LIFTING TECHNIQUES FOR DATA-DRIVEN IDENTIFICATION OF NONLINEAR SYSTEMS

By

Moad Abudia

Bachelor of Science in Mechanical Engineering
Oklahoma State University
Stillwater, OK
2016

Master of Science in Mechanical Engineering
Oklahoma State University
Stillwater, OK
2019

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2026

LIFTING TECHNIQUES FOR DATA-DRIVEN IDENTIFICATION OF NONLINEAR SYSTEMS

Dissertation Approved:

Dr. Rushikesh Kamalapurkar

Dissertation Advisor

Dr. Kursat Kara

Chair

Dr. He Bai

Dr. Alan Noell

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Kamalapurkar.

I would like to thank my family for their patience.

I would like to thank the Air Force Office of Scientific Research, National Science Foundation, and of course Oklahoma State University for the funding and facilities to conduct this research.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: MOAD ABUDIA

Date of Degree: JULY, 2026

Title of Study: LIFTING TECHNIQUES FOR DATA-DRIVEN IDENTIFICATION OF NONLINEAR SYSTEMS

Major Field: MECHANICAL ENGINEERING

Abstract:

With the current abundance of data, methods for effective use of data for system analysis and control are needed. System identification (SysID) is a numerical method that uses a set of trajectories of a dynamical system and produces a model that represents the unknown system. In this dissertation, we study data-driven, equation-free identification of dynamical systems. the objectives of the study are twofold. The first objective is data-driven prediction, i.e., utilization of a set of recorded trajectories to generate new trajectories or to extend the recorded trajectories into the future. The second objective is modeling for data-driven control, i.e., obtaining a model capable of anticipating the response of a system to new control inputs. For nonlinear systems and systems that are described by PDEs, the structure of the model is often too complex to be developed using first principles. To overcome the difficulty posed by the unknown structure of the system, equation-free SysID approaches such as neural networks, autoregressive moving averages, proper orthogonal decomposition have been developed over the past decade. In this dissertation we focus on lifting techniques.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation

The increasing availability of high-resolution trajectory, sensor, and input-output data has made data-driven modeling an essential tool for the analysis and control of dynamical systems. System identification (SysID) seeks to construct a mathematical model from observed data so that the resulting model can reproduce, predict, or control the behavior of the underlying system. For linear time-invariant (LTI) systems, this problem is comparatively well understood: mature frequency-domain, prediction-error, subspace, state-space, and behavioral methods are available, and the main modeling questions often reduce to selecting an appropriate model order and estimating a finite number of parameters [1–5].

The nonlinear setting is fundamentally more difficult. Nonlinear systems do not possess a universal finite-dimensional structure analogous to the LTI state-space form, and systems governed by partial differential equations, such as fluid flows, deformable solids, transport processes, and thermal systems, may have very high-dimensional or infinite-dimensional state descriptions. When the equations of motion are known up to unknown parameters, nonlinear system identification can sometimes be reduced to parameter estimation, and adaptive or prediction-error methods can be applied. However, in many applications the functional structure of the model is not known *a priori*. In that case, the objective is not merely to estimate unknown coefficients, but to infer a useful representation of the dynamics from data.

The objectives of this dissertation are twofold. The first objective is data-driven pre-

diction: given a set of recorded trajectories, construct a model capable of extending those trajectories into the future or predicting new trajectories from new initial conditions. The second objective is modeling for data-driven control: given state and input data, construct a model capable of anticipating the response of a system to new control signals. A model with this property can be used for controller design, motion planning, stability analysis, parameter identification, and the discovery of approximately optimal trajectories with respect to a prescribed cost. The difficulty is that the identified model must be expressive enough to represent nonlinear behavior while remaining structured enough to support analysis and computation.

A common strategy is to replace the search for a nonlinear model in the original state space with the search for a linear or bilinear representation in a lifted space. A lifting map embeds the state, the trajectory, or the observable functions of a system into a higher-dimensional space where the evolution of the lifted variables has a simpler structure. In the most favorable case, a nonlinear finite-dimensional system can be represented as a infinite-dimensional linear system. Finite-dimensional approximations of that infinite-dimensional representation then lead to algorithms that resemble linear system identification, spectral decomposition, regression, or model reduction. In this sense, lifting is not merely a change of coordinates; it is a modeling principle that trades nonlinear complexity in the original variables for linear structure in a larger function space.

The lifting viewpoint has two classical origins. Carleman linearization lifts the state of a nonlinear system to the collection of monomials in the state and represents the nonlinear dynamics by an infinite-dimensional linear system [6]. Koopman operator theory instead lifts the problem from states to observables: the Koopman operator acts linearly on functions evaluated along trajectories of a nonlinear system [7]. These two constructions are complementary. Carleman lifting gives an explicit algebraic hierarchy for analytic or polynomial systems, while Koopman-based methods provide an operator-theoretic framework in which nonlinear dynamics are represented through linear action on observables.

The dissertation focuses on lifting techniques for data-driven identification of nonlinear systems, with particular emphasis on occupation kernels, Liouville operators, control Liouville operators, and Carleman lifting. The framework used throughout the dissertation may be organized according to the type of system and the lifted object. Figure 1 gives a schematic summary. Koopman-type approaches lift observables for discrete-time systems, Liouville-type approaches lift the infinitesimal evolution of observables for continuous-time systems, occupation-kernel methods lift trajectories to elements of reproducing kernel Hilbert spaces (RKHSs), and Carleman methods lift the state through monomials. Controlled systems require additional structure: control occupation kernels and vector-valued RKHSs incorporate the input signal, while control Liouville operators represent the infinitesimal action of control-affine dynamics.

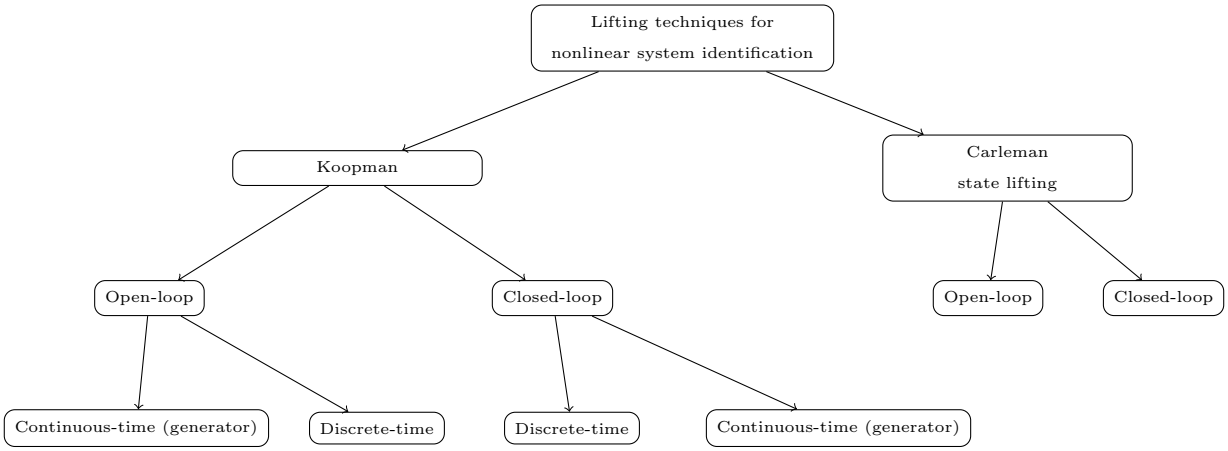

Figure 1: A taxonomy of the lifting techniques considered in this dissertation. The diagram organizes the methods by lifting framework, distinguishing Koopman/operator-theoretic lifting from Carleman state lifting, and then by whether the resulting formulation is open-loop or closed-loop. The figure also indicates whether the lifted model is formulated in discrete time or through a continuous-time generator.

The methods developed in the dissertation are motivated by several limitations of existing approaches. Classical numerical differentiation is sensitive to measurement noise. Dictionary-based lifting methods can be effective, but their success depends strongly on the

choice of dictionary. Koopman-based DMD methods are computationally attractive, but spectral convergence is subtle because convergence in the strong operator topology does not generally imply spectral convergence. Carleman linearization provides powerful error bounds when the model is known, but its classical use is not fully data-driven. Parameter identification methods are mature when the way parameters enter the dynamics is known, but many practical systems contain parameters whose coupling to the dynamics is difficult to write down explicitly. The goal of the dissertation is to address these limitations by developing lifting-based identification methods that retain the advantages of linear representations while providing stronger theoretical or numerical guarantees in settings relevant to nonlinear prediction, control, and parameter identification.

1.2 Literature Review

1.2.1 Classical system identification and nonlinear model learning

Classical system identification provides a deep foundation for data-driven modeling. For LTI systems, frequency-domain and time-domain approaches can identify transfer functions, impulse responses, state-space realizations, and input-output models from experimental data [1]. These methods are powerful because the model class is fixed, the role of inputs and outputs is clear, and the resulting models are directly compatible with linear control design. Behavioral methods further show that trajectories themselves contain enough information to represent the behavior of an LTI system under suitable richness conditions [2, 3]. The principal advantage of these approaches is that they are mature, computationally efficient, and accompanied by extensive statistical and control-theoretic analysis. Their main disadvantage, for the purposes of this dissertation, is that they are not designed to identify general nonlinear systems without imposing additional structure.

For nonlinear systems, the identification problem depends strongly on what is known in advance. When a parametric structure is available, the problem can be reduced to estimat-

ing unknown constants. Adaptive control and Lyapunov-based estimation methods provide strong tools in this setting, particularly when the uncertain dynamics admit a known regressor form [8–10]. The advantage of these methods is that they can be integrated with stability and control objectives. The disadvantage is that they usually require detailed knowledge of how the unknown parameters enter the dynamics. When that structural information is unavailable, one must use a more flexible model class.

Several equation-free nonlinear modeling techniques have been developed for this purpose, including neural networks, autoregressive moving-average models, proper orthogonal decomposition, and support vector machines [11–14]. These approaches can fit complicated input-output relationships and are useful in many applications. However, flexible model classes often introduce new difficulties: interpretability may be reduced, stability and error guarantees may be difficult to obtain, and the resulting model may not have an operator structure suitable for spectral analysis or controller synthesis. Sparse identification of nonlinear dynamics (SINDy) addresses interpretability by selecting a sparse model from a prescribed library of candidate functions [15, 16]. Its main strength is that it can produce explicit governing equations. Its main weakness is the dependence on the chosen library and the need to estimate time derivatives from data, a step that can be highly sensitive to noise. This motivates methods that avoid numerical differentiation and instead use integral information from trajectories.

1.2.2 Koopman operators, DMD, and EDMD

Dynamic mode decomposition (DMD) is one of the most influential lifting-based methods for data-driven analysis of nonlinear systems. DMD seeks a finite-rank representation of a transfer operator associated with a dynamical system using time-series measurements [17–20]. In its classical form, DMD constructs a best-fit linear map between snapshot matrices and decomposes the observed data into modes with associated growth rates or frequencies. The method is attractive because it is computationally simple, interpretable, and closely

related to spectral analysis.

The operator-theoretic interpretation of DMD is based on the Koopman operator. For a discrete-time system $x_{k+1} = F(x_k)$, the Koopman operator K_F acts on an observable g by

$$(K_F g)(x) = g(F(x)). \tag{1.2.1}$$

Although the state dynamics may be nonlinear, K_F is linear on a space of observables. This linearity is the central advantage of Koopman theory: nonlinear dynamics can be studied using spectral objects such as eigenvalues, eigenfunctions, and modes. Koopman-based DMD has been successful because it connects data-driven computation with a linear operator whose spectrum encodes temporal behavior [17, 21].

The classical Koopman-DMD framework also has limitations. First, a discrete-time Koopman operator requires a flow map or a fixed sampling interval. This is restrictive for continuous-time systems with finite escape times, irregularly sampled data, or multiple sensors with unequal sampling rates. Second, convergence of finite-rank Koopman approximations is delicate. Results such as [20] establish convergence in the strong operator topology under appropriate assumptions, but convergence in the strong operator topology does not imply spectral convergence [22]. Since DMD relies precisely on spectral information, convergence of the finite-rank operators alone is not sufficient to guarantee convergence of the DMD spectrum. Third, compactness of the Koopman operator is often needed for strong spectral approximation results, yet Koopman operators associated with nonlinear dynamics are typically noncompact in many natural settings [23, 24]. These issues motivate continuous-time and compact-operator alternatives.

Extended dynamic mode decomposition (EDMD) improves on classical DMD by choosing a dictionary of observables and constructing a finite-dimensional approximation of the Koopman operator on the span of that dictionary [21]. The advantage of EDMD is flexibility: nonlinear observables can be included directly, and the resulting method can approximate nonlinear systems more accurately than a linear fit in the original state coordinates. The main disadvantage is dictionary dependence. If the dictionary does not contain functions that

represent the relevant Koopman-invariant subspace, the approximation may require many observables or may produce spurious spectral information. Data-driven dictionary learning partially addresses this weakness. For example, machine-learning-based EDMD can select observables that reduce prediction error using fewer functions than a fixed library [25]. The trade-off is increased computational cost and the need for training and validation. Eigenfunction validation methods, such as those used in Koopman Reduced Order Nonlinear Identification and Control (KRONIC), can reduce spurious modes and improve reliability when the Gram matrix is rank deficient [26]; nevertheless, validation does not eliminate the more basic question of whether the selected finite-dimensional space yields a convergent spectral approximation.

Several controlled extensions of Koopman and DMD methods have also been proposed. Dynamic mode decomposition with control represents controlled nonlinear systems using input-output data and is computationally attractive for prediction and control [27]. EDMD with control generalizes this idea by incorporating dictionaries of observables and control-dependent terms [28]. These methods can provide good predictions and useful lifted models, but the spectral convergence and control-separation issues remain subtle. For discrete-time nonlinear control systems, one must represent both state evolution and input evolution; in some formulations this introduces a shift operator for the control signal [28]. Other approaches separate state and control using approximations or derive bilinear lifted forms. The Koopman canonical transform and the Koopman bilinear form provide continuous-time lifted models that are useful for feedback design [29, 30]. More recent bilinear Koopman formulations can produce stabilizing controllers under sufficiently rich data and suitable region-of-attraction assumptions [31]. Their advantage is direct relevance to control; their limitation is that the richness requirements may be difficult to satisfy for unstable systems or systems that cannot be safely excited in all input directions.

1.2.3 Liouville operators and continuous-time DMD

Continuous-time DMD replaces the discrete-time Koopman operator with an infinitesimal generator. For a continuous-time system $\dot{x} = f(x)$ and an observable g, the Liouville operator is formally given by

$$A_f g = \nabla g \cdot f. \tag{1.2.2}$$

In Koopman terminology, this operator is the generator of the Koopman semigroup when the semigroup is well defined. The Liouville formulation is natural for continuous-time data because it models infinitesimal evolution rather than a fixed-time flow map. It is therefore better suited for systems sampled irregularly, systems with finite escape times, and trajectories with different time horizons.

A major advantage of Liouville-operator DMD is that compactness can be recovered by choosing the domain and range appropriately. Some attempts to obtain compact operators for DMD multiply Koopman or Liouville operators by auxiliary compactifying operators [32, 33]. These methods are useful because compact operators have better spectral approximation properties, but the resulting compact operators only approximately represent the original dynamics. In contrast, compact Liouville operators can be constructed so that the operator truly corresponds to the continuous-time dynamics on appropriate RKHSs [34, 35]. This provides a direct path toward norm-convergent finite-rank approximations, which are stronger than strong-operator-topology convergence and more relevant for spectral methods.

The Liouville framework also connects naturally with occupation kernels. Rather than approximating derivatives from data, one can pair the dynamics with trajectory-dependent functions that represent integration along trajectories. This connection is central to the dissertation. It provides a way to construct finite-rank approximations of Liouville operators from trajectory data while reducing sensitivity to measurement noise. For controlled systems, the analogous object is the control Liouville operator, which acts on vector-valued observables

and incorporates the control input. The resulting framework links continuous-time DMD, RKHS methods, control occupation kernels, and finite-rank operator approximation.

1.2.4 Kernel methods, RKHSs, and occupation kernels

Kernel methods offer a complementary approach to lifting. Instead of explicitly selecting a finite dictionary of features, a positive definite kernel implicitly defines a possibly infinite-dimensional feature space, namely a reproducing kernel Hilbert space. Kernel methods have been widely adopted in system identification because they provide flexible nonparametric models, regularization, and a natural way to encode smoothness or prior information [36–38]. The advantage of kernel methods is that they combine expressive modeling with convex optimization in many settings. Their disadvantage is the need to choose kernels and tune hyperparameters, which can have a large effect on performance and remains a persistent difficulty [39].

