geq 0}$ to estimate the function $\mathbb{G}$.

As in the last two sections, the research monograph [90] describes numerous specific cases, some far more general than that briefly summarized here, that derive approximation rates for Koopman operators that depend on a family of linear approximation spaces $A^{r,s}(\mathbb{B})$.

The research manuscript in [90] gives conditions that guarantee that by constructing estimates $\mathcal{U}_{j,z}$ of the Koopman operator $\mathcal{U}$ using m IID samples described above and basis functions defined over a fixed grid having mesh resolution j, errors in expectation are bounded by expressions such as

$$\mathbb{E}_{\mathbb{P}^m}\left(\|\mathcal{U}f - \mathcal{U}_{j,z}f\|_U^2\right) \lesssim \lambda_j^r + \frac{j}{m}. \tag{16}$$

for all functions in the spectral approximation space $A_\lambda^{r,2}$ and r is a measure of smoothness of the unknown function $\mathcal{U}f$. Here $\{\lambda_j\}_{j \geq 1}$ is the sequence of positive eigenvalues converging to zero associated with an eigenfunction basis that defines the approximation space $A_\lambda^{r,2}$. This measure of expected error exhibits the the bias and variance tradeoff that is familiar from modern approaches for approximation theory and distribution free learning theory. Again, this result has no precedent in applications to consensus estimation nor in recent treatments of approximation in Koopman theory.

Of all of the methods for constructing approximations in Koopman theory, one of the most popular techniques is the enhanced domain decomposition method (EDMD). The foundations of this approach are discussed in [94, 95, 96]. Essentially, these works establish the convergence estimates of the Koopman operator *in a fixed finite dimensional space* of functions. We obtain a much sharper estimate of the rate of convergence. Using the strategy outlined in [90], we find that

$$\mathbb{E}_{\nu^m}\left(\|\mathcal{U}f - \mathcal{U}_{j,z}^{edmd}f\|_U\right) \lesssim \left(\frac{\log m}{m}\right)^{2r/(2r+1)}$$

for all functions that reside in the approximation space $A^{r,2}$. Here $\mathbb{E}_{\nu^m}(\cdot)$ is the expectation over the samples $z := \{(x_i, \tilde{y}_i)\}_{i \leq m}$ with respect to the product measure ν^m on Z^m. It is known that this estimate is the best possible rate of convergence (except for the $\log m$ term), and therefore it is referred to as a semi-optimal bound. [97]

In addition, this estimate can be extended to certain discrete nonlinear models of evolution that generate samples along a trajectory, one of the primary objectives of this research grant. As discussed in more detail in [106], an analogous result follows for some strongly mixing Markov chains. By modifying the rule derived in [100, 101] to employ the effective number of samples $e(m)$ as in Reference [106], it follows that

$$\mathbb{E}_{\mathbb{P}^m_{\{z\}}} \left(\|\mathcal{U}f - \mathcal{U}^{edmd}_{j,z} f\|_U \right) \lesssim \left(\frac{\log e(m)}{e(m)} \right)^{2r/(2r+1)}.$$

In this equation the expectation $\mathbb{E}_{\mathbb{P}^m_{\{z\}}}$ is the expectation over the first m steps of the Markov chain.

All of the results stated above, which extend and improve existing rates of convergence for approximation of Koopman operators (and therefore for approximations of our category of dynamical systems) are discussed in full detail in [90]. This long research manuscript has been uploaded to the report website.

8 Hardware Development

This research program worked on the further development of the the Terrestrial Unmanned Robots for Teamed Learning and Exploration (TURTLEs), under the direction of co-investigator Alex Leonessa, in addition to the theoretical work summarized in the last three sections. The TURTLEs consist of four identical ground robots, shown in Figure 4 Each unit has been completed to the same

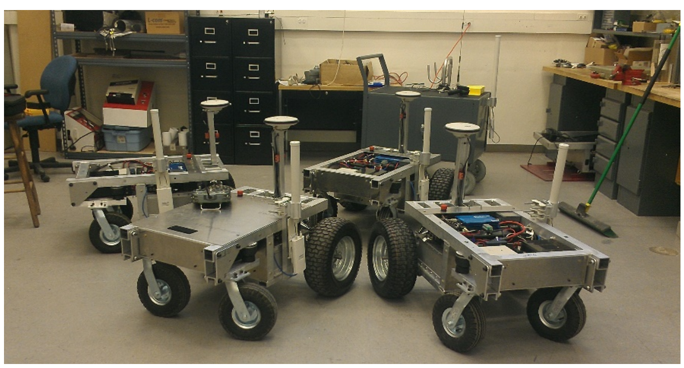


Figure 4: The TURTLEs Autonomous Vehicles Created in the MS Thesis research by Adam Shoemaker, 2015

degree, described in the following paragraphs. Each vehicle has likewise been tested in both in and out door environments on a variety of terrain. Figure 1 – Terrestrial Unmanned Robots for Teamed Learning and Exploration The base design includes a differentially steered setup with two brushless motors located in the rear. Each motor can continually produce 139 lbf-in of torque with peaks up to 257 lbf-in. This corresponds to a continuous draw of over 500 W per motor and speeds of up to 10 mph. The motors are controlled via a RoboteQ brushless controller that allows for RS-232, USB, and PWM signals to modulate speed. The motors and controller are powered by 8 LiFePO4 battery cells in series. These cells have a capacity of 100 A-hr and typically allow for field operation time in excess of 10 hours. The battery management system facilitates fully automated charging, allowing for hands off upkeep during down time. The

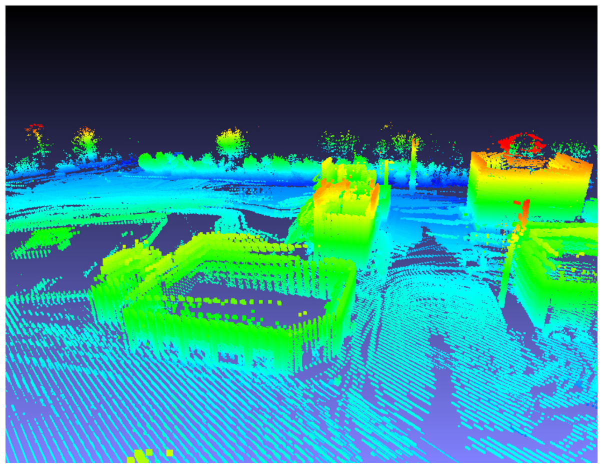


Figure 5: Site Map of Terrain Developed in Collaboration between Adam Shoemaker, and Professors Alex Leonessa and Kevin Kochersburger