Occupation kernels adapt RKHS ideas to trajectory data. Given a trajectory $\gamma : [0, T] \to \mathbb{R}^n$ and an RKHS H of functions, the occupation kernel $\Gamma_\gamma \in H$ represents the integral functional

$$\langle h, \Gamma_\gamma \rangle_H = \int_0^T h(\gamma(t))\, dt. \tag{1.2.3}$$

If the unknown vector field $f = (f_1, \ldots, f_n)^\top$ has components in H and γ satisfies $\dot{\gamma} = f(\gamma)$, then

$$\langle f_i, \Gamma_\gamma \rangle_H = \int_0^T f_i(\gamma(t))\, dt = \int_0^T \dot{\gamma}_i(t)\, dt = \gamma_i(T) - \gamma_i(0). \tag{1.2.4}$$

Thus, trajectory endpoint differences provide linear measurements of the unknown dynamics in the RKHS without requiring numerical differentiation [40]. This is the key advantage of occupation-kernel methods: they transform derivative-based identification into integration-based identification, which is substantially less sensitive to high-frequency measurement noise.

Occupation kernels generalize occupation measures in a function-theoretic way, analogous to the way in which kernel functions can be viewed as regularized point evaluations

or function-theoretic analogs of delta distributions. Occupation measures have been used extensively in optimal control and reachability [41–43]. Occupation kernels inherit the integral representation of occupation measures while living inside an RKHS, making them directly usable as basis functions for approximation. They have been used for motion tomography [44, 45], fractional-order nonlinear system identification [46], and the construction of finite-rank representations of Liouville operators [47]. Their limitations are those typical of RKHS methods: performance depends on the chosen kernel, the method may scale poorly with the number of trajectories, and theoretical guarantees generally require the unknown dynamics to have sufficient regularity relative to the chosen RKHS.

Control occupation kernels extend occupation kernels to controlled systems. For a control-affine system

$$\dot{x} = f(x) + g(x)u, \tag{1.2.5}$$

the unknown dynamics include both the drift f and the control effectiveness matrix g. A control occupation kernel incorporates both the state trajectory and the control signal in a vector-valued RKHS so that the inner product with the unknown row $\begin{pmatrix}(f)_j & (g)_j\end{pmatrix}$ produces an integral measurement of the controlled dynamics [48]. The advantage is that open-loop controlled trajectory data can be used directly. The main challenge is that the RKHS must now encode vector-valued structure, and the quality of the identified model depends on both state-space coverage and richness of the control signals.

1.2.5 Carleman lifting and finite-dimensional truncation

Carleman linearization is another classical lifting technique. It maps a finite-dimensional nonlinear system into an infinite-dimensional linear system by augmenting the state with monomials of the original state variables [6]. For polynomial systems, this embedding has an explicit algebraic structure. For analytic systems, a related hierarchy can be obtained through Taylor or Maclaurin expansions. The primary advantage of Carleman lifting is that it gives a direct linear representation of nonlinear dynamics in lifted coordinates. This makes

linear-system tools available for approximation, reachability, and control.

The main limitation of Carleman lifting is truncation. The exact lifted system is infinite dimensional, while computations require a finite-dimensional truncation. The number of monomials grows rapidly with the state dimension and truncation order, which can lead to high computational cost. Moreover, truncation errors depend on the system, the region of interest, and the time horizon. For known polynomial nonlinearities, explicit truncation-error estimates have been developed [49]. These estimates can be used for reachability analysis of weakly nonlinear systems, where the nonlinear terms are dominated by the linear terms [50]. Carleman lifting has also been used to approximate optimal feedback controllers through finite-dimensional optimization problems [51]. These works show the theoretical value of Carleman methods, but they generally assume that the nonlinear vector field is known.

More recent work provides computable error bounds for Carleman linearization of known nonlinear systems on regions of attraction [52]. The advantage of these results is that the truncation order can be related to a desired approximation error and time horizon. The limitation, from the standpoint of system identification, is that the model-based Carleman construction starts with a known nonlinear vector field. This dissertation uses these theoretical results as a foundation for data-driven Carleman lifting: the truncation order is selected using prescribed-error analysis, and the parameters of the truncated lifted linear system are estimated from trajectory data.

1.2.6 Parameter identification and operator-theoretic models

Parameter estimation is a mature area with classical roots in process identification, prediction-error methods, gray-box modeling, recursive estimation, and adaptive control [1, 8, 53]. In many nonlinear-control applications, parameter update laws are designed using known regressors, as in adaptive robot manipulator control [9]. These methods are powerful when the relationship between the unknown parameters and the dynamics is known. Their limitation is precisely this structural requirement.

Modern approaches also often encode parameter dependence in advance. Unknown parameters may enter linearly through known regressors [54], nonlinearly through a prescribed neural network architecture [55], or through a selected set of basis functions [56]. Bayesian methods and neural inverse-problem methods provide additional tools for recovering parameters from data [57–59]. These methods can quantify uncertainty or handle high-dimensional inverse problems, but parameter recovery is often separated from the construction of an operator-theoretic model of the closed-loop dynamics.

Many practical systems involve parameters whose coupling to the dynamics is difficult to express explicitly. A vehicle carrying an unknown payload, a robot manipulating an unknown object, or a building thermal system subject to unknown occupancy patterns may experience parameter-dependent changes that are nonlinear and hard to model from first principles [60–63]. This motivates an operator-theoretic approach in which unknown parameters are treated as latent components of an augmented state and identified directly from input-output data. The DCLDMD framework provides a foundation for this idea by constructing finite-rank approximations of closed-loop control Liouville operators using kernel propagation and multiplication operators [33, 64]. The advantage is that parameter identification can be integrated with lifted dynamical modeling. The limitation is that the resulting parameter estimates depend on excitation, data coverage, and the validity of the augmented-state representation.

1.3 Contributions

This dissertation develops lifting-based methods for data-driven identification, prediction, control-oriented modeling, and parameter identification of nonlinear systems. The contributions are organized around two complementary lifting mechanisms: occupation-kernel/Liouville lifting and Carleman lifting. The occupation-kernel and Liouville chapters emphasize continuous-time, RKHS-based, and operator-theoretic identification. The Carleman chapter emphasizes prescribed-error finite-dimensional lifted linear models. Together, these contributions ad-

dress the limitations identified above: sensitivity to numerical differentiation, lack of spectral convergence guarantees in Koopman-DMD approximations, difficulty incorporating controls into continuous-time lifted models, dependence of Carleman methods on known vector fields, and the need for known parameter-dynamics coupling in classical parameter identification.

The first contribution is the development of control occupation kernel regression (COKR) for nonlinear higher-order control-affine systems. It introduces higher-order control occupation kernels and uses them as basis functions in a regularized regression problem for systems of the form

$$\frac{d^s x}{dt^s} = f(x) + g(x)u. \tag{1.3.1}$$

The approach extends occupation-kernel regression for uncontrolled or fractional-order systems [46] and control occupation kernels for first-order controlled systems [48]. Its main advantage is that it identifies the drift and control effectiveness terms using integral information from trajectories, avoiding numerical differentiation of noisy data. The method also leads to a finite-dimensional matrix problem through a representer-theorem argument and provides cumulative error bounds through a functional ridge-regression formulation. Numerical examples on the Duffing oscillator and a two-link robot manipulator demonstrate its ability to estimate nonlinear control-affine dynamics and to support control-oriented simulation.

The second contribution is a data-driven Carleman lifting method for nonlinear system identification with guaranteed error bounds. Existing Carleman results provide useful truncation-error estimates when the nonlinear model is known [49, 52]. This dissertation adapts those ideas to the identification setting. Given trajectory data, a desired time horizon, and information about the decay of the Maclaurin expansion, the method selects a truncation order and estimates the finite-dimensional lifted linear dynamics from data. The resulting identified lifted system is constructed so that its trajectory remains within a prescribed error bound of the trajectory of the nonlinear system over the desired horizon. This contribution is significant because it combines the interpretability and error analysis of model-based Car-

leman lifting with data-driven estimation of the lifted linear system.

The third contribution is a dynamic mode decomposition method for continuous-time nonlinear control-affine systems using compact control Liouville operators. Koopman-based DMD methods provide useful finite-dimensional approximations, but spectral convergence is not guaranteed by strong-operator-topology convergence alone [20, 22]. This dissertation addresses that issue by using control Liouville operators on appropriate RKHSs and deriving finite-rank approximations that converge in norm to the true compact operator under suitable richness assumptions. Control occupation kernels and vector-valued RKHSs allow the state and control data to be incorporated directly. The resulting SCLDMD algorithm yields an operator-theoretic model for control-affine systems and is validated on a controlled Duffing oscillator.

The fourth contribution is the derivation of a norm-convergent continuous-time DMD formulation and its equivalence with occupation kernel regression. Previous compactification approaches produce useful DMD algorithms but may approximate modified operators rather than the true operator associated with the dynamics [32, 33]. This dissertation uses compact Liouville operators that correspond directly to the continuous-time dynamics and develops a finite-rank representation of the Liouville operator itself. The resulting DMD model is shown to be equivalent to the unregularized occupation-kernel regression model. This equivalence clarifies the relationship between two apparently different identification methods: spectral finite-rank approximation of Liouville operators and RKHS regression using occupation kernels.

The fifth contribution is an operator-theoretic approach to parameter identification using discrete control Liouville dynamic mode decomposition. Classical and modern parameter-identification techniques often require prior knowledge of how the parameters enter the dynamics [1, 9, 53, 55]. This dissertation relaxes that requirement by augmenting the system state with unknown parameters and embedding the augmented dynamics into an RKHS and vector-valued RKHS operator framework. The resulting DCLDMDPID algorithm identifies

a lifted model that can predict trajectories associated with candidate parameter values, and parameter estimates are selected by comparing predicted and observed trajectories. Numerical experiments on a controlled Duffing oscillator demonstrate simultaneous recovery of state evolution and unknown parameters from open-loop data.

The final contribution is a unified presentation of lifting techniques for data-driven nonlinear identification. The dissertation connects Koopman and DMD methods, Liouville operators, occupation kernels, control occupation kernels, vector-valued RKHSs, Carleman lifting, and parameter identification within a single narrative. The literature review emphasizes not only what each method can do, but also where each method is limited: DMD is efficient but may lack spectral convergence; EDMD is flexible but dictionary-dependent; kernel methods are expressive but require kernel and hyperparameter choices; occupation kernels avoid numerical differentiation but rely on RKHS regularity and data coverage; Carleman lifting provides error bounds but can suffer from combinatorial growth and usually assumes model knowledge; and classical parameter identification is mature but often requires known parameter structure. The methods developed in this dissertation are designed to preserve the strengths of these approaches while addressing several of their limitations through integral data representations, compact Liouville operators, control-aware RKHS constructions, and prescribed-error lifted linear modeling.

1.4 Dissertation Organization

Chapter II provides the background material needed for the dissertation, including classical DMD, RKHSs, control occupation kernels, vector-valued RKHSs, Liouville operators, occupation kernels, and Carleman lifting. Chapter III develops control occupation kernel regression for higher-order nonlinear control-affine systems and presents numerical examples involving the Duffing oscillator and a two-link robot manipulator. Chapter IV develops the data-driven Carleman lifting method with prescribed error bounds and demonstrates the method on a nonlinear oscillator example. Chapter V develops singular control Liouville

dynamic mode decomposition for continuous-time control-affine systems. Chapter VI studies norm-convergent Liouville DMD and establishes its equivalence with occupation kernel regression. Chapter VII extends discrete control Liouville DMD to parameter identification by augmenting the state with unknown physical parameters. The dissertation concludes by summarizing the main results and identifying directions for future work.

CHAPTER II

BACKGROUND

This chapter collects the background material used throughout the dissertation. The presentation follows the operator-theoretic structure used in the later chapters. We begin with dynamic mode decomposition (DMD), extended dynamic mode decomposition (EDMD), and their connection with the Koopman operator. We then introduce reproducing kernel Hilbert spaces (RKHSs), explain why the usual Koopman framework has limitations for spectral approximation, and introduce continuous-time Liouville operators as an alternative. Occupation kernels are then presented as trajectory-dependent RKHS elements that connect data to Liouville operators through integration rather than numerical differentiation. Finally, the chapter extends these ideas to vector-valued RKHSs, controlled systems, control occupation kernels, and Carleman lifting.

2.1 Dynamic Mode Decomposition and Koopman Operator Theory

Consider a continuous-time dynamical system

$$\dot{x} = f(x), \tag{2.1.1}$$

where $x(t) \in \mathbb{R}^n$ is the state and $f : \mathbb{R}^n \to \mathbb{R}^n$ is the vector field. If the system admits a flow map $F_{\Delta t}$ over a fixed sampling interval $\Delta t > 0$, then the sampled states satisfy

$$x_{k+1} = F_{\Delta t}(x_k) = x_k + \int_{k\Delta t}^{(k+1)\Delta t} f(x(t))\, dt, \tag{2.1.2}$$

where $x_k = x(k\Delta t)$. Given a sequence of snapshots $x_1, x_2, \ldots, x_m$, define

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{m-1} \end{bmatrix}, \qquad X' = \begin{bmatrix} x_2 & x_3 & \cdots & x_m \end{bmatrix}.$$

The objective of classical DMD is to find a finite-dimensional linear map A such that

$$X' \approx AX. \tag{2.1.3}$$

A least-squares solution is

$$A = X'X^{\dagger}, \tag{2.1.4}$$

where $X^{\dagger}$ is the Moore-Penrose pseudoinverse. The eigenvectors of A are referred to as DMD modes, and the associated eigenvalues describe the growth, decay, or oscillation rates of the fitted linear model.

DMD is closely connected to the Koopman operator. For a discrete-time dynamical system $x_{k+1} = F(x_k)$, the Koopman operator K_F acts on scalar observables g by

$$(K_F g)(x) = g(F(x)). \tag{2.1.5}$$

Although F may be nonlinear on the state space, K_F is linear on a space of observables. This is the main operator-theoretic insight behind Koopman-based data-driven modeling: nonlinear state evolution is represented as linear evolution of functions along trajectories.

Classical DMD may be interpreted as a finite-dimensional approximation of the Koopman operator on the span of a selected set of observables. In the simplest case, the observables are the coordinate projection functions, so that DMD fits a linear model directly in the state coordinates. Extended dynamic mode decomposition (EDMD) generalizes this idea by choosing a dictionary of observables

$$\psi(x) = \begin{bmatrix} \psi_1(x) & \psi_2(x) & \cdots & \psi_N(x) \end{bmatrix}^{\top},$$

and computing a matrix approximation of the Koopman operator on $\operatorname{span}\{\psi_1, \ldots, \psi_N\}$ [21]. If this span is invariant under the Koopman operator, then EDMD can recover an exact finite-dimensional representation on that subspace. In practice, the span is only approximately invariant, so the quality of the approximation depends strongly on the selected dictionary.

The Koopman/EDMD viewpoint is powerful because it turns nonlinear modeling into a linear operator approximation problem. However, several limitations are important for

the rest of the dissertation. First, the discrete-time Koopman operator requires a flow map over a fixed sampling interval. This may be restrictive for systems with finite escape times, irregularly sampled data, or multiple sensors sampled at unequal frequencies. Second, the boundedness and compactness of the Koopman operator depend on the selected function space. Koopman operators are often noncompact, and finite-rank approximations that converge strongly do not necessarily yield convergent spectra [20, 22]. Since DMD uses eigenvalues and eigenfunctions of finite-rank approximations, spectral convergence is a central issue. These limitations motivate continuous-time generator methods, Liouville operators, and compact RKHS-based formulations.

2.2 Reproducing Kernel Hilbert Spaces

A Hilbert space generalizes Euclidean space to spaces whose elements may be functions or sequences. The inner product structure allows one to define norms, orthogonality, projections, and adjoints. A reproducing kernel Hilbert space is a Hilbert space of functions in which pointwise evaluation is continuous.

Definition 2.2.1 *A reproducing kernel Hilbert space (RKHS) H over a set X is a Hilbert space of functions from X to $\mathbb{R}$ such that, for each $x \in X$, the evaluation functional $E_x : H \to \mathbb{R}$ defined by $E_x g := g(x)$ is bounded.*

By the Riesz representation theorem, for each $x \in X$ there exists a function $K_x \in H$ such that

$$\langle g, K_x \rangle_H = g(x), \qquad g \in H. \tag{2.2.1}$$

The function K_x is the kernel function centered at x, and

$$K(x, y) = \langle K_y, K_x \rangle_H \tag{2.2.2}$$

is the reproducing kernel of H [65]. Equivalently, $K(x, y) = K_y(x)$. A symmetric function $K : X \times X \to \mathbb{R}$ is positive semidefinite if the Gram matrix $(K(x_i, x_j))_{i,j=1}^M$ is positive semidefinite

for every finite collection $\{x_1, \ldots, x_M\} \subset X$. The Aronszajn-Moore theorem states that every positive semidefinite kernel induces a unique RKHS with K as its reproducing kernel [66].