batteries are also interfaced with an onboard voltage regulator that distributes 3.3V, 5V, and 12V to the computers and peripheral devices. This configuration allows for the quick implementation of new equipment to the vehicle without power configuration changes. This power supply, which can also be run on a standard 120V outlet, currently serves to power an onboard computer, network router, GPS/INS system, and 5 GHz radio. The onboard computer is a Pandora Mini PC manufactured by Cappuccino PC. Each unit has an i-7 Intel processor, 80 GB SSD, 8 GB of RAM, and runs a Windows 7 environment. In addition to onboard sensors, the computers are connected to an onboard router which allows each computer to communicate with one another and a base station via the 5 GHz radio. The 5 GHz radio is part of the Rocket M series manufactured by Ubiquiti Networks. It acts as a wireless Ethernet bridge and has a throughput of 150+ Mbps. Moreover, each radio has a range of several miles allowing for efficient and wide scale implementation. Each radio automatically connects with the base station, allowing for communication to a base station computer or between robots using the base station as a relay. The GPS/INS system is the SPAN-CPT manufactured by NovAtel. This unit is rated to an accuracy of $\pm$ 1.5 m and, in practice, is generally accurate to less than $\pm$ 1 m. Moreover, the units can be linked to the base station for RTK corrections, allowing for accuracies of ± 2 cm. Heading accuracy falls within $\pm$ 0.5^o. The GPS/INS communicates over USB or RS-232 and can relay position and IMU data up to 20 Hz and 100 Hz, respectively. In addition to these built in systems, each TURTLE has mechanical mounting points for an Ibeo Lux 4 LIDAR, which are installed when needed. These devices, which communicate over Ethernet, collect scan data across 110^o at up to 200m away. They have additional obstacle detection features built in. Their maximum scanning rate is 50 Hz. Any two of the four robots can optionally be fitted with Velodyne HDL-32E 3-D LIDARs as well. These sensors feature 32 beams spanning 40^o vertically. The 32 beams rotate 360^o on an axis at 10 Hz, producing 700,000 points per second that describe the surroundings. The beams produce distance measurements with accuracies of $\pm$ 2cm at a range of up to 100 m. In addition to the LIDARs, each robot is designed to be optionally fitted from a collection of the following: one Point Grey Bumblebee XB3 Stereo Camera, one Point Grey BumbleBee2 Stereo Camera, one Ladybug5 Spherical Camera, or four Basler Scout GigE HD Cameras. These vision systems allow for the variable collection of environmental data specific to the project.

The TURTLEs vehicle team was used by Adam Shoemaker in collaboration with Professor Kevin Kochersburger (VT) during field testing of collaborative ground and air vehicle characterizations terrain and buildings. The results of collaborative testing are depicted in Figure 5. While Mr. Adam Shoemaker unexpectedly departed without completing his PhD, it is anticipated that the TURTLEs team will play an key role in validating the results of the analysis summarized in the technical descriptions above.

9 Appendix

9.1 Reproducing Kernel Hilbert Spaces

This section presents a short overview of the background theory for reproducing kernel Hilbert Spaces. A thorough treatment can be found in [51, 52]. This background will enable the careful study of the above evolution problems. In this report we make use of several properties of *real* reproducing kernel Hilbert (RKH) spaces defined over a set $\mathcal{X} \subseteq \mathbb{R}^d$. First we discuss an RKH space $H_\mathcal{X}$ of real-valued functions and subsequently discuss a special class of vector-valued RKH spaces. An RKH space $H_\mathcal{X}$ of functions that map $\mathcal{X} \to \mathbb{R}$ is defined in terms of a real-valued, continuous, and symmetric function $\mathfrak{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ of positive type that is referred to as the kernel underlying the RKH space. When we say that $\mathfrak{K}$ is of positive type, this means that the collocation matrix $\mathbb{K}_m := [\mathfrak{K}(x_i, x_j)] \in \mathbb{R}^{m\times m}$ is positive semidefinite for every selection of points $\{x_i\}_{1\leq i\leq m} \subset \mathcal{X}$ with $m \in \mathbb{N}$. The kernel is of strictly positive type if all such collocation matrices are strictly positive definite for collections of distinct points $\{x_i\}_{1\leq i\leq m} \subset \mathcal{X}$. The function $\mathfrak{K}_x := \mathfrak{K}(x, \cdot)$ is known as the kernel function at x, and the inner product of two such functions $\mathfrak{K}_x, \mathfrak{K}_y$ is defined to be $(\mathfrak{K}_x, \mathfrak{K}_y)_{H_\mathcal{X}} := k(x, y)$ for each $x, y \in \mathcal{X}$. The RKH space $H_\mathcal{X}$ is the closed finite linear span of the set of functions $\{\mathfrak{K}_x \mid x \in \mathcal{X}\}$, that is,

$$H_\mathcal{X} := \overline{\{\mathfrak{K}_x \mid x \in \mathcal{X}\}} := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; f = \lim_{N\to\infty} \sum_{i=1}^{N} \alpha_{N,i} \mathfrak{K}_{x_{N,i}} \right. \\ \left. \text{with } \alpha_{N,i} \in \mathbb{R} \text{ and } x_{N,i} \in \mathcal{X} \right\}.$$

The closure in this definition is taken with respect to the norm $\|\cdot\|_{H_\mathcal{X}} := \sqrt{(\cdot,\cdot)_{H_\mathcal{X}}}$. It is well-known [51, 52, 50] that with this construction the reproducing property

$$E_x(f) := (\mathfrak{K}_x, f)_{H_\mathcal{X}} = f(x)$$

holds for all $f \in H_\mathcal{X}$ and $x \in \mathcal{X}$. Here $E_x : H_\mathcal{X} \to \mathbb{R}$ is the evaluation operator.

It is known [80, 81] that if for some positive constant $\bar{\mathfrak{K}}$, we have $\sup_{x\in\mathcal{X}} \mathfrak{K}(x, x) \leq \bar{\mathfrak{K}} < \infty$, then the evaluation operator E_x is a uniformly (in x) bounded linear operator, $\|E_x\| \leq C$, for some constant C. This results follows from the fact that

$$|E_x f| = |f(x)| = |(\mathfrak{K}_x, f)_H| \leq \|K_x\|_H \|f\|_H = \sqrt{k(x, x)}\|f\|_H \leq \sqrt{\bar{\mathfrak{K}}}\|f\|_H.$$

The above inequalities also imply that $\|f\|_{C(\mathcal{X})} \lesssim \|f\|_{H_\mathcal{X}}$, and therefore $H_\mathcal{X} \subseteq C(\mathcal{X})$. We consider only kernels $\mathfrak{K}$ on $\mathcal{X}$ for which such a constant $\bar{\mathfrak{K}}$ exists. For our calculations below, the adjoint operator $E_x^* : \mathbb{R} \to H_\mathcal{X}$ is needed, and a quick calculation shows that

$$E_x^* : \alpha \in \mathbb{R} \mapsto \alpha \mathfrak{K}_x \in H_\mathcal{X}$$

for each $x \in \mathcal{X}$. Since $\|E_x^*\| = \|E_x\|$, we know that the adjoint is a uniformly (in x) bounded linear operator, $\|E_x^*\| \leq \sqrt{\mathfrak{K}}$.