RKHSs are useful in data-driven identification because they allow functions to be manipulated through kernel evaluations. In particular, snapshots may be embedded into H through the kernel map

$$x \mapsto K(\cdot, x).$$

The span of the kernel functions $\{K(\cdot, x) : x \in X\}$ is dense in H. Indeed, if $h \in H$ is orthogonal to every $K(\cdot, x)$, then $h(x) = \langle h, K(\cdot, x)\rangle_H = 0$ for all $x \in X$, hence $h \equiv 0$ on X. Therefore, finite linear combinations of kernel functions provide a natural approximation class for functions in H.

2.3 Continuous-Time Operators and Liouville Operators

The continuous-time counterpart of the Koopman operator is its infinitesimal generator. For the system (2.1.1), the associated Liouville operator is formally defined by

$$A_f g = \nabla g \cdot f, \tag{2.3.1}$$

where g is an observable. More precisely, if H_d and H_r are Hilbert spaces of functions, then a Liouville operator with symbol f may be defined as

$$A_f : \mathcal{D}(A_f) \subset H_d \to H_r, \qquad \mathcal{D}(A_f) = \{g \in H_d : A_f g \in H_r\}. \tag{2.3.2}$$

The action of A_f describes the time derivative of an observable along trajectories because

$$\frac{d}{dt} g(x(t)) = \nabla g(x(t)) \cdot \dot{x}(t) = \nabla g(x(t)) \cdot f(x(t)) = (A_f g)(x(t)). \tag{2.3.3}$$

Thus, Liouville operators provide a continuous-time operator-theoretic representation of the dynamics without requiring a fixed-time flow map.

A key issue for DMD-type spectral methods is whether finite-rank approximations converge to the underlying operator in a topology strong enough to control spectra. Compact

operators are particularly useful in this respect. In RKHSs generated by exponential dot product kernels, Liouville operators can be compact when the domain and range spaces are chosen appropriately. Let

$$K_\mu(x,y) = \exp\left(\frac{x^\top y}{\mu}\right), \qquad \mu > 0. \tag{2.3.4}$$

In one variable, the native space of this kernel is the restriction of the Bargmann-Fock space to the real line, denoted by $F^2_\mu(\mathbb{R})$. Its elements are functions of the form

$$h(x) = \sum_{k=0}^{\infty} a_k x^k,$$

where

$$\sum_{k=0}^{\infty} |a_k|^2 \mu^k k! < \infty, \qquad \|h\|^2_\mu = \sum_{k=0}^{\infty} |a_k|^2 \mu^k k!.$$

For functions on $\mathbb{R}^n$, the corresponding space is denoted by $F^2_\mu(\mathbb{R}^n)$. With multi-index notation, for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$,

$$|\alpha| = \sum_{i=1}^{n} \alpha_i, \qquad \alpha! = \prod_{i=1}^{n} \alpha_i!, \qquad x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i}.$$

The monomials $x^\alpha \mu^{|\alpha|}/\sqrt{\alpha!}$ form an orthonormal basis for $F^2_\mu(\mathbb{R}^n)$.

If $\mu_2 < \mu_1$, then differential operators from $F^2_{\mu_1}(\mathbb{R}^n)$ to $F^2_{\mu_2}(\mathbb{R}^n)$ are compact, and multiplication operators with polynomial symbols are bounded [34]. Consequently, if f is componentwise polynomial, then

$$A_f : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_2}(\mathbb{R}^n), \qquad A_f g = \nabla g \cdot f,$$

is compact. This compact Liouville formulation addresses one of the main weaknesses of direct Koopman-DMD approximations: it provides an operator setting in which finite-rank approximations can converge in operator norm, which is better suited for spectral approximation than strong convergence alone.

2.4 Occupation Kernels and Their Interaction with Liouville Operators

Occupation kernels encode trajectory information as elements of an RKHS. Let H be an RKHS of continuous functions on a compact set $X \subset \mathbb{R}^n$, and let $\gamma : [0, T] \to X$ be a continuous trajectory. The functional

$$g \mapsto \int_0^T g(\gamma(t))\, dt$$

is bounded on H. Therefore, by the Riesz representation theorem, there exists a unique function $\Gamma_\gamma \in H$ such that

$$\langle g, \Gamma_\gamma \rangle_H = \int_0^T g(\gamma(t))\, dt, \qquad g \in H. \tag{2.4.1}$$

The function Γ_γ is called the occupation kernel corresponding to γ in H [40, 67].

The reproducing property gives an explicit formula for evaluating an occupation kernel:

$$\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t))\, dt. \tag{2.4.2}$$

Moreover, if $\gamma_i : [0, T_i] \to X$ and $\gamma_j : [0, T_j] \to X$, then

$$\langle \Gamma_{\gamma_j}, \Gamma_{\gamma_i} \rangle_H = \int_0^{T_i} \int_0^{T_j} K(\gamma_i(\tau), \gamma_j(t))\, dt\, d\tau. \tag{2.4.3}$$

Thus, the Gram matrices needed for computation can be assembled using only kernel evaluations along measured trajectories.

Occupation kernels are especially useful for system identification because they replace numerical differentiation with integration. Suppose γ is a solution of $\dot{x} = f(x)$ and the components f_i of f belong to H. Then

$$\langle f_i, \Gamma_\gamma \rangle_H = \int_0^T f_i(\gamma(t))\, dt = \int_0^T \dot{\gamma}_i(t)\, dt = \gamma_i(T) - \gamma_i(0). \tag{2.4.4}$$

The right-hand side depends only on trajectory endpoints, not on numerical estimates of $\dot{\gamma}$. This is the fundamental reason occupation-kernel methods are less sensitive to measurement noise than derivative-based regression methods.

Occupation kernels also interact directly with Liouville operators. If $\gamma : [0, T] \to X$ is a solution of $\dot{x} = f(x)$, then

$$A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)), \tag{2.4.5}$$

under the appropriate domain assumptions [33]. This identity follows from the chain rule and the definition of the occupation kernel:

$$\langle A_f g, \Gamma_\gamma \rangle_H = \int_0^T (A_f g)(\gamma(t))\, dt = \int_0^T \frac{d}{dt} g(\gamma(t))\, dt = g(\gamma(T)) - g(\gamma(0)).$$

Using the reproducing property, $g(\gamma(T)) - g(\gamma(0)) = \langle g, K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H$, which yields (2.4.5). This identity is the basis for Liouville DMD: occupation kernels represent the range side of the finite-rank approximation, while endpoint kernel differences represent the action of the adjoint operator on data.

2.5 Vector-Valued RKHSs

Controlled systems require a function space that can represent vector-valued quantities. Let $\mathcal{Y}$ be a Hilbert space and let X be a set.

Definition 2.5.1 *A vector-valued reproducing kernel Hilbert space (vvRKHS) H over X with output space $\mathcal{Y}$ is a Hilbert space of functions from X to $\mathcal{Y}$ such that, for each $x \in X$ and $v \in \mathcal{Y}$, the functional*

$$h \mapsto \langle h(x), v \rangle_{\mathcal{Y}}$$

is bounded.

By the Riesz representation theorem, for each $x \in X$ and $v \in \mathcal{Y}$ there exists a function $K_{x,v} \in H$ such that

$$\langle h(x), v \rangle_{\mathcal{Y}} = \langle h, K_{x,v} \rangle_H, \qquad h \in H. \tag{2.5.1}$$

The mapping $v \mapsto K_{x,v}$ is linear; hence it may be represented as an operator K_x by writing $K_x v := K_{x,v}$. The operator-valued kernel associated with H is

$$K(x, y) = K_y^* K_x. \tag{2.5.2}$$

In the control-affine setting used in this dissertation, $\mathcal{Y} = \mathbb{R}^{m+1}$ is viewed as a space of row vectors, $X = \mathbb{R}^n$, and elements of H are continuously differentiable row-vector-valued functions. With this convention, the action of the kernel operator is written as vK_x for $v \in \mathbb{R}^{m+1}$.

2.6 Controlled Systems, Control Occupation Kernels, and Control Liouville Operators

Consider a continuous-time control-affine system

$$\dot{x} = f(x) + g(x)u, \tag{2.6.1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $f : \mathbb{R}^n \to \mathbb{R}^n$ is the drift vector field, and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the control effectiveness matrix. Let $\theta : [0, T] \to \mathbb{R}^n$ be a state signal and $u : [0, T] \to \mathbb{R}^m$ be a control signal. The control occupation kernel associated with (θ, u) is the unique element $\Gamma_{\theta,u} \in H$ satisfying

$$\langle h, \Gamma_{\theta,u} \rangle_H = \int_0^T h(\theta(t)) \begin{pmatrix} 1 \\ u(t) \end{pmatrix} dt, \qquad h \in H. \tag{2.6.2}$$

Control occupation kernels were introduced to incorporate input data directly into the occupation-kernel framework [48].

For higher-order systems of the form

$$\frac{d^s x}{dt^s} = f(x) + g(x)u, \tag{2.6.3}$$

the control occupation kernel can be generalized using Cauchy's iterated integral formula. For $s \in \mathbb{N}$, define $\Gamma^{(s)}_{\theta,u} \in H$ by

$$\left\langle h, \Gamma^{(s)}_{\theta,u} \right\rangle_H = \frac{1}{(s-1)!} \int_0^T (T-t)^{s-1} h(\theta(t)) \begin{pmatrix} 1 \\ u(t) \end{pmatrix} dt. \tag{2.6.4}$$

If γ_u is a controlled trajectory satisfying (2.6.3) and the row vectors $\begin{pmatrix} (f)_j & (g)_j \end{pmatrix}$ belong to H, then

$$\left\langle \begin{pmatrix} (f)_j & (g)_j \end{pmatrix}, \Gamma^{(s)}_{\gamma_u,u} \right\rangle_H = \left(\gamma_u(T) - \sum_{\ell=0}^{s-1} \frac{T^\ell}{\ell!} \frac{d^\ell \gamma_u}{dt^\ell}(0) \right)_j . \tag{2.6.5}$$

For $s = 1$, this reduces to the first-order control-affine case. For $s > 1$, it allows certain higher-order systems to be treated without augmenting the state with derivatives.

The control Liouville operator is the operator counterpart of (2.6.1). For a vvRKHS H of row-vector-valued functions, the control Liouville operator with symbol (f, g) is defined by

$$A_{f,g} h(x) = \nabla_x h(x) \begin{pmatrix} f(x) & g(x) \end{pmatrix}, \tag{2.6.6}$$

with canonical domain

$$\mathcal{D}(A_{f,g}) = \{h \in H : A_{f,g} h \in H\}. \tag{2.6.7}$$

The interaction between control occupation kernels and the control Liouville operator parallels the autonomous identity (2.4.5). If γ_u is a controlled trajectory of (2.6.1) under the admissible control signal u, then

$$A^*_{f,g} \Gamma_{\gamma_u,u} = K(\cdot, \gamma_u(T)) - K(\cdot, \gamma_u(0)) \tag{2.6.8}$$

under the appropriate domain assumptions [68]. Indeed,

$$\begin{aligned} \langle A_{f,g} h, \Gamma_{\gamma_u,u} \rangle_H &= \int_0^T \nabla h(\gamma_u(t)) \begin{pmatrix} f(\gamma_u(t)) & g(\gamma_u(t)) \end{pmatrix} \begin{pmatrix} 1 \\ u(t) \end{pmatrix} dt \\ &= \int_0^T \nabla h(\gamma_u(t)) \dot{\gamma}_u(t)\, dt = h(\gamma_u(T)) - h(\gamma_u(0)). \end{aligned}$$

Using the reproducing property yields (2.6.8). This identity motivates the use of control occupation kernels in control Liouville DMD and in control occupation kernel regression. If the vvRKHS is universal and the measured trajectories and control signals are sufficiently rich, then continuous drift and control-effectiveness functions can be approximated using finite combinations of control occupation kernels.

2.7 Discrete Control Liouville Models and Parameter Augmentation

The same RKHS and vvRKHS constructions can also be used for discrete-time controlled systems. Consider a discrete-time control-affine system

$$z_{k+1} = f_0(z_k, \rho) + g_0(z_k, \rho)u_k, \tag{2.7.1}$$

where $\rho \in \mathbb{R}^p$ denotes a vector of physical parameters. When the dependence of f_0 and g_0 on ρ is unknown, the parameters may be treated as latent components of an augmented state. The augmented variable is embedded into an RKHS through a kernel map, and the input dependence is encoded through a vvRKHS. The resulting discrete control Liouville representation provides a data-driven model that can be evaluated at candidate parameter values and compared with an observed query trajectory. This framework is used later in the dissertation for parameter identification: the candidate parameter vector whose predicted trajectory best matches the query data is selected as the parameter estimate.

This parameter-augmented viewpoint differs from classical parameter identification, where one typically assumes that unknown parameters enter through known regressors or known nonlinear functions. Here, the RKHS model is trained to encode the relationship between parameters, inputs, and state transitions directly from data.

2.8 Carleman Lifting

The dissertation also uses Carleman lifting, which is a state-lifting technique rather than an observable-lifting technique. Consider an analytic nonlinear system

$$\dot{x} = f(t, x), \tag{2.8.1}$$

where $x \in \mathbb{R}^d$ and $f(t, 0) = 0$. If f admits a Maclaurin expansion in the state variable,

$$f(t, x) = \sum_{\alpha \in \mathbb{Z}_+^d \setminus \{0\}} f_\alpha(t) x^\alpha, \tag{2.8.2}$$

then Carleman lifting maps x to the collection of monomials in x. For truncation order N, the lifted state contains all unrepeated monomials up to order N:

$$z = l_N(x) = \begin{bmatrix} x_1 & \cdots & x_d & x_1^2 & x_1x_2 & \cdots & x_d^2 & \cdots \end{bmatrix}^\top. \tag{2.8.3}$$

In the infinite-dimensional limit, the lifted state satisfies a linear system

$$\dot{z} = Az. \tag{2.8.4}$$

In computation, the lifted system is truncated to a finite dimension. Under suitable decay assumptions on the Maclaurin coefficients, the trajectory of the truncated lifted system can approximate the trajectory of the original nonlinear system over a prescribed time interval [6,52]. This provides a complementary lifting framework to the Koopman/Liouville methods discussed above: instead of lifting observables into an RKHS, Carleman lifting constructs an explicit monomial state embedding.

For data-driven identification, one estimates the finite-dimensional lifted matrix A from measured trajectories. If $\gamma_i : [0, T] \to \mathbb{R}^d$ are measured trajectories and $l_N(\gamma_i(t))$ are their lifted versions, then the integrated form of the lifted linear system suggests the least-squares approximation

$$\widehat{A} = (\Gamma(T) - \Gamma(0))\, I_\Gamma^\dagger, \tag{2.8.5}$$

where $\Gamma(t)$ concatenates the lifted measured trajectories and $I_\Gamma = \int_0^T \Gamma(\tau)\, d\tau$. The Carleman chapter uses this construction together with prescribed-error analysis to select a truncation order and to identify a lifted linear model whose projected trajectories approximate the original nonlinear trajectories over a desired time horizon.

2.9 Summary

The material in this chapter provides the common mathematical foundation for the dissertation. DMD and EDMD approximate Koopman-type operators in finite-dimensional spaces of observables. RKHSs supply a function-space framework in which kernel evaluations, projections, and trajectory-dependent functions can be computed from data. Liouville

operators provide a continuous-time alternative to fixed-time Koopman operators and can be compact when the domain and range RKHSs are chosen appropriately. Occupation kernels turn trajectory integrals into RKHS inner products and connect directly to adjoints of Liouville operators. Vector-valued RKHSs, control occupation kernels, and control Liouville operators extend these ideas to control-affine systems. Finally, Carleman lifting provides a complementary monomial state-lifting framework with finite-dimensional truncations and prescribed-error analysis.