In some examples we are interested in function spaces that are built from the restrictions of functions in $H_\mathcal{X}$. We denote the set of functions

$$R_\Omega(H_\mathcal{X}) := \left\{ f : \Omega \to \mathbb{R} \mid f = g\Big|_\Omega \text{ with } g \in H_\mathcal{X} \right\}.$$

It is shown in [52, 51] that $R_\Omega(H_\mathcal{X})$ is in fact a RKH in its own right, and the kernel on this space is $\mathfrak{K}_{R_\Omega}(x,y) := \mathfrak{K}|_{\Omega\times\Omega}(x,y)$ for all $x, y \in \Omega$ where $\mathfrak{K}$ is the kernel that defines $H_\mathcal{X}$. Often we will also need to construct the closed subspaces

$$H_\Omega := \overline{\{\mathfrak{K}_x \mid x \in \Omega\}} \subseteq H_\mathcal{X} \tag{17}$$

defined in terms of the *indexing subset* $\Omega \subseteq \mathcal{X}$. We emphasize that the indexing set is not necessarily the same as the set over which the functions in the RKH space are defined. When the kernel $\mathfrak{K}$ is assumed to be supported on $\mathcal{X} \times \mathcal{X}$, the functions in $H_\mathcal{X}$ and H_Ω are supported on $\mathcal{X}$, while the indexing sets are $\mathcal{X}$ and Ω, respectively. Thus, in this report, the functions in such a space H_Ω are not to be interpreted as restrictions of functions in $H_\mathcal{X}$, that is, the spaces $H_\Omega \subset H_\mathcal{X}$ and $R_\Omega(H_\mathcal{X})$ are usually not synonymous. It is, however, again a standard construction that the subspace $H_\Omega \subset H_\mathcal{X}$ is a RKH space in its own right. Let Π_Ω be the $H_\mathcal{X}$-orthonormal projection of $H_\mathcal{X}$ onto the closed subspace H_Ω. The inner product on H_Ω is given by $(\cdot,\cdot)_{H_\Omega} := (P_\Omega(\cdot), (\cdot))_{H_\mathcal{X}}$. The kernel $\mathfrak{K}_\Omega$ of H_Ω is given by $\mathfrak{K}_\Omega(x,y) := (P_\Omega \mathfrak{K}_x)(y)$ for all $x, y \in \mathcal{X}$. In view of the above, we let H be a generic RKH space, and it may be either $H_\mathcal{X}, H_\Omega$, or $R_\Omega(H_\mathcal{X})$.

The general theory of vector-valued RKH spaces can be found in [75, **?**], and we denote $\mathbb{H} := H^N$. That is, in this short report we restrict attention to product spaces such as

$$\mathbb{H}_\mathcal{X} := \bigtimes_{1\leq i\leq N} H_\mathcal{X} := H_\mathcal{X} \times \cdots \times H_\mathcal{X}.$$

Let us consider the example of $\mathbb{H} := \mathbb{H}_\mathcal{X}$ in a bit more detail, since the other examples follow similarly. Since each coordinate function f_i in $\mathbb{F} := \{f_1, \ldots, f_N\} \in \mathbb{H}_\mathcal{X}$ lies in the RKH space $H_\mathcal{X}$ that has the same kernel $\mathfrak{K}$, the matrix of evaluation operators $\mathbb{E}_X := \text{diag}\{E_{x_i} \mid 1 \leq i \leq N\}$ defines the map $\mathbb{E}_X : \mathbb{H}_\mathcal{X} \to \mathbb{R}^N$ with $\mathbb{E}_X(\mathbb{F}) := \mathbb{F}(X) = \{E_{x_1} f_1, \ldots, E_{x_N} f_N\}^T$ for each $X := \{x_1, \ldots, x_N\} \in \mathbb{R}^d$ and $\mathbb{F} \in \mathbb{H}_\mathcal{X}$. To find the adjoint, for each $X \in \Omega$ we have

$$\begin{aligned}(\mathbb{E}_X \mathbb{F}, \alpha)_{\mathbb{R}^N} = \alpha^T \mathbb{F}(X) &= \sum_{1\leq i\leq N} (E_{x_i} f_i, \alpha_i)_\mathbb{R} = \sum_{1\leq i\leq N} (\mathfrak{K}_{x_i} \alpha_i, f_i)_H = (\{\alpha_i \mathfrak{K}_{x_i}\}_{1\leq i\leq N}, \mathbb{F})_{\mathbb{H}_\mathcal{X}}, \\ &= (\mathbb{E}_X^*, \mathbb{F})_{\mathbb{H}_\mathcal{X}}\end{aligned}$$

for all $\mathbb{F} \in \mathbb{H}$. Thus, the adjoint $\mathbb{E}_X^* : \mathbb{R}^N \to \mathbb{H}$ is given by $\mathbb{E}_X^* \alpha := \mathfrak{K}_X \alpha \in \mathbb{H}$ for $\alpha \in \mathbb{R}^N$ where $\mathfrak{K}_X := \text{diag}\{\mathfrak{K}_{x_i} \mid 1 \leq i \leq N\}$. Carefully note that $\boldsymbol{\mathfrak{K}}_X$ is a square diagonal matrix with the scalar-valued kernels $\mathfrak{K}_{x_i}$ on the main diagonal.

Before concluding this section, we make one last assumption regarding the family of kernels $\mathfrak{K}$ in this report. We say that the RKH space $H_{\mathcal{X}}$ includes a *rich family of bump functions* if for any $x \in \mathcal{X}$ and any open ball $B_r(x) := \{y \in \mathcal{X} \mid d_{\mathcal{X}}(y, x) < r\}$ there is a bump function $b_{x,r} \in H_{\mathcal{X}}$ such that $b_{x,r} : \mathcal{X} \to [0, 1]$, $b_{x,r}(x) = 1$, the support of $b_{x,r}$ is contained in a compact set $C \subset B_r(x)$. An assumption of this type is not esoteric as it might at first seem. It is often required in applications that use RKH spaces that the kernel exhibits good "separation properties." For instance, the assumption of the existence of bump functions is similar to, but stronger than, the assumption in [76] where it is needed to establish measurability of subsets of $\Omega \subset \mathcal{X}$ in terms of the kernel. In [76] a kernel is said to separate the closed sets if for each closed set C and point $x \notin C$ there is a function $f \in H_{\mathcal{X}}$ such that $f(x) \neq 0$ and $f(y) = 0$ for all $y \in C$. It is clear that if $H_{\mathcal{X}}$ has a rich family of bump functions then it separates the closed sets. The richness of bump functions condition has appeared in reports that seek to construct best approximations by functions in an RKH space in [74]. In this reference it is noted that any RKH space that contains $C_0^\infty(\mathcal{X})$, the continuous functions on $\mathcal{X}$ having compact support contained in $\mathcal{X}$, admits a rich family of bump functions. Here we will see that this assumption enables the study of the relation between a positive limit set and a persistently excited set.

9.2 Persistence and Positive Limit Sets

The next two results illustrate simple, and intuitive relationships between persistently excited sets, positive orbits $\Gamma^+(X_0)$, and the positive limit set $\omega^+(X_0)$. First we have the following relationship between the forward orbit $\Gamma^+(X_0)$ when the set $\mathcal{X} \equiv \Omega$ in the PE Definition 2.

Theorem 5 *Let $H_{\mathcal{X}}$ be the RKH space of functions generated by a kernel $\mathfrak{K}$ over the domain $\mathcal{X}$ and suppose that this RKH space includes a rich family of bump functions. If the PE condition in Definition 2 holds for the domain $\Omega = \mathcal{X}$, then the forward orbit $\Gamma^+(X_0)$ is dense in Ω. That is, we have*

$$X \in \Omega = \mathcal{X} \quad \Longrightarrow \quad \exists \{t_k\}_{k \in \mathbb{N}} \ \ \textit{with} \ \ \lim_{k \to \infty} S(t_k) X_0 \to X.$$

Note that Theorem 5 does not require that the set of times $t_k \to \infty$. Recall, on the other hand, that the positive limit set $\omega^+(X_0)$ is contained in the closure of all accumulation points, of the orbit $\Gamma^+(X_0)$ for sequences of the form $\{S(\tau_k)X_0\}$, as $\tau_k \to \infty$. Next, we discuss a relationship of the positive limit set $\omega^+(X_0)$ and a PE set $\mathcal{X} = \Omega$ in Definition 2.

Theorem 6 *Let $H_{\mathcal{X}}$ be the RKH space of functions over $\mathcal{X}$ generated by a kernel $\mathfrak{K}$ and suppose that this RKH space includes a rich family of bump functions. If the PE condition in Definition 2 holds for $\Omega = \mathcal{X}$, then*

$$\mathcal{X} = \omega^+(X_0).$$

The next example shows that much more complicated systems can be studied with Definition 2

Example 1 *We now consider a case where the semigroup $\{S(\tau)\}_{\tau \geq 0}$ defining the true solution of the governing equations is a flow on the unit circle S^1. It is known that S^1 is, of course, a compact Riemannian manifold that is an embedded submanifold of $\mathbb{R}^2$. We can view this evolution as a special case of a more general situation.*

Let us set $\mathcal{X} := \mathbb{M}$ with $\mathbb{M}$ any $k-$dimensional, smooth, compact, Riemannian manifold. The Sobolev space $H^{r,2}(\mathbb{M})$ of real-valued functions $f : \mathbb{M} \rightarrow \mathbb{R}$ over the manifold $\mathbb{M}$ consists of all functions in $f \in L^2(\mathbb{M})$ such that all the (covariant) distributional derivatives $\nabla^s f$ for $s \leq r$ are elements in $L^2(\mathbb{M})$. We assume that $r > d/2$, which by the Sobolev embedding theorem guarantees that $H^{r,2}(\mathbb{M}) \hookrightarrow C(\mathbb{M})$, that is, this Sobolev space is contained in the continuous functions and the imbedding is continuous. Since we have

$$|E_X f| = |f(X)| \leq \|f\|_{C(\Omega)} \lesssim \|f\|_{H^{r,2}(\mathbb{M})}$$

*for each $X \in \mathbb{M}$ and $f \in H^{r,2}(\mathbb{M})$, each evaluation functional $E_X : H^{r,2}(\mathbb{M}) \rightarrow \mathbb{R}$ is bounded. But $H^{r,2}(\mathbb{M})$ is a Hilbert space; boundedness of all its evaluation functionals implies that $H^{r,2}(\mathbb{M})$ is a RKH space. Following the analysis in [**?**, **?**], it is possible to define the RKH space $H^{\mathfrak{K},r}(\mathbb{M})$ in terms of the Matern-Sobolev kernel $\mathfrak{K}_{\mathbb{M},r}$ of smoothness $r > 0$ that is the fundamental solution of the elliptic differential operator equation $\sum_{1 \leq \ell \leq r} (\nabla^\ell)^* \nabla^\ell \mathfrak{K}_{\mathbb{M},r} = \delta$ where ∇ is the covariant derivative operator over the manifold $\mathbb{M}$ and δ denotes the Dirac distribution. It has been shown that $H^{\mathfrak{K},r}(\mathbb{M}) \approx H^{r,2}(\mathbb{M})$ for the chosen range $r > d/2$. See reference [74] for the case when $\mathbb{M}$ is a smooth Riemannian manifold or see reference [**?**] for the special case $\mathbb{M} := \mathbb{R}^d$. Moreover, it is well-known that the Sobolev space $H^{s,2}(\mathbb{M})$ contains the space $C_0^\infty(\mathbb{M})$. This means that the RKH space $H^{\mathfrak{K},r}(\mathbb{M})$ contains a rich family of bubble functions. Now define the vector-valued function space $\mathbb{H}^{\mathfrak{K},r}(\mathbb{M}) := (H^{\mathfrak{K},r}(\mathbb{M}))^k$. We conclude that if the motion over the manifold $\mathbb{M}$ satisfies the persistency condition in Definition 2, then $\mathbb{M} := \omega^+(X_0)$.*

This example illustrates that the newly introduced persistency condition can be applicable, *in principal*, to the study of certain evolutions over smooth compact Riemannian manifolds. Still, the analysis in the example above is fairly abstract. Perhaps more importantly, it is not a simple task to come up with a closed form expression for the Sobolev-Matern kernel. Of course this can be done for motions on S^1 or S^2, but is not readily accomplished for some arbitrary manifold $\mathbb{M}$. The definition of the space $H^{\mathfrak{K},r}(\mathbb{M})$ is intrinsic in the above example : it depends on the domain $\mathbb{M}$ and the family of charts used to define the manifold.