CHAPTER III

CONTROL OCCUPATION KERNEL REGRESSION FOR NONLINEAR CONTROL-AFFINE SYSTEMS

3.1 Introduction

This chapter presents an algorithm for obtaining an approximation of a nonlinear high order control affine dynamical system. Controlled trajectories of the system are leveraged as the central unit of information via embedding them in vector-valued reproducing kernel Hilbert space (vvRKHS). The trajectories are embedded as the so-called higher order control occupation kernels which represent an operator on the vvRKHS corresponding to iterated integration after multiplication by a given controller. The solution to the system identification problem is then the unique solution of an infinite dimensional regularized regression problem. The representer theorem is then used to express the solution as finite linear combination of these occupation kernels, which converts an infinite dimensional optimization problem to a finite dimensional optimization problem. The vector valued structure of the Hilbert space allows for simultaneous approximation of the drift and control effectiveness components of the control affine system. Several experiments are performed to demonstrate the effectiveness of the developed approach.

3.2 Problem Statement

The objective of this chapter is to learn an unknown higher order control affine system from observed integrable control signals and controlled trajectories, $\{u_j : [0, T_j] \to \mathbb{R}^m\}_{j=1}^M$ and

$\{\gamma_{u_j} : [0, T_j] \to \mathbb{R}^n\}_{j=1}^M$, respectively, satisfying

$$\frac{d^s \gamma_{u_j}}{dt^s}(t) = f(\gamma_{u_j}(t)) + g(\gamma_{u_j}(t))u_j(t) \tag{3.2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is the drift function, $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the control effectiveness matrix, and $s \in \mathbb{N}$ is the order of the system.

For several selections of f, g, and s, the formulation in (3.2.1) specializes to several different system identification problems. When $g \equiv 0$, the problem reduces to determining the unknown dynamics f for a higher order dynamical system. When $s = 1$ the problem becomes a system identification problem for a first order control-affine system. If $s = 1$ and $g \equiv 0$, the method developed in this chapter agrees with that of [46] for integer-order systems with $q = 1$.

Systems of the form (3.2.1) encompass $s-$th order linear systems and Euler-Lagrange models with invertible inertia matrices, and hence, represent a wide class of physical plants, including but not limited to robotic manipulators and autonomous ground, aerial, and underwater vehicles.

In order to facilitate the description of the controlled dynamical system in terms of operators, a vector valued Reproducing Kernel Hilbert Space (vvRKHS) framework is utilized in this chapter.

3.3 Higher Order Occupation Kernels

We use the vvRKHS notation and first-order control occupation kernel construction introduced in Chapter II. The only new object needed in this chapter is the higher-order control occupation kernel.

Definition 3.3.1 *For a given continuous signal $\theta : [0, T] \to \mathbb{R}^n$, and a vector-valued RKHS as above, the* control occupation kernel of order $s \in \mathbb{N}$ corresponding to θ *in H, denoted by $\Gamma_{\theta,u}^{(s)}$ is given as the unique function that represents the bounded functional $h \mapsto \frac{1}{(s-1)!}\int_0^T (T -$*

$t)^{s-1}h(\theta(t))\begin{pmatrix}1\\u(t)\end{pmatrix}dt$ *as*

$$\langle h, \Gamma^{(s)}_{\theta,u}\rangle_H = \frac{1}{(s-1)!}\int_0^T (T-t)^{s-1}h(\theta(t))\begin{pmatrix}1\\u(t)\end{pmatrix}dt.$$

This definition enables the treatment of particular higher order dynamical systems without state augmentation via the following proposition.

Proposition 3.3.1 *If H is a vvRKHS consisting of $\mathbb{R}^{m+1}$ (row) valued continuous functions over $\mathbb{R}^n$, $\gamma_u : [0,T] \to \mathbb{R}^n$ is a controlled trajectory with continuous controller $u : [0,T] \to \mathbb{R}^m$ satisfying, for all $t \in [0,T]$, $\frac{d^s\gamma_u}{dt^s}(t) = f(\gamma_u(t)) + g(\gamma_u(t))u(t)$, and $\begin{pmatrix}(f)_j & (g)_j\end{pmatrix} \in H$ for each $j = 1,\ldots,n$, then $\langle \begin{pmatrix}(f)_j & (g)_j\end{pmatrix}, \Gamma^{(s)}_{\theta_u,u}\rangle_H = \left(\gamma_u(T) - \sum_{\ell=0}^{s-1}\frac{T^\ell}{\ell!}\frac{d^\ell\gamma_u}{dt^\ell}(0)\right)_j$.*

Proof. Since $\begin{pmatrix}(f(\gamma_u(t)))_j & (g(\gamma_u(t)))_j\end{pmatrix}\begin{pmatrix}1\\u(t)\end{pmatrix} = (f(\gamma_u(t)))_j + (g(\gamma_u(t)))_j u(t) = \left(\frac{d^s\gamma_u}{dt^s}(t)\right)_j$, and $\Gamma^{(s)}_{\gamma_u,u}$ implements Cauchy's iterated integral formula through the inner product of the Hilbert space, the proposition follows through iterated application of the fundamental theorem of calculus. ■

While Proposition 3.3.1 follows from a simple application of the fundamental theorem of calculus, it sets the stage for a powerful approximation routine leveraging higher order control occupation kernels. These kernels can be implemented directly using a direct interpolation approach, or they can arise naturally in a regularized regression problem.

Section 3.4 explores a regularized regression approach for learning higher order control affine dynamical systems, where the dynamics may be expressed as a linear combination of higher order control occupation kernels. Therefore, it is necessary to have a cogent method for evaluating the higher order control occupation kernels, and this may be realized through inner products with the kernels e_jK_x, where e_j is the j-th cardinal basis function for $\mathbb{R}^{m+1}$.

Proposition 3.3.2 *Fix $s \in \mathbb{N}$. Let θ and u be continuous signals from $[0,T]$ to $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Let H be a vector valued RKHS as above, then*

$$\Gamma^{(s)}_{\theta,u}(x) = \frac{1}{(s-1)!}\int_0^T (T-t)^{s-1}\begin{pmatrix}1 & u^\top(t)\end{pmatrix} K_{\theta(t)}(x)dt. \tag{3.3.1}$$

Proof. Consider $\langle \Gamma^{(s)}_{\theta,u}(x), v\rangle_{\mathcal{Y}} = \langle \Gamma^{(s)}_{\theta,u}, vK_x\rangle_H$, which follows from the definition of a vector valued RKHS. Symmetrically, since H is a real valued RKHS, it follows that

$$\begin{aligned}\langle \Gamma^{(s)}_{\theta,u}(x),v\rangle_{\mathcal{Y}} &= \langle \Gamma^{(s)}_{\theta,u}, vK_x\rangle_H = \langle vK_x, \Gamma^{(s)}_{\theta,u}\rangle_H = \\ &\frac{1}{(s-1)!}\int_0^T (T-t)^{s-1}[vK_x](\theta(t))\begin{pmatrix}1\\ u(t)\end{pmatrix}dt \\ = &\left\langle \frac{1}{(s-1)!}\int_0^T (T-t)^{s-1}\begin{pmatrix}1 & u(t)^\top\end{pmatrix}K_{\theta(t)}(x)dt, v\right\rangle_{\mathbb{R}^{m+1}}.\end{aligned}$$

Hence, the equality holds for all $v \in \mathbb{R}^{m+1}$, and the result follows. ■

If variable length trajectories are admitted, then it can be shown that the span of the occupation kernels corresponding to trajectories θ_u that result from the application of the control u to the system in (3.2.1) is dense in H.

Proposition 3.3.3 *For any order s, the span of the set*

$$A_s \coloneqq \left\{ \Gamma^{(s)}_{\gamma_{u,\gamma_0},u} \;\middle|\; \begin{array}{c}\gamma_{u,\gamma_0} \in C([0,T_u];\mathbb{R}^n), \gamma_0 \in \mathbb{R}^n, \\ u \in C([0,T_u];\mathbb{R}^m), T_u \in [0,T]\end{array}\right\}$$

is dense in H, where γ_{u,γ_0} is a solution of (3.2.1)*, under the control u and starting from γ_0.*

Proof. Select γ_0 such that $h(\gamma_0) \neq 0$. Continuity of h and γ_{u,γ_0} can then be invoked to conclude that for any constant control signal $u(t) = b$ such that $h(\gamma_0)\begin{pmatrix}1\\ b\end{pmatrix} \neq 0$, there exists a $T_u > 0$ for which $\int_0^{T_u} h(\gamma_{u,\gamma_0}(t))\begin{pmatrix}1\\ b\end{pmatrix} dt \neq 0$. For example, one can select $T_u = \min\left\{T, \inf_t\left\{h(\gamma_{u,\gamma_0}(t))\begin{pmatrix}1\\ b\end{pmatrix} = 0\right\}\right\}$. A straightforward extension of the above argument

to the higher order case allows us to conclude that for any order s and any nonzero h, there exist γ_0, u, and T_u such that $\left\langle h, \Gamma^{(s)}_{\gamma_u,\gamma_0,u} \right\rangle_H \neq 0$. That is, $H \cap A_s^\perp = \{0\}$, and as a result, $(A_s^\perp)^\perp = H$. Since $(A_s^\perp)^\perp = \overline{\operatorname{span} A_s}$, we conclude that $H = \overline{\operatorname{span} A_s}$. ■

Proposition 3.3.3 motivates the use of control occupation kernels for system identification. If H is universal, then any continuous function can be approximated, uniformly over any compact set, by a function in H, and any function in H can be approximated using a linear combination of sufficiently many control occupation kernels.

3.4 A Regression Method using Control Occupation Kernels for Control Affine Systems

The control occupation kernel regression (COKR) method developed in this chapter solves a regularized regression problem posed over the vector valued RKHS, H. Through the representer theorem, it is shown that higher order control occupation kernels arise naturally as a collection of basis functions for approximating the control affine dynamics.

$$\begin{aligned}\left\langle \begin{pmatrix} (f)_i & (g)_i \end{pmatrix}, \Gamma^{(s)}_{\gamma_u,u} \right\rangle_H &= \left(\gamma_u(T) - \sum_{\ell=0}^{s-1} \frac{T^\ell}{\ell!} \frac{d^\ell \gamma_u}{dt^\ell}(0) \right)_i \\ &= \frac{1}{(s-1)!} \int_0^T (T-t)^{s-1} \left((f(\gamma_u(t)))_i + (g(\gamma_u(t)))_i u(t) \right) dt,\end{aligned}$$

given $\lambda > 0$ and the controllers and controlled trajectories from Section 3.2, the regularized regression problem to determine an approximation of the i-th row of f and g within the vector valued RKHS is given as

$$\begin{aligned}\min_{\begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix} \in H} \sum_{j=1}^{M} \Bigg[&\left\langle \begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix}, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle_H \\ &- \left(\gamma_{u_j}(T_j) - \sum_{\ell=0}^{s-1} \frac{T_j^\ell}{\ell!} \frac{d^\ell \gamma_{u_j}}{dt^\ell}(0) \right)_i \Bigg]^2 + \lambda \left\| \begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix} \right\|_H^2, \end{aligned} \tag{3.4.1}$$

where $\lambda > 0$ is a regularization parameter, and

$$\frac{1}{(s-1)!} \int_0^{T_j} (T_j - t)^{s-1} \Big((\hat{f})_i(\gamma_{u_j}(t)) + (\hat{g})_i(\gamma_{u_j}(t)) u_j(t) \Big) dt$$

is equal to the inner product $\left\langle \begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix}, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle_H$. Using the Representer Theorem for occupation kernels (cf. [69] and [46, Proposition 1]), the minimizer of (3.4.1) may be expressed as a linear combination of occupation kernels

$$\begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix} = w_{i,1}\Gamma^{(s)}_{\gamma_{u_1},u_1} + \cdots + w_{i,M}\Gamma^{(s)}_{\gamma_{u_M},u_M}.$$

Using this representation of the minimizer, the inner products in the optimization problem can be computed as

$$\langle \begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix}, \Gamma^{(s)}_{\gamma_{u_j},u_j} \rangle_H = \left\langle \sum_{k=1}^{M} w_{i,k}\Gamma^{(s)}_{\gamma_{u_k},u_k}, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle_H = (w)_i (G)^j$$

where $(w)_i \coloneqq \begin{pmatrix} w_{i,1} & \dots & w_{i,M} \end{pmatrix}$ denotes the $i-$th row of the weight matrix $w \in \mathbb{R}^{n\times M}$, $(G)^j$ denotes the $j-$th column of the Gram matrix $G \in \mathbb{R}^{M\times M}$, whose $i,j-$th element, denoted by $G_{i,j}$, is given by $G_{i,j} = \left\langle \Gamma^{(s)}_{\gamma_{u_i},u_i}, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle_H$. Furthermore, the norm in the optimization problem can be computed as $\| \begin{pmatrix} (\hat{f})_i & (\hat{g})_i \end{pmatrix} \|^2_H = (w)_i G (w)_i^\top$. Hence, the resolution (3.4.1) reduces to the finite dimensional convex optimization problem

$$\min_{(w)_i \in \mathbb{R}^{1\times M}} \sum_{j=1}^{M} \left(\left[(w)_i (G)^j - (D)_i^j \right]^2 + \lambda (w)_i G (w)_i^\top \right). \tag{3.4.2}$$

where $(D)_i^j$ is the $i,j-$th element of the end point difference matrix $D \in \mathbb{R}^{n\times M}$, whose $j-$th column is $(D)^j = \gamma_{u_j}(T_j) - \sum_{\ell=0}^{s-1} \frac{T_j^\ell}{\ell!} \frac{d^\ell \gamma_{u_j}}{dt^\ell}(0)$.

The optimization problem in (3.4.2) is written with respect to a single dimension in the state space. Once all the state space dimensions are concatenated, the combined problem can be solved via the resolution of the linear system

$$(G + \lambda I_M)\, w^\top = D^\top. \tag{3.4.3}$$

The resultant approximation is given as $\begin{pmatrix} \hat{f}(x) & \hat{g}(x) \end{pmatrix} = \sum_{j=1}^{M} (w)^j \Gamma^{(s)}_{\gamma_{u_j},u_j}(x)$ where $(w)^j$ denotes the $j-$th column of w.

3.5 Functional ridge regression

3.5.1 COKR and kernel ridge regression

The COKR technique can be interpreted as a generalized kernel ridge regression (KRR) technique. Given a scalar-valued RKHS $\tilde{H}$ with reproducing kernel $\tilde{K}$, functions $\{K(\cdot, x_i)\}_{i=1}^M \subset \tilde{H}$, and outputs $\{y_i\}_{i=1}^M \subset \mathbb{R}$ that presumably satisfy $y_i = \langle h, \tilde{K}(\cdot, x_i)\rangle_{\tilde{H}} = h(x_i)$, kernel ridge regression solves the minimization problem

$$\min_{h\in\tilde{H}} \left\{ \sum_{i=1}^{M} \left[\langle h, \tilde{K}(\cdot, x_i)\rangle_{\tilde{H}} - y_i \right]^2 + \lambda \|h\|_{\tilde{H}}^2 \right\}.$$

The representer theorem then implies that given a new function $\tilde{K}(\cdot, x)$, the KRR predictor satisfies $\delta_{KRR} \coloneqq \langle h_{KRR}, \tilde{K}(\cdot, x)\rangle_{\tilde{H}} = Y^\top (\tilde{G}+\lambda I_M)^{-1} (\langle \tilde{K}(\cdot, x_1), \tilde{K}(\cdot, x)\rangle_{\tilde{H}} \ \dots \ \langle \tilde{K}(\cdot, x_M), \tilde{K}(\cdot, x)\rangle_{\tilde{H}})^\top$, where Y is a column vector comprised of the outputs y_i (see, for example, [70, Appendix A]).

In the developed COKR method, we are given a set of functions $\{\Gamma^s_{\gamma_{u_i}, u_i}\}_{i=1}^M$ and a set of constants $(D)_i^j$ that satisfy $\left\langle (h)_i, \Gamma^{(s)}_{\gamma_{u_j}, u_j} \right\rangle_H = (D)_i^j$ with $(h)_i = \begin{pmatrix} (f)_i & (g)_i \end{pmatrix}$. We then proceed to solve

$$\min_{(h)_i \in H} \sum_{j=1}^{M} \left(\left[\left\langle (h)_i, \Gamma^{(s)}_{\gamma_{u_j}, u_j} \right\rangle_H - (D)_i^j \right]^2 + \lambda \|(h)_i\|_H^2 \right).$$

Using the representer theorem, we can also show that given a new function $\Gamma^s_{\gamma_u, u}$, the COKR predictor $(\hat{h})_i$ satisfies

$$(\delta)_i \coloneqq \langle (\hat{h})_i, \Gamma^s_{\gamma_u, u}\rangle_H = (D)_i (G + \lambda I_M)^{-1} \begin{pmatrix} \langle \Gamma^{(s)}_{\gamma_{u_1}, u_1}, \Gamma^{(s)}_{\gamma_u, u}\rangle_H & \dots & \langle \Gamma^{(s)}_{\gamma_{u_M}, u_M}, \Gamma^{(s)}_{\gamma_u, u}\rangle_H \end{pmatrix}^\top.$$

The COKR method can therefore be seen as a generalization of KRR, where the control occupation kernels replace the reproducing kernels.

The developed generalization is useful in problems where direct samples of functions to be approximated are not available for training. If inner products of the unknown functions against a set of basis functions can be computed, and if that set of basis functions satisfies the representer theorem, then the generalized KRR method can be used to generate useful

estimates of the unknown functions. The affine system identification problem studied in this chapter is one example of approximation problems that have this property. The price paid for the generalization is that instead of performance guarantees in terms of function evaluation, we get performance guarantees in terms of inner product evaluation against the set of basis used for the representation. Performance in terms of function evaluation then depends on the selected set of basis, and needs to be analyzed separately.

Performance guarantees developed for KRR in terms of function evaluation can be interpreted as performance guarantees in terms of inner product evaluation via the reproducing property. Such re-interpretation allows generalization of these performance guarantees to COKR. For instance, for a set of j trajectories, let $G_j \in \mathbb{R}^{j\times j}$ denote the Gram matrix of the control occupation kernels, $D_j \in \mathbb{R}^{n\times j}$ denote the end point difference matrix, and $\Gamma_j \in \mathbb{R}^{j-1}$ denote the column vector $\begin{pmatrix} \langle \Gamma^{(s)}_{\gamma_{u_1},u_1}, \Gamma^{(s)}_{\gamma_{u_j},u_j}\rangle_H & \dots & \langle \Gamma^{(s)}_{\gamma_{u_{j-1}},u_{j-1}}, \Gamma^{(s)}_{\gamma_{u_j},u_j}\rangle_H \end{pmatrix}^\top$. The first $j-1$ trajectories can then be used to predict the i-th component of the end point difference for the $j-$th trajectory, denoted by $(D_j)_i^j$, using COKR. The predicted end point difference is given by $(\delta_j)_i \coloneqq \langle \hat{h}_i, \Gamma^s_{\gamma_{u_j},u_j}\rangle_H = (D_{j-1})_i(G_{j-1}+\lambda I_{j-1})^{-1}\Gamma_j$. A straightforward adaptation of the proof of [71, Theorem 3] then establishes the following proposition

Proposition 3.5.1 *The cumulative end point difference prediction loss satisfies*

$$\sum_{j=1}^{M} \frac{((\delta_j)_i - (D_j)_i^j)^2}{1+\nu_j} = \min_{(h)_i\in H} \left(\sum_{j=1}^{M} \left[\left\langle (h)_i, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle_H - (D_j)_i^j \right]^2 + \lambda \|(h)_i\|_H^2 \right),$$

where $\nu_j = \frac{\langle \Gamma^{(s)}_{\gamma_{u_j},u_j}, \Gamma^{(s)}_{\gamma_{u_j},u_j}\rangle_H - \Gamma_j^\top (G_{j-1}+\lambda I_{j-1})^{-1}\Gamma_j}{\lambda}$.

Furthermore, similar to KRR, it can be shown that in the limit as M goes to infinity, the COKR predictor is at least as good as any continuous function for prediction of end point differences.

Proposition 3.5.2 *If H is universal and X is compact then for any continuous function $h : X \to \mathbb{R}^{1\times m+1}$,*

$$\limsup_{M\to\infty} \frac{1}{M} \left\{ \sum_{j=1}^{M} \left((\delta_j)_i - (D_j)_i^j\right)^2 - \sum_{j=1}^{M} \left[\left\langle h, \Gamma^{(s)}_{\gamma_{u_j},u_j} \right\rangle - (D_j)_i^j \right]^2 \right\} \leq 0$$

Proof. Follows from Proposition 3.5.5 below. ■

3.5.2 A generalization of kernel ridge regression

The system identification problem can be generalized to other applications where the objective is to approximate an unknown function g using the data $\{(F_j, y_j)\}_{j=1}^M$ where $\{F_j\}_{j=1}^M \subset H^*$ are bounded linear functionals such that $[F_j](g) = y_j$. In the functional ridge regression (FRR) approach, we aim to find a function $\hat{g} \in H$ that solves

$$\min_{g \in H} \sum_{i=1}^{M} \left([F_i](g) - y_i \right)^2 + \lambda \|g\|_H^2. \tag{3.5.1}$$

The FRR problem admits a solution in terms of functions that represent the functionals in a RKHS through the Riesz representation theorem.

Proposition 3.5.3 *If f_i is the representative of F_i for all i then there exist constants $\{c_i\}_{i=1}^M \subseteq \mathbb{R}$ such that the minimizer of* (3.5.1) *is given by $\hat{g} = \sum_{i=1}^M c_i f_i$.*

Proof. write $g = g_1 + g_2$, with $g_1 \in \text{span}\{f_i\}_{i=1}^M := S$ and $g_2 \in S^\perp$. Then, the first term in the cost, $\sum_{i=1}^M \left([F_i](g) - y_i \right)^2 = \sum_{i=1}^M \left(\langle g_1, f_i \rangle + \langle g_2, f_i \rangle - y_i \right)^2 = \sum_{i=1}^M \left(\langle g_1, f_i \rangle - y_i \right)^2$, is independent of g_2. The second term is given by $\|g\|_H^2 = \langle g_1 + g_2, g_1 + g_2 \rangle_H = \langle g_1, g_1 \rangle_H + 2\langle g_1, g_2 \rangle_H + \langle g_2, g_2 \rangle = \|g_1\|_H^2 + \|g_2\|_H^2$. Since the first term is independent of g_2 and the second term is monotonic in $\|g_2\|_H^2$, for any $\hat{g}$ that minimizes the cost, $\|g_2\|_H^2 = 0$. ■

The vector of constants $c = \begin{pmatrix} c_1, & \ldots, & c_M \end{pmatrix}^\top$ in the FRR minimizer are then given by the resolution of the linear system

$$(G_f + \lambda I_M) c = y,$$

where $y = \begin{pmatrix} y_1, & \ldots, & y_M \end{pmatrix}^\top$ denotes the vector of measurements and $G_f \in \mathbb{R}^{M \times M}$ is the Gram matrix of the representing functions $\{f_i\}_{i=1}^M$, whose $i,j-$th element is $\langle f_i, f_j \rangle_H$. Given a new $F \in H^*$, with representative $f \in H$, the action of F on the unknown function g can then be approximated as $[F](g) \approx [F](\hat{g}) = c^\top \begin{pmatrix} \langle f, f_1 \rangle_H, & \ldots, & \langle f, f_M \rangle_H \end{pmatrix}^\top$. Note that for

implementation of FRR, all that is required is the ability to compute inner products between the representing functions.

The COKR problem studied in this chapter is an example of FRR where the functionals F_j are the integration functionals $\begin{bmatrix} f & g \end{bmatrix} \mapsto \left\langle \begin{bmatrix} f & g \end{bmatrix}, \Gamma_{\gamma_{u_j}, u_j} \right\rangle_H$ and the representatives are the control occupation kernels $\Gamma_{\gamma_{u_j}, u_j}$. The KRR problem is another example of FRR, where the functionals F_j are the evaluation functionals $f \mapsto [E_{x_j}](f) = f(x_j)$ and the representatives are data-centered kernels $K(\cdot, x_j)$.

The FRR framework applies to other problems, similar to the system identification problem studied in this chapter, where sensors measure cumulative information about an unknown scalar or vector field. For example sensors measure cumulative effects of fields in applications such as radiation exposure monitoring [72], computed tomography [73], and motion tomography [44].

Consider a mobile sensor making M sensing runs, where in the $i-$th sensing run, the average value $y_i = \sum_{j=1}^{N} \frac{g(x_{i,j})}{N}$ of an unknown scalar field g over N measurement locations $\{x_{i,j}\}_{j=1}^{N}$ is returned. The corresponding functional is $[F_i](g) = \sum_{j=1}^{N} \frac{g(x_{i,j})}{N}$. It is easy to see that the representing function of this functional in an RKHS with reproducing kernel k is given by $f_i = \sum_{j=1}^{N} \frac{k(\cdot, x_{i,j})}{N}$. Furthermore, the inner products needed for the FRR solution can be computed as $\langle f_i, f_j \rangle_H = \langle \sum_{k=1}^{N} \frac{k(x_{i,k}, \cdot)}{N}, \sum_{k=1}^{N} \frac{k(x_{j,k}, \cdot)}{N} \rangle_H = \sum_{k,l=1}^{N} \frac{k(x_{i,k}, x_{j,l})}{N^2}$. As such, given a new set of sensing locations $\{z_j\}_{j=1}^{N}$, the corresponding functional $[F_z](g) = \sum_{j=1}^{N} \frac{g(z_j)}{N}$, and its representing function $f_z = \sum_{j=1}^{N} \frac{k(\cdot, z_j)}{N}$, the average value of the unknown scalar field over the sensing locations can be predicted using the FRR predictor as $\sum_{j=1}^{N} \frac{g(z_j)}{N} \approx \begin{pmatrix} \langle f_z, f_1 \rangle_H, & \ldots, & \langle f, f_M \rangle_H \end{pmatrix}^{\top}$

Another example is the continuous time version of the previous example where the sensor returns the average value of a scalar field g for an M number of trajectories. each trajectory has an interval $[0, T]$, and the measurement is $y_i = \frac{1}{T} \int_0^T g(x_i(t))dt$, $i = 1, \ldots, M$. The corresponding Functional is $[F_i](g) = \frac{1}{T} \int_0^T g(x_i(t))dt$, which results in the representing function in the RKHS with the reproducing kernel k of $f_i = \frac{1}{T} \int_0^T k(x_i(t), \cdot)dt$. Additionally,

the inner products needed for the FRR solution are $\langle f_i, f_j \rangle_H = \frac{1}{T^2} \int_0^T \int_0^T k(x_i(t), x_j(\tau)) dt d\tau$. Therefore, given a new trajectory z, the corresponding functional $[F_z](g) = \frac{1}{T} \int_0^T g(z(t)) dt$, and its representing function $f_i = \frac{1}{T} \int_0^T k(z(t), \cdot) dt$, the average value of the unknown scalar field g over the trajectory z can be predicted using the FRR predictor as $\frac{1}{T} \int_0^T g(z(t)) dt \approx \begin{pmatrix} \langle f_z, f_1 \rangle_H, & \ldots, & \langle f, f_M \rangle_H \end{pmatrix}^\top$.

3.5.3 Cumulative error bounds for FRR

In this section, we show that cumulative error guarantees for KRR derived in results such as [74, Theorem 11.11] also hold for FRR. To that end, let C with the supremum norm $\|\cdot\|_\infty$ denote the Banach space of continuous functions from a compact set X to $\mathbb{R}^n$ and let H be a RKHS of $\mathbb{R}^n-$valued continuous functions defined on X. Let $\{F_j\}_{j=1}^\infty \subset C^*$ be a family of functionals that satisfies the following assumption.

Assumption 3.5.1 *There exist positive constants $\overline{F} < \infty$ such that for all $j \in \{1, \ldots, \infty\}$, $f \in C$, and $g \in H$, $|[F_j](f)| \leq \overline{F} \|f\|_\infty$ and $|[F_j](g)| \leq \overline{F} \|g\|_H$. Furthermore, there exists a positive constant $\overline{Y} < \infty$ such that for all $j \in \{1, \ldots, \infty\}$, $|y_j| \leq \overline{Y}$.*

If the measurements y_j correspond to the action of the functionals on a continuous scalar field g defined on a compact set, then the existence of $\overline{Y}$ follows from uniform boundedness of the functionals. To facilitate the convergence analysis, let $\hat{g}_j = \arg\min_{g \in H} \sum_{i=1}^{j-1} \left([F_i](g) - y_i \right)^2 + \lambda \|g\|_H^2$ be the FRR minimizer and let $[F_j](\hat{g}_j)$ be the FRR prediction of the action of the $j-$th functional on the unknown function g, both computed using the first $j-1$ functionals and the corresponding $j-1$ measurements. Under assumption 3.5.1, proposition 3.5.1 can be generalized to obtain the following result.

Proposition 3.5.4 *The cumulative prediction loss satisfies*

$$\sum_{j=1}^{M} \frac{\left(y_j - [F_j](\hat{g}_j) \right)^2}{1 + \nu_j} = \min_{g \in H} \sum_{j=1}^{M} \left([F_j](g) - y_j \right)^2 + \lambda \|g\|_H^2,$$

where $\nu_j = \frac{\langle f_j, f_j \rangle_H - \overline{f}_j^\top \left(G_{f,j-1} + \lambda I\right)^{-1} \overline{f}_j}{\lambda}$, $\overline{f}_j^\top = \begin{pmatrix} \langle f_1, f_j \rangle_H & \cdots & \langle f_{j-1}, f_j \rangle_H \end{pmatrix}$, *and* $G_{f,j-1}$ *is the Gram matrix of the first* $j-1$ *representing functions.*

Proof. The proof follows from a straightforward adaptation of the proof of [71, Theorem 3]. ■

Additionally, it can be shown that if the RKHS H is universal, then in the limit as M goes to infinity, the FRR predictor is at least as good as any continuous function in the cumulative prediction error metric.

Proposition 3.5.5 *If* H *is universal,* X *is compact, and assumption 3.5.1 is satisfied then for any continuous function* $h : X \to \mathbb{R}^{1\times m+1}$,

$$\limsup_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \left(\left([F_j] \left(\hat{g}_j \right) - y_j \right)^2 - \left([F_j] \left(h \right) - y_j \right)^2 \right) \leq 0.$$

Proof. Given $\epsilon > 0$ and a continuous function g, let $\overline{g}_\epsilon \in H$ be a function that satisfies $\sup_{x\in X} \|g(x) - \overline{g}_\epsilon(x)\| \leq \epsilon$. The probabilistic interpretation in section 5 of [71] implies that $\langle f_j, f_j \rangle_H - \overline{f}_j^\top \left(G_{f,j-1} + \lambda I\right)^{-1} \overline{f}_j > 0$ and as a result, proposition 3.5.4 implies that

$$\sum_{j=1}^{M} \left(y_j - [F_j] \left(\hat{g}_j \right) \right)^2 < \min_{h\in H} \sum_{j=1}^{M} \left([F_j] \left(h \right) - y_j \right)^2 + \lambda \|h\|_H^2.$$

In particular, with $h = \overline{g}_\epsilon$,

$$\sum_{j=1}^{M} \left(y_j - [F_j] \left(\hat{g}_j \right) \right)^2 - \sum_{j=1}^{M} \left([F_j] \left(g \right) - y_j \right)^2 + \sum_{j=1}^{M} \left(\left([F_j] \left(g \right) - y_j \right)^2 - \left([F_j] \left(\overline{g}_\epsilon \right) - y_j \right)^2 \right) < \lambda \|\overline{g}_\epsilon\|_H^2,$$

Expanding the squares in the last term on the left hand side, using linearity of F_j and the identity $\left([F_j] \left(g \right) \right)^2 - \left([F_j] \left(\overline{g}_\epsilon \right) \right)^2 = \left([F_j] \left(g + \overline{g}_\epsilon \right) \right) \left([F_j] \left(g - \overline{g}_\epsilon \right) \right)$.