The next example illustrates that even if the form of the manifold $\mathbb{M}$ is unknown, it is possible to come up with constructions of a kernel for $\mathbb{M}$ that is extrinsic in the sense that it is defined by the restriction of some known kernel on a larger domain. This line of attack may be useful to the study of unknown or uncertain dynamical systems in that the persistency of excitation condition is cast in terms of the kernel on the larger space in this case. We examine this case in the next example.

Example 2 *In this example we are interested in an evolution over the domain* $\Omega := \mathbb{M} \subseteq \mathcal{X} := \mathbb{R}^d$*, where here* $\mathbb{M}$ *is a smooth, k-dimensional, compact Riemannian, embedded submanifold of* $\mathbb{R}^d$*. Suppose that we can observe the states* $X(t) := S(t)X_0 \subset \mathbb{R}^d$ *for each* $t \geq 0$ *of the semiflow, we know that the orbit* $\Gamma^+(X_0)$ *is a subset of some manifold* $\mathbb{M}$*, but we do not know the exact form of the manifold* $\mathbb{M}$*. Actually, in many problems of interest in deriving a "reduced complexity" approximation of some high-dimensional model, assumptions of this sort are common. Since the manifold* $\mathbb{M}$ *is unknown, it is not possible to derive a closed form expression for a kernel intrinsic to* $\mathbb{M}$*. However, the definitions of kernels as the restrictions of kernels over a larger domain has a long history in the theory of RkH spaces. [52, 51] There are a host of closed form expressions for the kernels of RKH spaces defined over* $\mathcal{X} := \mathbb{R}^d$*. For instance, the Matern-Sobolev kernels over* $\mathbb{R}^d$ *are given for* $r > d/2$ *by*

$$\mathfrak{K}(x,y) = \mathfrak{K}(\|x-y\|_{\mathbb{R}^d}) = \mathfrak{K}(\xi) = \frac{2^{1-(r-d/2)}}{\Gamma(r-d/2)}\xi^{r-d/2}\mathfrak{B}_{r-d/2}(\xi)$$

for all $x, y \in \mathbb{R}^d$ *with* $\mathfrak{B}_{r-d/2}$ *the Bessel function of order* $r - d/2$*. As in the last example, we have* $H^{\mathfrak{K},r}(\mathbb{R}^d) \approx H^{r,2}(\mathbb{R}^d)$ *under the condition that* $r > d/2$*. We can define a candidate* $\mathfrak{K}_{\mathbb{M}}$ *for a kernel over* $\mathbb{M}$ *from the restriction*

$$\mathfrak{K}_{\mathbb{M}}(x,y) := \mathfrak{K}\Big|_{\mathbb{M}\times\mathbb{M}}(x,y)$$

for all $x, y \in \mathbb{M}$*. Denote the RKH space generated by the kernel* $\mathfrak{K}_{\mathbb{M}}$ *as* $H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M})$*. At this point we do not yet have a rigorous notion of exactly how smooth the restricted functions in* $H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M})$ *are. But from Lemma 4 of [**?**], we know that the* $H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M}) = T(H^{\mathfrak{K},r}(\mathbb{R}^d))$ *where* T *is the trace operator*

$$\begin{aligned} T &: H^{\mathfrak{K},r}(\mathbb{R}^d) \to H^{\mathfrak{K}_{\mathbb{M}},r}(\mathbb{M}), \\ T &: f \to f|_{\mathbb{M}}. \end{aligned}$$

In other words the trace operator T *maps a function on* $\mathbb{R}^d$ *to its restriction to* $\mathbb{M}$*, and it is synonymous with the restriction operator* R_Ω *introduced earlier. It is more common in references on RKH spaces to refer to the restriction operator [51, 52], while in references that discuss Sobolev spaces it is much more commonly referred to as the trace operator. From Proposition 2 of [**?**], under the standing assumptions on* $\mathbb{M}$*, the trace operator* $T : f \to f|_{\mathbb{M}}$ *is a continuous operator from* $H^{r,2}(\mathbb{R}^d)$ *onto* $H^{r-(d-k)/2,2}(\mathbb{M})$ *for* $r > (d-k)/2$ *and* $1 \leq k \leq d$*. In summary then, if we choose the kernel* $\mathfrak{K}$ *on* $\mathbb{R}^d$ *with a sufficiently large smoothness index* r *so that* $r > r - (d-k)/2 > d/2$*, we have*

$$H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M}) = T(H^{\mathfrak{K},r}(\mathbb{R}^d) \approx T(H^{r,2}(\mathbb{R}^d)) \approx H^{r-(d-k)/2,2}(\mathbb{M}).$$

This set of equivalencies gives a precise notion of the smoothness or regularity of the restricted functions in the RKH space $H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M})$: the RKH space over $\mathbb{M}$ is equivalent to the Sobolev space having smoothness $r-(d-k)/2$ when the original kernel $\mathfrak{K}$ is r-regular. As in the last example, we set the vector-valued function space $\mathbb{H}^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M}) = (H^{\mathfrak{K}_{\mathbb{M}}}(\mathbb{M}))^d$. Again, we see that since $C_0^\infty(\mathbb{M}) \subset H^{\tau,2}(\mathbb{M})$ for any $\tau > 0$, the RKH space contains a rich set of bubble functions. If the original semiflow is persistently exciting as in Definition 3, with

$$\Omega := \mathbb{M}, \qquad \mathcal{X} := \mathbb{R}^d, \qquad R_\Omega := T,$$

then $\omega^+(X_0) = \mathbb{M}$. Note that the statement of persistence in Definition 3 is expressed in terms of the kernel $\mathfrak{K}_{\mathbb{M}} := \mathfrak{K}|_{\mathbb{M}\times\mathbb{M}}$, which can be used for computations since a closed form for $\mathfrak{K}$ is known on $\mathcal{X}$.

References

[1] M. Andreasson, D.V. Dimarogonas, H. Sandberg, K.H. Johansson, *Distributed control of networked dynamical systems: Static feedback, integral action and consensus,* **IEEE Trans. Automat. Control,** 59, 7 , 2014, pp. 1750-1764.

[2] J. Baumeister, W. Scondo, M.A. Demetriou, and I.G. Rosen, *On-Line Parameter Estimation for Infinite Dimensional Dynamical Systems*, *SIAM J. Control. Optim.*, Vol. 35, No. 2, pp. 678–713, 1997.

[3] Alain Berlinet and Christine Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer, 2004.

[4] Parag Bobade, Suprotim Majumdar, Savio Pereira, Andrew J. Kurdila, and John B. Ferris, *Adaptive Estimation in Reproducing Kernel Hilbert Spaces*, IEEE American Control Conference 2017.