Using Assumption 3.5.1, we obtain

$$\sum_{j=1}^{M} \left(y_j - [F_j] \left(\hat{g}_j \right) \right)^2 - \sum_{j=1}^{M} \left([F_j] \left(g \right) - y_j \right)^2 < \lambda \|\overline{g}_\epsilon\|_H^2 + \sum_{j=1}^{M} \overline{F}^2 \left\| g + \overline{g}_\epsilon \right\|_\infty \left\| g - \overline{g}_\epsilon \right\|_\infty + 2\overline{F} \left\| \overline{g}_\epsilon - g \right\|_\infty \overline{Y}$$

Adding and subtracting g to the first term in the second row,

$$\sum_{j=1}^{M} \left(y_j - [F_j]\left(\hat{g}_j\right)\right)^2 - \sum_{j=1}^{M} \left([F_j]\left(g\right) - y_j\right)^2 < \lambda \|\overline{g}_\epsilon\|_H^2 + \sum_{j=1}^{M} \left(\overline{F}^2 2\left\|g\right\|_\infty + 2\overline{FY}\right) \left\|\overline{g}_\epsilon - g\right\|_\infty + \left\|\overline{g}_\epsilon - g\right\|_\infty^2 .$$

Since $\sup_{x \in X} \|g(x) - \overline{g}_\epsilon(x)\| \leq \epsilon$,

$$\frac{1}{M}\left(\sum_{j=1}^{M} \left(y_j - [F_j]\left(\hat{g}_j\right)\right)^2 - \sum_{j=1}^{M} \left([F_j]\left(g\right) - y_j\right)^2\right) < \frac{\lambda}{M}\|\overline{g}_\epsilon\|_H^2 + \left(\overline{F}^2 \left\|2g\right\|_\infty + 2\overline{FY}\right)\epsilon + \epsilon^2$$

Taking the limit superior as $M \to \infty$, and then taking the limit as $\epsilon \to 0$, the proposition is established. ■

Proposition 3.5.5 shows that the FRR solution converges to a function that performs better than any continuous function, with respect to the cumulative average prediction error. However, proposition 3.5.5 does not imply that the functions $\hat{g}_j$ converge to g in the supremum norm, a limitation FRR shares with KRR. The following section nevertheless provides an effective tool for approximation of nonlinear dynamical systems.

3.6 Numerical Examples

3.6.1 Higher order system - an academic example

This experiment utilizes the second order one dimensional nonlinear model of the Duffing oscillator given by

$$\ddot{x} = (x - x^3) + (2 + \sin(x))u, \tag{3.6.1}$$

where $f(x) = (x - x^3)$ is the drift function, $g(x) = 2 + \sin(x)$ is the control effectiveness function, and u is the controller. To approximate the system dynamics, 169 trajectories of the system are recorded, along with the corresponding control signals, starting from a grid of initial conditions, under a control signal that is composed of the sum of three sinusoidal signals with randomly generated frequencies and coefficients. Each trajectory

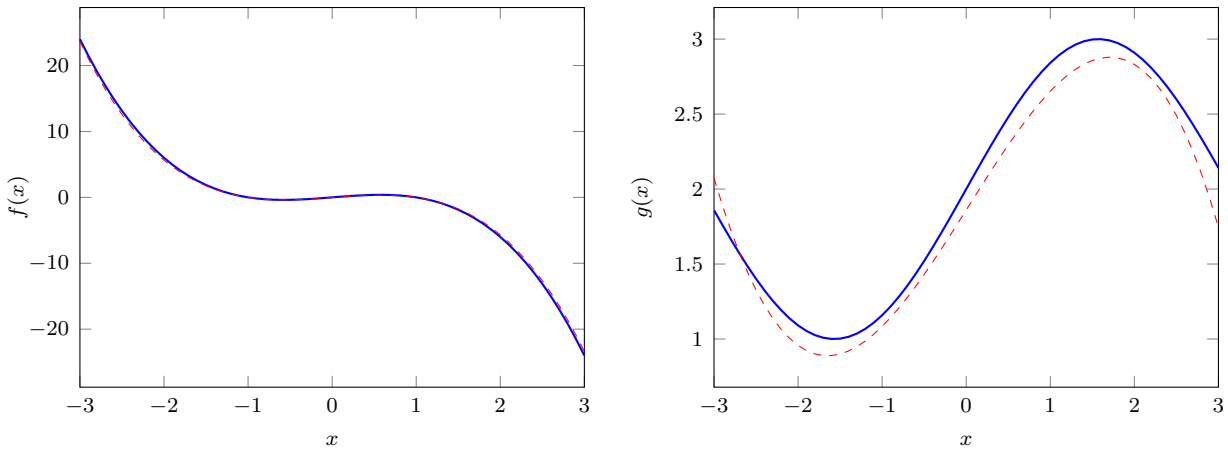

Figure 2: The evaluation of f and $\hat{f}$ (left) and g and $\hat{g}$ (right) for the Duffing oscillator with measurement noise. The solid blue lines represent the true values of the functions and the dotted red lines represent the COKR estimates.

is corrupted by Gaussian measurement noise with standard deviation 0.001. The initial velocities are obtained by numerically differentiating the measured noisy trajectories. The recorded trajectories and control signals are then utilized to approximate f and g.

Figure 2 shows the estimated and actual values of the drift function and the estimated and actual values of the control effectiveness function. Figure 3 shows the drift approximation error, $\tilde{f}$, as a function of x and the control effectiveness approximation error, $\tilde{g}$, as a function of x.

A Monte-Carlo simulation with 1000 trials is conducted to evaluate robustness of the developed method to excitation signals and sensor noise, each trial uses 169 trajectories generated by a control signal that is composed of the sum of three sinusoidal signals with randomly generated frequencies and coefficients. Each trajectory is corrupted by Gaussian measurement noise with standard deviation 0.001. The initial velocities are obtained by numerically differentiating the measured noisy trajectories. For each trial, the max of $\left|\tilde{f}\right|$ over the interval $[-3, 3]$ are collected using COKR and SINDYc for comparison. Box plots for the Monte-Carlo trials are presented in Figure 5 for the maximum errors $\max_{x\in[-3,3]} \left|\tilde{f}\right|$ and $\max_{x\in[-3,3]} |\tilde{g}|$, respectively.

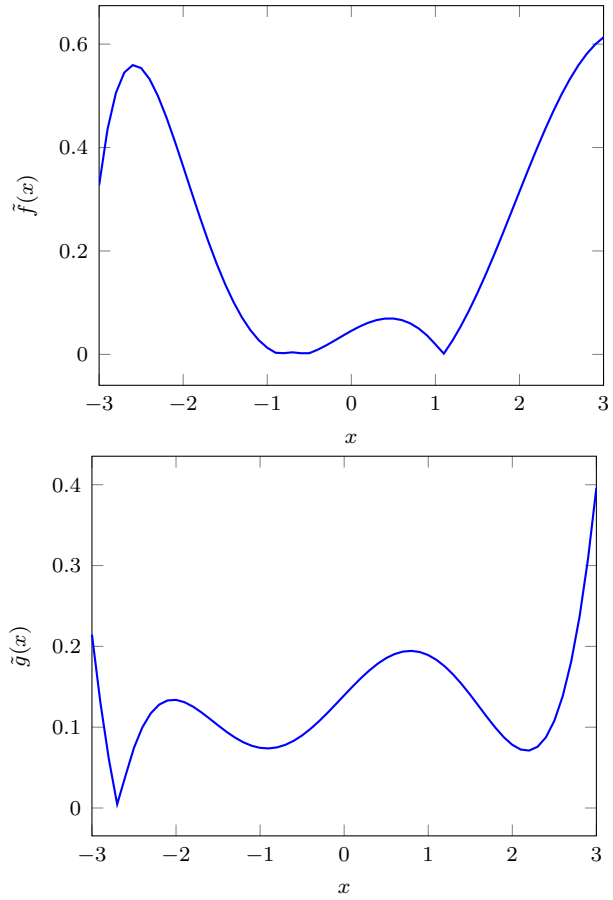

Figure 3: The evaluation of $\left|\tilde{f}\right|$ (left) and $|\tilde{g}|$ (right) for the Duffing oscillator with measurement noise.

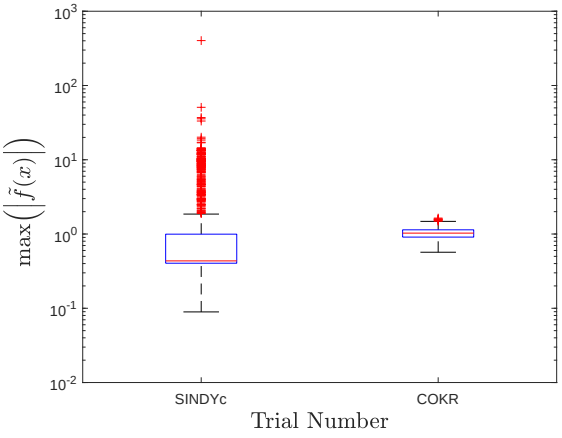

Figure 4: Monte-Carlo results of the maximum of $\left|\tilde{f}\right|$ over 1000 trials

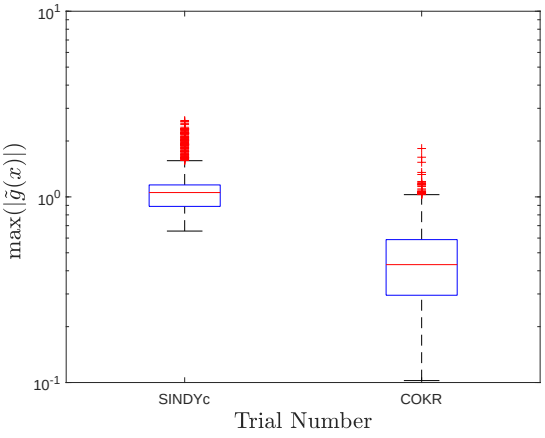

Figure 5: Monte-Carlo results of the maximum of $|\tilde{g}|$ over 1000 trials

3.6.2 Control-affine model of a two-link robot manipulator

This experiment utilizes the first order four dimensional model of a two-link robot manipulator given by

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -M^{-1}(x_1)C(x_1, x_2) \end{pmatrix} + \begin{pmatrix} 0_{2\times 2} \\ M^{-1}(x_1) \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}, \tag{3.6.2}$$

where $x_1 := \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$, $x_2 := \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix}$, q_1, and q_2 are the angular positions of the two links, respectively, $M(x_1) = \begin{pmatrix} p_1 + 2p_3\cos(q_2) & p_2 + p_3\cos(q_2) \\ p_2 + p_3\cos(q_2) & p_2 \end{pmatrix}$ is the inertia matrix, $C(x_1, x_2) = (V(x_1, x_2) + F_d)\, x_2 + F_s(x_2)$, where $F_d = \begin{pmatrix} f_{d_1} & 0 \\ 0 & f_{d_2} \end{pmatrix}$ denotes the viscous friction at the two joints, $F_s(x_2) = \begin{pmatrix} f_{s_1}\tanh(\dot{q}_1) \\ f_{s_2}\tanh(\dot{q}_2) \end{pmatrix}$ denotes the static friction at the two joints, $V(x_1, x_2) = \begin{pmatrix} -p_3\sin(q_2)\dot{q}_2 & -p_3\ \sin(q_2)(\dot{q}_1 + \dot{q}_2) \\ p_3\sin(q_2)\cdot\dot{q}_1 & 0 \end{pmatrix}$ is the centrifugal-Coriolis matrix and τ_1 and τ_2 are the control torques applied by the two motors. The parameters are given by $p_1 = 3.473$, $p_2 = 0.196$, $p_3 = 0.242$, $f_{d_1} = 5.3$, $f_{d_2} = 1.1$, $f_{s_1} = 8.45$, and $f_{s_2} = 2.35$.

Trajectories of the system are collected, starting from 1000 different initial conditions, sampled using pseudorandom Halton sampling over a cube of side 6 centered at the origin of $\mathbb{R}^4$. Each trajectory is recorded using control signals τ_1 and τ_2, comprised of a sum of three sinusoidal signals with randomly generated frequencies and coefficients. For each run, the control signals and the trajectories of the system are recorded to generate the database that is used to approximate f and g. Since this system has two controllers τ_1 and τ_2, the control effectiveness matrix g is decomposed into g_1 and g_2, where g_1 is the first row of g and g_2 is the second row of g.

The approximation error between the identified system and the actual system is computed

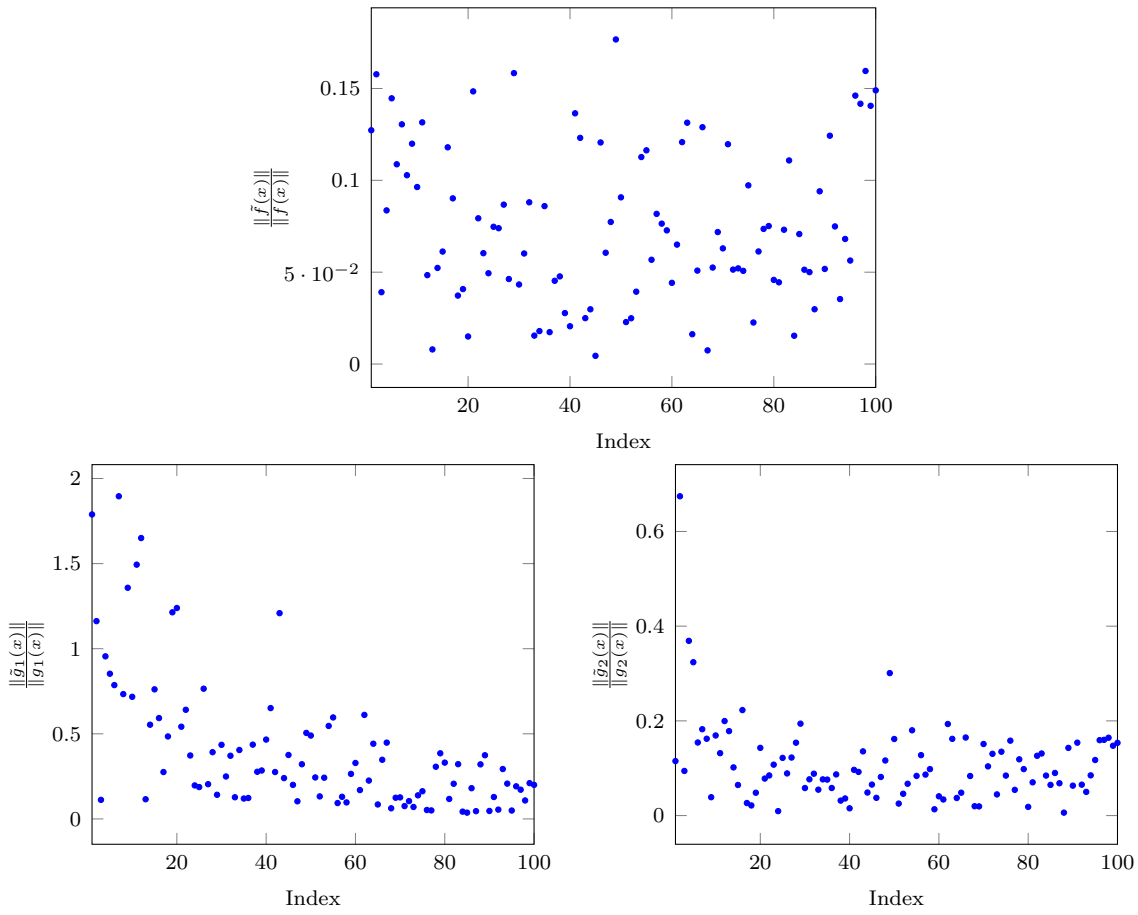

Figure 6: The relative errors $\frac{\|\tilde{f}(x)\|}{\|f(x)\|}$ (top), $\frac{\|\tilde{g}_1(x)\|}{\|g_1(x)\|}$ (bottom left), and $\frac{\|\tilde{g}_2(x)\|}{\|g_2(x)\|}$ (bottom right) evaluated at 100 points, indexed by decreasing distance from the origin.

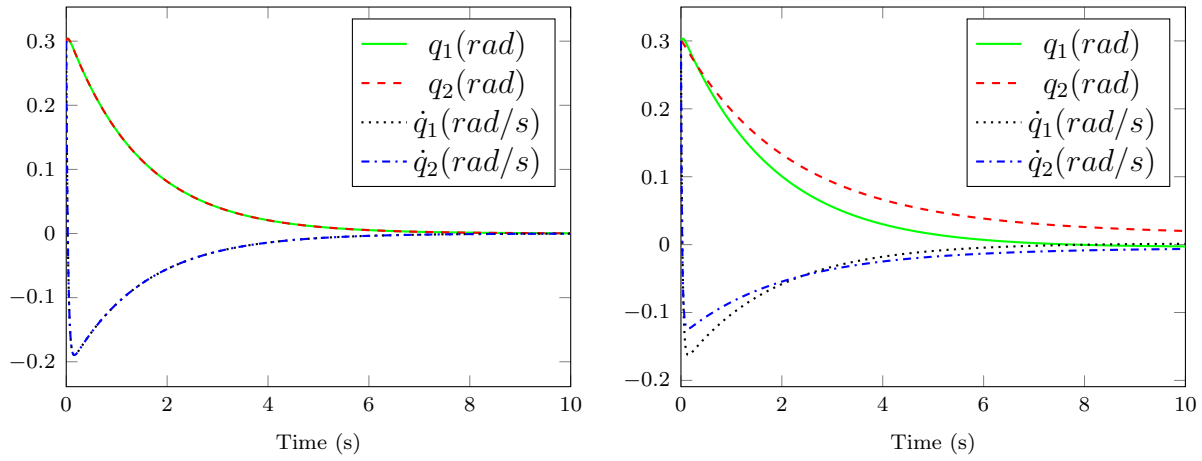

Figure 7: Regulation result using the true model (left) and estimated model (right) for computed torque control.

through evaluation at 100 sample points sampled using pseudorandom Halton sampling over a cube of side 4 centered at the origin of $\mathbb{R}^4$. The functions f, $\hat{f}$, g, and $\hat{g}$ are evaluated at each vertex of the grid to yield the approximation errors $\tilde{f}$, $\tilde{g_1}$, and $\tilde{g_2}$.