[5] Parag Bobade, Suprotim Majumdar, Savio Pereira, Andrew J. Kurdila, and John B. Ferris, *Adaptive estimation for nonlinear systems using reproducing kernel Hilbert spaces*, Advances in Computational Mathematics, Springer, 2018.

[6] X Bu, Z Hou, H Zhang, *Consensus of hybrid multi-agent systems*, **IEEE transactions on neural networks**, 2018.

[7] Yiran Cao, Toshiki Oguchi, Paul Verhoeckx, and Henk Nijmeijer, *Consensus Control for a Multiagent System with Time Delays*, Mathematical Problems in Engineering, 2017.

L. Cheng, Z.-G. Hou, M. Tan, Y. Lin, W. Zhang, *Solving a modified consensus problem of linear multi-agent systems*, **Automatica,** 47, 10, 2011, pp. 2218-2223

[8] M.A. Demetriou, *Adaptive Parameter Estimation of Abstract Parabolic and Hyperbolic Distributed Parameter Systems,* Ph.D. thesis, Departments of Electrical-Systems and Mathematics, University of Southern California, Los Angeles, CA, 1993.

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

[10] M.A. Demetriou and I.G. Rosen, *On the Persistence of Excitation in the Adaptive Identification of Distributed Parameter Systems*, IEEE Trans. Automat. Control, Vol. 39, pp. 1117–1123, 1994.

[11] Zhaoda Deng, Jessica Gregory, and Andrew Kurdila, *Learning Theory with Consensus in Reproducing Kernel Hilbert Spaces*, IEEE American Control Conference 2012.

[12] L. Ding, Q.L. Han, X. Ge, X.M. Zhang, *An overview of recent advances in event-triggered consensus of multiagent systems,* **IEEE Transactions on Cybernetics**, 48, 4, April, 2018, pp. 1110 - 1123.

[13] L. Ding, W.X. Zheng, G Guo, *Network-based practical set consensus of multi-agent systems subject to input saturation,* **Automatica**, 2018.

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

Y. Feng, S. Xu, B. Zhang, *Group consensus control for double-integrator dynamic multiagent systems with fixed communication topology, Internat. J. Robust Nonlinear Control,* 24, 3, 2014, pp. 532-547.

[15] Y. Hong, G. Chen, L. Bushnell, *Distributed observers design for leader-following control of multi-agent networks,* **Automatica,** 44, 3, 2008, pp. 846-850.

[16] Yiguang Hong, Jiangping Hua, Linxin Gao, *Tracking control for multi-agent consensus with an active leader and variable topology,* **Automatica**, Volume 42, Issue 7, July 2006, Pages 1177-1182.

[17] A. Jadbabaie, J. Lin, A.S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules,* **IEEE Trans. Automat. Control,** 48, 6, 2003, pp. 988-1001.

[18] Xiaozheng Jin, Shaofan Wang, Jiahu Qin, Wei Xing Zheng , Yu Kang, *Adaptive Fault-Tolerant Consensus for a Class of Uncertain Nonlinear Second-Order Multi-Agent Systems With Circuit Implementation*, **IEEE Transactions on Automatic Control**, 2018.

[19] Z. Li, Z. Duan, G. Chen, L. Huang, *Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint,* **IEEE Trans. Circuits Syst.**, 57, 1, 2010, pp. 213-224.

[20] Z. Li, Z. Duan, F.L. Lewis, *Distributed robust consensus control of multi-agent systems with heterogeneous matching uncertainties,* **Automatica,** 50, 3, 2014, pp. 883-889.

[21] S. Li, Y. Jun, W. Chen, X. Chen, *Generalized extended state observer based control for systems with mismatched uncertainties,* **IEEE Trans. Ind. Electron.**, 59, 12, 2012, pp. 4792-4802.

[22] Z. Li, X. Liu, W. Ren, L. Xie, *Distributed tracking control for linear multiagent systems with a leader of bounded unknown input,* **IEEE Trans. Automat. Control,** 58, 2, 2013, pp. 518-523

[23] Z. Li, W. Ren, X. Liu, L. Xie, *Distributed consensus of linear multi-agent systems with adaptive dynamic protocols,* **Automatica**, 49, 7, 2013, pp. 1986-1995.

[24] S. Li, J. Yang, W.-H. Chen, X. Chen, *Generalized extended state observer based control for systems with mismatched uncertainties, IEEE Trans. Ind. Electron.*, 59, 12, 2012, pp. 4792-4802.

[25] P. Lin, Y. Jia, L. Li, *Distributed robust consensus control in directed networks of agents with time-delay,* **Systems Control Lett.,** 57, 2008, pp. 643-653.

[26] Y. Liu, Y. Jia, *Consensus problem of high-order multi-agent systems with external disturbances: An analysis approach, Internat. J. Robust Nonlinear Control,* 20 (14) (2010), pp. 1579-1593.

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

[28]

[29] Angelia Nedic, Asuman Ozdaglar, Pablo A. Parrilo, *Constrained Consensus and Optimization in Multi-Agent Networks,* **IEEE Transactions on Automatic Control**, 55 , 4 , April, 2010.

[30] Michael Demetriou and Stephen Nestinger, *Adaptive consensus estimation of multiagent systems*, IEEE Conference on Decision and Control 2011.

[31] W. Ni and D. Cheng, *Leader-following consensus of multi-agent systems under fixed and switching topologies,* **Systems Control Lett.**, 59, 3, 2010, pp. 209-217.

[32] W. Ni, L. Xiao, X. Chun "Consensus controllability, observability and robust design for leader-following linear multi-agent system," *Automatica*, 49 (7) (2013), pp. 2099-2105.

[33] B Ning, J Jin, J Zheng, Z Man, *Finite-time and fixed-time leader-following consensus for multi-agent systems with discontinuous inherent dynamics,* **International Journal of Control**, 2018.

[34] R. Olfati-Saber, R. Murray *Consensus problems in networks of agents with switching topology and time delays,* **IEEE Trans. Automat. Control,** 49, 9, 2004, pp. 1520-1533.

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

[36]

[37] Gabriele Perozzi, Denis Efimov, Jean-Marc Biannic, Laurent Planckaert, Patricia Coton, *Wind estimation algorithm for quadrotors using detailed aerodynamic coefficients*, IEEE American Control Conference 2018.

[38] Jiahu Qin, Qichao Ma, Yang Shi, and Long Wang, *Recent Advances in Consensus of multiagent Systems: A Brief Survey*, IEEE Transactions on Industrial Electronics, 2017.