Figure 6 shows the relative errors $\frac{\left\|\tilde{f}(x)\right\|}{\|f(x)\|}$, $\frac{\|\tilde{g_1}(x)\|}{\|g_1(x)\|}$, and $\frac{\|\tilde{g_2}(x)\|}{\|g_2(x)\|}$, respectively as a function of the node index.

To demonstrate the utility of the model developed via the novel system identification method, we simulate the actual two-link robot manipulator using a fourth order Runge-Kutta method with a computed torque controller designed to regulate the angular positions and velocities to zero. The computed torque controller is calculated using the estimated model, and is given by

$$\tau = \hat{M}(q)[-K_v\dot{q} - K_p q] + \hat{C}(q, \dot{q}) \tag{3.6.3}$$

where $\hat{M}(q)$ and $\hat{C}(q, \dot{q})$ are recovered from the estimated data-driven drift and control-effectiveness functions $\hat{f}$ and $\hat{g}$, and $K_p = \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix}$, $K_v = \begin{pmatrix} 30 & 0 \\ 0 & 30 \end{pmatrix}$ are feedback gains. Figure 7 shows the performance of the computed torque regulator used on the true two-link robot manipulator model.

3.7 Discussion

In the first numerical experiment, where the Duffing oscillator system is identified, the approximation $\hat{f}$ is nearly identical to the true f as seen in Figure 2 and the maximum error is 0.61, which is a 2.6% error as seen in Figure 3. The approximation $\hat{g}$ captures the underlining structure of g where $g(x) = 2 + \sin(x)$, but $\hat{g}$ deviates slightly from the true value as seen in Figure 2 with a maximum error of 0.4 which is a 23% error as seen in Figure 3. The maximum errors are a function of kernel type, shape parameters, number and spatial coverage of the recorded data, and measurement noise, and can be reduced through segmentation of the recorded trajectories to generate more data points and increase the resolution of the approximations.

	COKR		SINDYc	
Quantity	Mean	SD	Mean	SD
$\max \lvert\tilde{f}(x)\rvert$	1.0322	0.1765	2.5224	13.3437
$\max \lvert\tilde{g}(x)\rvert$	0.4718	0.2346	1.1095	0.3417

Table 1: Comparison of the Monte Carlo simulation results for COKR and SINDYc.

The Monte-Carlo trials summarized in Figure 5 demonstrate the robustness of the developed system identification method to excitation signals and measurement noise. Figure 2 show the full error plots of a representative sample trial.

Figure 5 indicates that COKR is less sensitive to measurement noise than SINDYc. In Table 1 the mean and standard deviation of $\max\left(\left|\tilde{f}(x)\right|\right)$ and $\max(|\tilde{g}(x)|)$ for COKR and SINDYc are listed, which shows that COKR is more precise than SINDYc. This result is expected in a scenario where measurement noise is present, since COKR uses numerical integration where as SINDYc uses numerical differentiation. Although The median of the error measurement of SINDYc for f in 5 is lower than that of COKR, SINDYc results in a significant number of outliers, where the worst preforming run produces an error measurement that is two orders of magnitude higher than COKR.

In the second numerical experiment, where the model of a two-link robot manipulator is identified, Figure 6 indicates that the estimation errors get worse as one approaches the boundary of the domain covered by the trajectories, and better as one approaches the origin. While the errors are hard to visualize as a function of distance from the training data in four dimensions, the trends observed in Figure 6 could heuristically be attributed to limited coverage of the corners of the domain.

The identified system $\begin{pmatrix} \hat{f} & \hat{g} \end{pmatrix}$ is used to compute the necessary torque to regulate the actual two-link robot manipulator, which results in a performance similar to a computed torque controller implemented using exact model knowledge , as seen in Figure 7. The main downside of regulating the system using $\begin{pmatrix} \hat{f} & \hat{g} \end{pmatrix}$ is the long computation time necessary to

evaluate $\begin{pmatrix} \hat{f} & \hat{g} \end{pmatrix}$ at any given x, which highlights the need for optimization of the evaluation function to improve computational performance.

3.8 Conclusion

A data-driven control occupation kernel regression method is developed in this chapter for identification of nonlinear affine control systems, where identification of higher order systems is possible without numerical differentiation for higher order systems. As a result, as indicated by the numerical experiments, the developed method is robust to measurement noise. While the model developed using this method can be used to compute the feedforward component of a controller, depending on the number of trajectories used for modeling, evaluation of the model at a given state can be computationally expensive. Further research is required to develop a more efficient evaluation method to render it useful for real-time feedback.

A distinct advantage of the use of an occupation kernel basis to approximate the dynamics of the system is that under the assumptions of the framework developed in this chapter, the solution of the minimization problem in (3.4.1) is guaranteed to be a linear combination of the occupation kernels by the representer theorem. This means that the occupation kernels are natural basis functions arising from a dynamics context, and in principle they should perform better than less structured bases. In particular, through the representer theorem, it is clear that the occupation kernel basis will, in general, result in models that yield a lower modeling error in the Hilbert space norm than any other generic basis, resulting in better generalization of the estimates away from the training data.

CHAPTER IV

CARLEMAN LIFTING FOR NONLINEAR SYSTEM IDENTIFICATION WITH GUARANTEED ERROR BOUNDS

4.1 Introduction

This chapter concerns identification of uncontrolled or closed-loop nonlinear systems using a set of trajectories that are generated by the system in a domain of attraction. The objective is to ensure that the trajectories of the identified systems are close to the trajectories of the real system, as quantified by an error bound that is *prescribed a priori*. A majority of existing methods for nonlinear system identification rely on techniques such as neural networks, autoregressive moving averages, and spectral decomposition that do not provide systematic approaches to meet predefined error bounds. The developed method is based on Carleman linearization-based lifting of the nonlinear system to an infinite dimensional linear system. The linear system is then truncated to a suitable order, computed based on the prescribed error bound, and parameters of the truncated linear system are estimated from data. The effectiveness of the technique is demonstrated by identifying an approximation of the Van der Pol oscillator from data within a prescribed error bound.

4.2 Problem Statement

Consider an unknown dynamical system of the form

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}) \tag{4.2.1}$$

where $\mathbf{f} : \left(\mathbb{R}^+ \times \mathbb{R}^d\right) \to \mathbb{R}^d$ is a vector-valued real analytic function of several variables, $\mathbb{R}^+ \coloneqq \{y \in \mathbb{R} : y \geq 0\}$, $\mathbf{x}(0) = \mathbf{x}_0$, and $\mathbf{f}(t, 0) = 0$. Given a set of observed trajectories, $X = \{\gamma\}_{i=1}^m$, generated from (4.2.1) and an error bound $\Delta > 0$, the objective is to develop a systematic technique to either

1. construct another dynamical system of the form $\dot{\hat{z}} = g(t, \hat{z})$ such that the error $\|\mathbf{x}(t) - \hat{z} \mid_d (t)\|$ between the trajectories of the $z-$system and the trajectories of (4.2.1) is less than Δ for $t \in [0, \tau^*]$, for some $\tau^* > 0$, where $\hat{z} \mid_d$ is the truncation of $\hat{z}$ to the first d dimensions, or

2. conclude that construction of such a system, using the particular method developed in this chapter, is not possible.

The problem, as formulated above, is difficult to solve for general nonlinear systems. In this chapter, inspired by [52], the formulation is restricted to a subclass of nonlinear systems, defined by the following assumptions.

Assumption 4.2.1 *The vector field* $\mathbf{f}(t, \mathbf{x})$ *admits a Maclaurin expansion about the state vector* x

$$\mathbf{f}(t, \mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{Z}_+^d} \mathbf{f}_{\boldsymbol{\alpha}}(t)\mathbf{x}^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\alpha} \in \mathbb{Z}_+^d \setminus \{0\}} \mathbf{f}_{\boldsymbol{\alpha}}(t)\mathbf{x}^{\boldsymbol{\alpha}}, t \in \mathbb{R}^+. \tag{4.2.2}$$

Assumption 4.2.2 *The Maclaurin expansion coefficients satisfy the exponential decay property,*

$$\sup_{t \geq 0} \sum_{|\alpha| = n} \|\mathbf{f}_{\boldsymbol{\alpha}}(t)\|_\infty \leq CR^{-n}, n \geq 0 \tag{4.2.3}$$

where C *and* R *are positive constants.*

While this assumption is restrictive, a large subclass of systems falls under this category, since polynomial, trigonometric, and exponential functions have Maclaurin expansions with

exponential decay. The idea is to use Carleman lifting [6, 52] to lift the nonlinear system into an infinite-dimensional linear system such that the truncation of the trajectory of the linear system to the first d dimensions approximates the trajectory of the system in (4.2.1), in infinity norm. The infinite-dimensional linear system is then truncated to yield a finite-dimensional linear system such that the error between the projected trajectories and the original trajectories is less than the given error bound, Δ.

4.3 Data-driven Carleman Lifting

The nonlinear system is approximated by a linear system in terms of a lifted state z which consists of unrepeated monomials of x up to order N. For example, if $N = 3$ then

$$z = l(x) = \begin{array}{c}[x_1, \ldots, x_d, x_1^2, x_1x_2, \ldots, x_1x_d, x_2^2, \ldots, x_d^2, \\ x_1^3, x_1^2x_2, \ldots, x_1x_2x_3, \ldots, x_d^3]^T,\end{array}$$

where $l : \mathbb{R}^n \to \mathbb{R}^{\mathbb{M}}$ denotes the lifting map and $\mathbb{M}$ is the number of monomials.

The results of [52] indicate that for any system that satisfies Assumptions 4.2.1 and 4.2.2, a trajectory in a region of attraction can be approximated, with arbitrary accuracy, by truncation of solutions of

$$\dot{z} = Az \tag{4.3.1}$$

to the first d dimensions, where computation of A requires complete knowledge of the system dynamics, $\mathbf{f}$.

In this chapter, a data-driven approach to generate the linear system is developed. Given a set of trajectories of the nonlinear system, denoted by $\{\gamma_i\} : [0, T] \to \mathbb{R}^n$, that satisfy $\|\gamma_i(t)\| \leq M$ for all $t \in [0, T]$, the objective is to generate an estimate, $\hat{A}$, of the system matrix A, such that $\sup_{t\in[0,\tau^*]} \|\mathbf{x}(t) - \hat{z} \mid_d (t)\| \leq \Delta$, where $\hat{z} \mid_d$ is the $d-$dimensional truncation of the solution $\hat{z}$ of $\dot{\hat{z}} = \hat{A}\hat{z}$ starting from $\hat{z}(0) = l(x(0))$.

Let z_i denote the trajectory of (4.3.1) starting from the initial condition $z_i(0) = l(\gamma_i(0))$. Let

$$\epsilon_i(t) \coloneqq l(\gamma_i(t)) - z_i(t) \tag{4.3.2}$$

denote the error between the trajectories of the model-based Carleman linearization in (4.3.1) and the trajectories of (4.2.1). The lifted measured trajectories and the trajectories of the linear system are then related by

$$\frac{d}{dt}l(\gamma_i(t)) = \dot{z}_i(t) + \dot{\epsilon}_i(t) = Az_i(t) + \dot{\epsilon}_i(t). \tag{4.3.3}$$

Integrating (4.3.3), we get

$$\begin{aligned}\int_0^T \frac{d}{dt}l(\gamma_i(t)) &= \int_0^T Az_i(\tau)d\tau + \epsilon_i(T) \\ &= A\int_0^T \left(l(\gamma_i(\tau)) - \epsilon_i(\tau)\right)d\tau + \epsilon_i(T).\end{aligned} \tag{4.3.4}$$

Note that $\epsilon_i(0) = 0$. Concatenating all the measured trajectories into a vector $\Gamma(t) = [l(\gamma_1(t)), \ldots, l(\gamma_m(t))]^T$, and letting $I = \int_0^T Z(\tau)d\tau$, $I_\epsilon = \int_0^T \epsilon(\tau)d\tau$ and $I_\Gamma = \int_0^T \Gamma(\tau)d\tau$, we get the relationships $I_\Gamma = I + I_\epsilon$ and

$$\Gamma(T) - \Gamma(0) = AI_\Gamma + AI_\epsilon + \epsilon(T), \tag{4.3.5}$$

where $\epsilon(t) = [\epsilon_1(t), \ldots, \epsilon_m(t)]^T$.

Provided the matrix I_Γ is full rank, a data-driven approximation $\hat{A}$ of the lifted matrix A can be computed using the least-squares solution of (4.3.5) as

$$\hat{A} = \left(\Gamma(T) - \Gamma(0)\right) I_\Gamma^\dagger, \tag{4.3.6}$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo-inverse.

Once the $\hat{A}$ matrix is computed from data, trajectories of the linear system $\dot{\hat{z}} = \hat{A}\hat{z}$ can be computed and truncated to the first d dimensions to estimate the trajectories of the nonlinear system.

4.4 Error Analysis and Prescribed Error Approximation

There are two sources of error between the trajectories of the identified system and the trajectories of the nonlinear system. The first is the error ϵ introduced in (4.3.2), between

the trajectories of the nonlinear system and the model-based Carleman linearized system. The second is the error $\delta(t) \coloneqq z(t) - \hat{z}(t)$, between the trajectories of the model-based Carleman linearized system and the identified linear system, starting from the same initial conditions. The former is quantified in [52] as follows.

Assumption 4.4.1 *The bound M on the measured trajectories satisfies*

$$M < \frac{R}{e},$$

where R is introduced in (4.2.3) and e is the base of the natural logarithm.

Under Assumptions 4.2.1 - 4.4.1, there exists a τ^* for every $N \geq 1$, such that

$$\|\epsilon\,|_d\,(t)\| \leq D\mu^N, \forall t \in [0, \tau^*]. \tag{4.4.1}$$

In (4.4.1),

$$0 < D = \frac{D_M M}{C_0 R},$$
$$D_M = C_0 R \left(1 - \left(\frac{M}{R}\right)\right)^{-1},$$
$$C_0 \leq C R^{-1},$$

and

$$\mu < \left(MeR^{-1}\right) e^{C_0 \tau^*} < 1.$$

The error bound in (4.4.1) can then be guaranteed over the time interval $[0, \tau^*]$ where

$$\tau^* < \frac{-\log(MeR^{-1})}{C_0}.$$

To quantify the error δ, between the trajectories of the identified system and the model-based Carleman linearized system, we write

$$\dot{\tilde{z}} = Az - \hat{A}\hat{z},$$

by adding and subtracting $A\hat{z}$, we get

$$\dot{\tilde{z}} = A\tilde{z} + (A - \hat{A})\hat{z}.$$

Using the variation of constants formula we get

$$\tilde{z}(\tau^*) = e^{A\tau^*}\tilde{z}(0) + \int_0^{\tau^*} e^{A(\tau^*-\tau)}(A-\hat{A})\hat{z}(\tau)d\tau,$$

using the triangle inequality and the Cauchy-Schwarz inequality then result in the bound

$$\|\tilde{z}(\tau^*)\| \leq \int_0^{\tau^*} \left\|e^{A(\tau^*-\tau)}\right\| \left\|A-\hat{A}\right\| \|\hat{z}(\tau)\| \, d\tau.$$

Since $\tilde{z}(0) = 0$, then

$$\|\tilde{z}(\tau^*)\| \leq t\overline{B}\overline{z}\overline{A}$$

where $\overline{B} = \max\left\{\left\|e^{A\tau^*}\right\|, 1\right\}$, $\overline{z} = \sup_{t\in[0,\tau^*]} \|\hat{z}(t)\|$, and $\overline{A} = \left\|A-\hat{A}\right\|$.

In order to calculate $\overline{A}$ we write

$$D(t) = Z(t) - Z(0).$$

From (4.3.1), it is the case that

$$A = DI^T(II^T)^{-1},$$

and from (4.3.6) we calculate the estimation

$$\hat{A} = (D+\epsilon)(I+I_\epsilon)^T[(I+I_\epsilon)(I+I_\epsilon)^T]^{-1}.$$

To compute a bound on δ, we need the following assumption.