[39] W. Ren, E. Atkins, *Distributed multi-vehicle coordinated control via local information exchange,* **Internat. J. Robust Nonlinear Control,** 17, 10–11, 2007, pp. 1002-1033.

[40] W. Ren, R.W. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control,* Springer, New York (2008)

[41] Wie Ren, R. W. Beard, and E. M. Atkins, *A survey of consensus problems in multiagent coordination*, IEEE American Control Conference 2005.

[42] W. Ren, Y. Cao, *Distributed Coordination of Multi-Agent Networks,* Springer, New York (2011).

[43] J Gonzalez-Rocha, CA Woolsey, C Sultan, S de Wekker, N Rose, *Measuring atmospheric winds from quadrotor motion*, AIAA Atmospheric Flight Mechanics Conference 2017.

[44] Federico Rossi, Saptarshi Bandyopadhyay, Michael Wolf, and Marco Pavone, *Review of multiagent Algorithms for Collective Behavior: a Structural Taxonomy*, arXiv:1803.05464v1 [cs.RO], March, 2018.

[45] Teymur Sadikhov, Michael A. Demetriou, Wassim M. Haddad, and Tansel Yucelen, *Adaptive Estimation Using Multiagent Network Identifiers With Undirected and Directed Graph Topologies*, Journal of Dynamic Systems, Measurement, and Control, March 2014.

[46] Saburou Saitoh and Yoshihiro Sawano, *Theory of Reproducing Kernels and Applications*, Springer, 2016.

[47] Alexandre Seuret, Dimos V. Dimarogonas and Karl H. Johansson, *Consensus under Communication Delays*, IEEE Conference on Decision and Control 2008.

[48] Srdjan S. Stanković, Miloš S.Stanković, Dušan M.Stipanović, *Consensus based overlapping decentralized estimation with missing observations and communication faults*, **Automatica**, 45, 6, June, 2009, Pages 1397-1406.

[49] Srdjan S. Stankovic ; Miloš S. Stankovic ; Dušan M. Stipanovic, *Decentralized Parameter Estimation by Consensus Based Stochastic Approximation,* **IEEE Transactions on Automatic Control**, 56, 3, March, 2011, pp. 531 - 543.

[50] Z. Tang, T. Huang, J. Shao, J. Hu, *Leader-following consensus for multiagent systems via sampled-data control,* **IET Control Theory Appl.,** 5, 14, 2011, pp. 1658-1665.

[51] B Tian, H Lu, Z Zuo, W Yang, *Fixed-time leader-follower output feedback consensus for second-order multiagent systems,* **EEE transactions on cybernetics,** 2018.

[52] J.A. Walker, *Abstract Dynamical Systems and Evolution Equations*, Springer, 1980.

[53] X. Wang, S Li, M. Z. Q. Chen, *Composite Backstepping Consensus Algorithms of Leader–Follower Higher-Order Nonlinear Multiagent Systems Subject to Mismatched Disturbances*, **IEEE transactions on cybernetics**, 48, 6, June, 2018, pp. 1935 - 1946.

[54] G. Wen, Z. Duan, W. Yu, G. Chen, *Consensus of multi-agent systems with nonlinear dynamics and sampled-data information: a delayed-input approach,* **Internat. J. Robust Nonlinear Control,** 23, 6, 2013, pp. 602-619

[55] Y Wu, Z Wang, S Ding, H Zhang, *Leader–follower consensus of multi-agent systems in directed networks with actuator faults*, **Neurocomputing,** 2018.

[56] Peng Yang, Randy A. Freeman, Kevin M. Lynch, *Multi-Agent Coordination by Decentralized Estimation and Control,* **IEEE Trans. on Automatic Control**, 53, 1, 2008.

[57] H. Yang, Z. Zhang, S. Zhang, *Consensus of second-order multi-agent systems with exogenous disturbances,* **Internat. J. Robust Nonlinear Control,** 21, 9 , 2010, pp. 945-956.

[58] W. Yu, G. Chen, W. Ren, J. Kurths, W.X. Zheng *Distributed higher order consensus protocols in multiagent dynamical systems,* **IEEE Trans. Circuits Syst.**, 58, 8, 2011, pp. 1924-1932.

[59] H. Zhang, F. Lewis, A. Das, *Optimal design for synchronization of cooperative systems: State feedback, observer and output feedback,* **IEEE Trans. Automat. Control**, 56, 8, 2013, pp. 1948-1952.

[60] Z. Zuo, B. Tian, M. Defoort, Z. Ding, *Fixed-time consensus tracking for multiagent systems with high-order integrator dynamics,* **IEEE Transactions on Automatic Control**, 63, 2, February, 2018, Page(s): 563 - 570.

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

[62] Petros A. Ioannou, Jing Sun, *Robust Adaptive Control*, Dover, 2012.

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

[64] B. Pasik-Duncan, *On the Consistency of a Least Squares Identification Procedure in Linear Evolution Systems,* Stochastics Report., Vol. 39, pp. 83–94, 1992.

[65] Holger Wendland, *Scattered data approximation*, Cambridge Monographs on Applied and Computational Mathematics, 17, Cambridge University Press, 2005.

[66] K.D Narendra and P.Kudva, *Stable adaptive schemes for identification and control*, IEEE Transactions on Systems, Man, and Cybernetics, 1974.

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

[68] 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 J.Control.Optim., 1977.

[69] A.J. Kurdila , F.J. Narowich F.J, and J. Ward, *Persistency of Excitation in Identification Using Radial Basis Function Approximants*, SIAM J.Control.Optim.,Vol 33, 1995.

[70] 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, Vol. 23, No. 2, February, 2012.

[71] H.A. Kingravi, G. Chowdhary,P.A. Vela, E.N. Johnson, *Reproducing kernel hilbert space approach for the online update of radial bases in neuro-adaptive control*, IEEE Transactions of Neural Network and Learning Systems, 2012.

[72] Joel B. Predd, Sanjeev R. Kulkarni, and H. Vincent Poor, *Distributed Learning in Wireless Sensor Networks*, arXiv:cs/0503072 [cs.IT], March 2005.

[73] J.B. Predd, S.R. Kulkarni, and H.V. Poor, *A Collaborative Training Algorithm for Distributed Learning*, IEEE Transactions on Information Theory, Vol. 55, No. 4, pp. 1856–1871, April 2009.

[74] Thomas Hangelbroek, Francis J. Narcowich, Christian Rieger, and Joseph D. Ward, *Direct and Inverse Results on Bounded Domains for Meshless Methods via Localized Bases on Manifolds*, arXiv:1406.1435 [math.NA], 2017.

[75] C. Carmeli, E. De Vito, A. Toigo, and V. Umanità, *Vector valued reproducing kernel Hilbert spaces and universality*, arXiv:0807.1659 [math.FA], 2008.

[76] Ernesto De Vito, Lorenzo Rosasco, and Alessandro Toigo, *Learning Sets with Separating Kernels*, arXiv:1204.3573 [stat.ML], 2014.

[77] Andrew Kurdila and Yu Lei, *Adaptive control via embedding in reproducing kernel Hilbert spaces*, IEEE American Control Conference 2013.

[78] Jessical Gregory, *A Rate of Convergence for Learning Theory with Consensus*, MS Thesis, Department of Mechanical Engineering, Virginia Tech, 2015.

[79] Parag Bobade, *Modeling, Approximation, and Control for a Class of Nonlinear Systems*, PhD Dissertation, Department of Engineering Science and Mechanics, Virginia Tech, 2017.

[80] Steve Smale and Ding-Xuan Zhou, *Online Learning with Markov Sampling*, Analysis and Applications, 2009.

[81] Steve Smale and Ding-Xuan Zhou, *Learning Theory Estimates via Integral Operators and Their Approximations*, Constructive Approximation, August 2007.

[82] H. Zhang, Z. Li, Z. Qu, and F.L. Lewis, "On Constructing Lyapunov Functions for multiagent Systems," *Automatica*, Vol. 58, pp. 39-42, 2015.

[83] H. Zhang, F. L. Lewis, Z. Qu, "Lyapunov, Adaptive, and Optimal Design Techniques for Cooperative Systems on Directed Communication Graphs," *IEEE Transactions on Industrial Electronics*, Vol. 59, No. 7, pp. 3016-3040, July, 2012.

[84] N. Barabanov and R. Ortega, " Global Consensus of Time-Varying Multiagent Systems Without Persistent Excitation Assumptions," *IEEE Transactions on Automatic Control*, Vol. 63, pp. 3835-3939, No. 11, November, 2018.

[85] R. Olfati-Saber and R. M. Murray, " Consensus Problems in Networks of Agents with Switching Topology and Time-Delays," *IEEE Transactions on Automatic Control,* Vol. 49, No. 9, pp. 1520-1533, September, 2004.

[86] B.D.O. Anderson, G. Shi, and J. Trumpf, "Convergence and State Reconstruction of Time-Varying multiagent Systems from Complete Observability Theory," *IEEE Transactions on Automatic Control*, pp. 2519-2523, Vol. 62, No. 5, May, 2017.

[87] N.R. Chowdhury and S. Sukumar, "Persistence Based Analysis of Consensus Protocols for Dynamic Graph Networks," `ArXiv:1404.1168v1`, April, 4, 2014.

[88] Parag Bobade, Dimitra Panagou, and Andrew J. Kurdila, *Multiagent Adaptive Estimation with Consensus in Reproducing Kernel Hilbert Spaces,* Manuscript 1009 submitted to the *2019 European Control Conference*, `https://www.overleaf.com/read/swxbnjvhnkdh`, November 16, 2018.

[89] Parag Bobade, Dimitra Panagou, and Andrew J. Kurdila, *Consensus and Persistency in Reproducing Kernel Hilbert Spaces,* preprint, `https://www.overleaf.com/read/ksmcfbvrbgpk`, 2019.

[90] Parag Bobade and Andrew J. Kurdila, *Koopman Theory and Linear Approximation Spaces*, preprint, `arXiv:1811.10809v1`, 2019.

[91] Milan Korda and Igor Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," *Journal of Nonlinear Science*, 2017, pp. 1–24.

[92] Nithin Govindrajan, and Ryan Mohr, and Shivkumar Chandrasekaran, and Igor Mezić, "On the approximation of Koopman spectra for measure preserving transformations," arXiv preprint `arXiv:1803.03920`, 2018.

[93] Milan Korda, and Mihai Putinar, and Igor Mezić, "Data-driven Spectral analysis of Koopman Operator," arXiv preprint `arXiv:1710.06532`, 2017.

[94] Dimitrios Giannakis, "Data-driven spectral decomposition and forecasting of ergodic dynamical systems," *Applied and Computational Harmonic Analysis*, September 2017.

[95] Stefan Klus, Patrick Gelb, Sebastian Peitz, and Christof Schutte, "Tensor Based Dynamic Mode Decomposition," `arXiv:1606.06625v2`, January, 2017.

[96] Stefan Klus, Ingmar Schuster, and Krikamol Muandet, Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces, arXiv preprint `arXiv:1712.01572v1`, December, 2017.

[97] Ronald DeVore, Gerard Kerkyacharian, Dominique Picard, and Vladimir Temlyakov, "Approxmation methods for supervised learning," *Foundations of Computational Mechanics*, Vol. 6, No. 1, pp 3-58, 2006.

[98] Ronald A. DeVore and George G. Lorentz, *Constructive Approximation*, Springer, 1993

[99] Ronald A. DeVore, "Nonlinear Approximation," *Acta Numerica*, Cambridge University Press, pp. 51—150, 1998.

[100] Peter Binev, and Albert Cohen, and Wolfgang Dahmen, and Ronald DeVore, "Universal Algorithms for Learning Theory Part I: Piecewise Constant Functions," 2005.

[101] Peter Binev, and Albert Cohen, and Wolfgang Dahmen, and Ronald DeVore "Universal Algorithms for Learning Theory Part II: Piecewise Polynomial Functions," 2006.

[102] Albrecht Pietsch, "Approximation spaces," *Journal of Approximation Theory*, 1981, Vol. 32, No. 2, pp. 115-134, 1981.

[103] Gerard Kerkyacharian and Dominque Picard "Regression in Random Design and Warped Wavelets," *Bernoulli*, Volume 10, Number 6, pp. 1053–1105, 2004.

[104] Gerard Kérkyacharian and Dominique Picard, "Entropy, universal coding, approximation, and bases properties," *Constructive Approximation,* Volume 20, Number 1, pp. 1–37, 2003.

[105] "A Distribution-Free Theory of Nonparametric Regression," L. Gyorfy, M. Kohler, A. Kryzak, and H. Walk, Springer, Berlin, 2002.

[106] "Near-Optimal Approximation Rates for Distribution Free Learning with Exponentially, Mixing Observations," Andrew J. Kurdila and Bin Xu, *Proceedings of the 2010 American Control Conference*, doi=10.1109/ACC.2010.5530863

[107] Sebastian Thrun, Wolfram Burgard, and Dieter Fox, *Probabalistic Robotics*, Massachussetts Institute of Technology, 2006.

[108] Mehran Mesbahi and Magnus Egerstedt, *Graph Theoretic Methods in Multiagent Networks*, Princeton University Press, 2010.