Assumption 4.4.2 *There exists a constant $\overline{I}$ such that $\left\|(I^TI)^{-1}\right\| \leq \overline{I}$.*

Note that the matrix I is comprised of integrals of the trajectories of the model-based Carleman linearized system, and as such, it is unknown. However, the matrix I_Γ can be computed, and is perturbed from I by I_ϵ. Using continuity of eigenvalues of matrices with respect to elements of the matrix, it can be concluded that for small enough I_ϵ, the difference between $\left\|(I^TI)^{-1}\right\|$ and $\left\|(I_\Gamma^TI_\Gamma)^{-1}\right\|$ is $O(I_\epsilon)$.

The bound $\overline{A}$ can then be estimated using the matrix inverse identity [75],

$$(Q+V)^{-1} = Q^{-1} - Q^{-1}V(Q+V)^{-1},$$

where Q and $(Q+V)$ are non-singular matrices. Applying the identity to the expression for $\hat{A}$, we get

$$\begin{aligned}\overline{A} \leq \left\|\overline{I}\right\| \left(\left\|I^T\right\| \|\epsilon\| + \left\|I_\epsilon^T\right\| \|D\| + \left\|I_\epsilon^T\right\| \|\epsilon\|\right) \\ + \left[\left\|\overline{I}\right\| \left(\left\|I^T\right\| \|I_\epsilon\| + \left\|I_\epsilon^T\right\| \|I\| + \|I_\epsilon\| \left\|I_\epsilon^T\right\|\right)\right] \\ \cdot \left[\left\|I^T\right\| \|D\| + \left\|I^T\right\| \|\epsilon\| + \left\|I_\epsilon^T\right\| \|D\| + \left\|I_\epsilon^T\right\| \|\epsilon\|\right], \quad (4.4.2)\end{aligned}$$

where $\left\|I^T\right\| \leq \left\|I_\Gamma^T\right\| + \left\|I_\epsilon^T\right\|$.

Since $\overline{B} = \max\{\left\|e^{A\tau^*}\right\|, 1\}$, an estimate of $\left\|e^{A\tau^*}\right\|$ is obtained by realizing that

$$e^{A\tau^*} = e^{(A-\hat{A}+\hat{A})\tau^*} = e^{(A-\hat{A})\tau^*} e^{\hat{A}\tau^*},$$

and as a result,

$$\overline{B} = \left\|e^{A\tau^*}\right\| \leq e^{\overline{A}\tau^*} \left\|e^{\hat{A}\tau^*}\right\|. \quad (4.4.3)$$

By using the triangle inequality we get the bound

$$\|\mathbf{x}(t) - \hat{z} \mid_d (t)\| \leq \|\epsilon \mid_d (t)\| + \|\tilde{z}\|,$$

which can be further simplified as

$$\|\mathbf{x}(t) - \hat{z} \mid_d (t)\| \leq D\mu^N + t\overline{B}\overline{z}\overline{A}, \quad \forall t \in [0, \tau^*]. \quad (4.4.4)$$

Once the upper bound is calculated as a function of N using (4.4.4), a search over $N = 1, 2, \ldots, \overline{N}$ can be conducted (see Algorithm 1), to yield

$$N^* = \arg\min_N (D\mu^N + \tau^*\overline{B}\overline{z}\overline{A}),$$

where N^* denotes the lifted order which produces an identified system with the smallest guaranteed error bound. If $D\mu^{N^*} + \tau^*\overline{B}\overline{z}\overline{A} > \Delta$, then the system identification method cannot produce a system that guarantees $\sup_{t\in[0,\tau^*]} \|x(t) - \hat{z} \mid_d (t)\| < \Delta$. Otherwise, if $D\mu^{N^*} + \tau^*\overline{B}\overline{z}\overline{A} \leq \Delta$, then a system can be identified using the truncation order N^* to guarantee $\sup_{t\in[0,\tau^*]} \|x(t) - \hat{z} \mid_d (t)\| < \Delta$.

Algorithm 1 Carleman system identification algorithm. In the algorithm, $\overline{N}$ is an upper bound on the truncation order, selected a priori.

Require: X, R, C, $\overline{N}$ and Δ

Ensure: $C_0 \leq CR^{-1}$

Ensure: $M > ||x||_\infty$

$D_M \leftarrow C_0 R \left(1 - \left(\frac{M}{R}\right)\right)^{-1}$

$D \leftarrow \frac{D_M M}{C_0 R}$

Ensure: $\tau^* < \frac{-\log(MeR^{-1})}{C_0}$

Ensure: $\mu < (MeR^{-1})\, e^{C_0 \tau^*} < 1$

$N \leftarrow 1$

while $N \neq \overline{N}$ **do**

Find $\overline{A}$ from (4.4.2)

Find $\overline{B}$ from (4.4.3)

$\Theta(N) \leftarrow D\mu^N + \tau^* \overline{A} \overline{z} \overline{B}$

$N \leftarrow N + 1$

end while

if $\min[\Theta(N)] \leq \Delta$ **then**

$N^* \leftarrow \arg\min_N [\Theta(N)]$

Return $\hat{A}$ using N^*

else if $\min[\Theta(N)] > \Delta$ **then**

Return "Failed"

end if

4.5 Simulation

To demonstrate the effectiveness of the system identification approach in Section 4.3, the Van der Pol Oscillator is used. Consider the Van der Pol Oscillator given by

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -x_1 - x_2 + x_2 x_1^2. \end{aligned} \tag{4.5.1}$$

A total of 209 trajectories of the system in (4.5.1), each over the time interval $[0, 10]$, are recorded starting from initial conditions that were uniformly sampled from the set $[-1, 1]\times[-1, 1]$. The trajectories are then lifted to different dimensions, with $N = 2, 3, \ldots, 12$ for system identification.

Using the lifted trajectories, (4.3.6) is used to determine the system matrix $\hat{A}$. The identified system matrix is used to simulate the identified system from $t_0 = 0$ to $t_f = 20$ seconds. The trajectories generated by the identified system are truncated to the first two dimensions and compared with the recorded trajectories of the nonlinear system in (4.5.1). For comparison, trajectories of the linear system obtained using the model-based Carleman linearization in [52] are also generated. Note that the system matrix $\hat{A}$ is computed directly using recorded data from the nonlinear system. The model-based Carleman linearization from [52] is used for comparison and analysis purposes only.

The resulting trajectories of the nonlinear, the model-based Carleman linearized, and the identified systems are shown in Figures 8-10. To quantify the performance, the errors $\sup_{t\in[0,\tau^*]} \|x(t) - \hat{z}\mid_2(t)\|$ between the measured trajectories and the identified system trajectories, and $\sup_{t\in[0,\tau^*]} \|x(t) - z\mid_2(t)\|$, between the measured trajectories and the model-based Carleman linearized trajectories, are plotted in Figure 11. The parameters introduced in Sections 4.2 and 4.4 are selected as $M = 1.5$, $R = 4.1$, $C = 33.7$, $\overline{N} = 13$ and $C_0 = 0.001$, all of which satisfy the conditions in Section 4.4 and in (4.2.3). This choice guarantees an error bound on $\|x(t) - \hat{z}\mid_2(t)\|$ which is given in (4.4.4) over $t \in [0, 0.2]$.

Figure 12 illustrates the data-driven bound on $\sup_{t\in[0,\tau^*]} \|x(t) - \hat{z}\,|_2\,(t)\|$, developed in (4.4.4), compared with the model-based bound on $\sup_{t\in[0,\tau^*]} \|x(t) - z\,|_2\,(t)\|$, developed in [52].

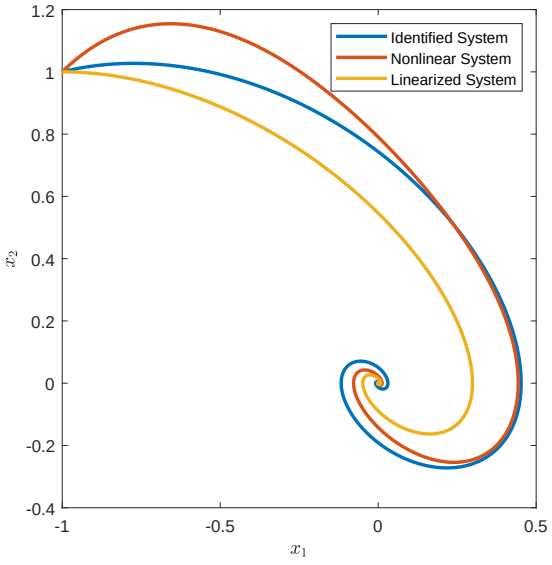

Figure 8: Trajectories of the nonlinear system, the identified system, and the model-based Carleman linearized system using $N = 2$.

4.6 Discussion

The results in Figures 8-11 show that for the model-based Carleman linearized system, increasing the order N produces a more accurate approximation of (4.5.1). In contrast, for the data-driven identified system increasing the order N past a certain point can result in an identified model that is inaccurate, as can be seen in Figure 10. As indicated by the bound in (4.4.4), while the Carleman linearization is guaranteed to get better with increasing truncation order, when used in conjunction with a system identification method, there is a critical point. The optimal truncation order can be estimated using the bound developed in (4.4.4), as illustrated by Figure 12.

Figure 11 shows that $N = 6$ produces the identified system with the lowest error, while

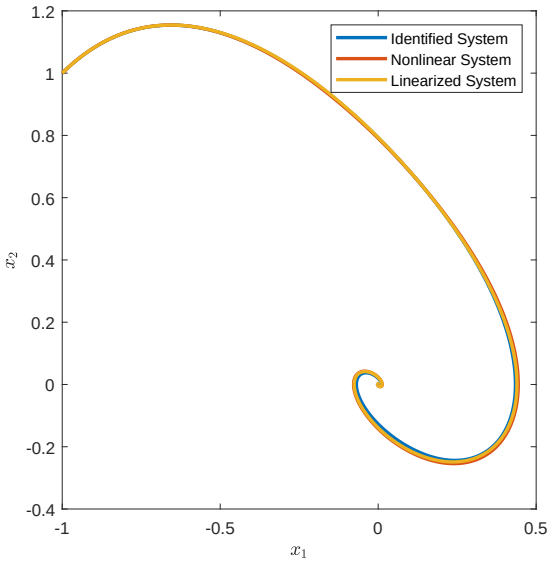

Figure 9: Trajectories of the nonlinear system, the identified system, and the model-based Carleman linearized system using $N = 5$.

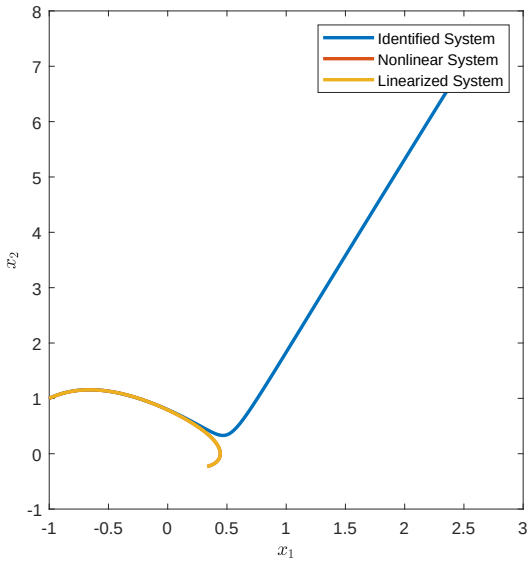

Figure 10: Trajectories of the nonlinear system, the identified system, and the model-based Carleman linearized system using $N = 11$.

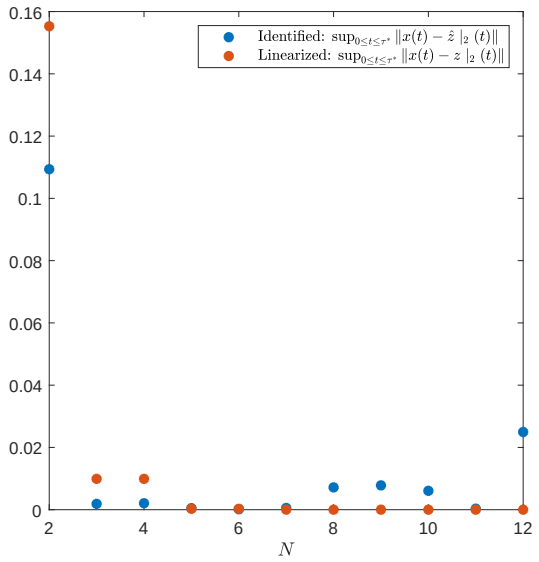

Figure 11: Comparison of errors $\sup_{t\in[0,\tau^*]}\|x(t)-\hat{z}\ |_2\ (t)\|$ between the measured trajectories and the identified trajectories, and $\sup_{t\in[0,\tau^*]}\|x(t)-z\ |_2\ (t)\|$, between the measured trajectories and the model-based Carleman linearized trajectories, for truncation orders $N=1,2,\ldots,12$.

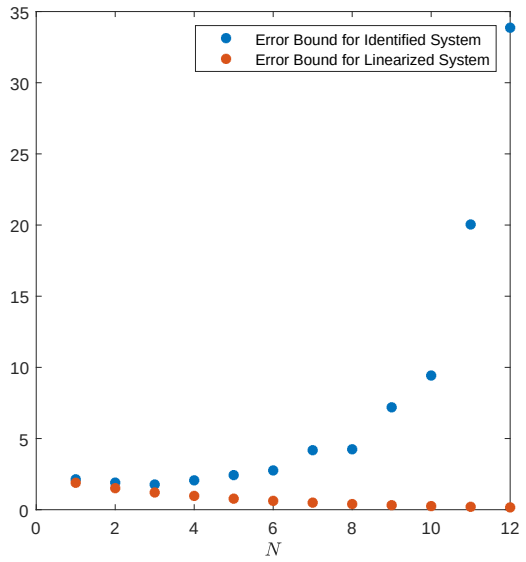

Figure 12: Comparison of analytical error bounds on $\sup_{t\in[0,\tau^*]} \|x(t) - \hat{z}\,|_2\,(t)\|$, the error between the trajectories of the nonlinear system and the identified system, and $\sup_{t\in[0,\tau^*]} \|x(t) - z\,|_2\,(t)\|$, the error between the trajectories of the nonlinear system and the model-based Carleman linearized system, for truncation orders $N = 1, 2, \ldots, 12$.

the analytical bound (see Figure 12) indicates that $N^* = 3$, meaning for $N = 3$ we get the smallest guaranteed error bound. Since the bound is conservative, some discrepancy between the exact numerical results and the analytical bound is expected.

To further explore the discrepancy, it is instructive to plot the actual errors between the full trajectories of the model-based Carleman linearized system and the data-driven identified linear system (see Figure 13) and the errors between the same trajectories, truncated to the first two dimensions (see Figure 14). The results indicate that while the full state error $\sup_{0\leq t\leq \tau^*} \|z(t) - \hat{z}(t)\|$ increases monotonically with N, the truncated state error $\sup_{0\leq t\leq \tau^*} \|z\mid_2(t) - \hat{z}\mid_2(t)\|$ initially decreases and then increases with an increasing N. The analysis presented in this chapter, that produces $N^* = 3$ is based on analyzing the full state error, which is observed to be an overly conservative bound for a subset of truncation orders. This suggests that development of an error bound for the truncated state error will reduce the discrepancy between the analytical and the experimental results.

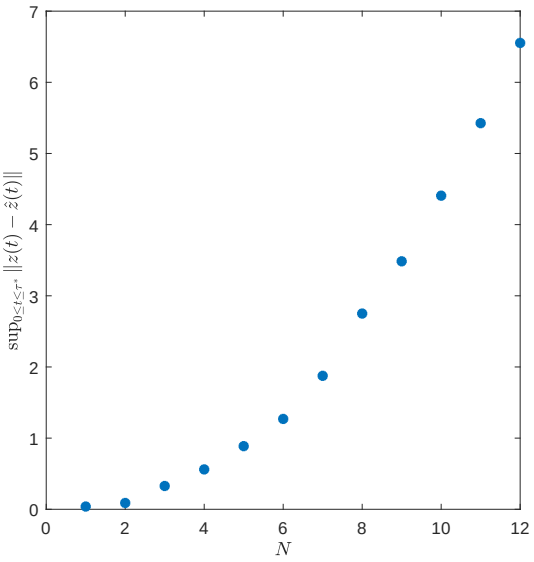

Figure 13: Maximum error norm of the full state between the linearized system and the identified system using $N = 1, 2, \ldots, 12$

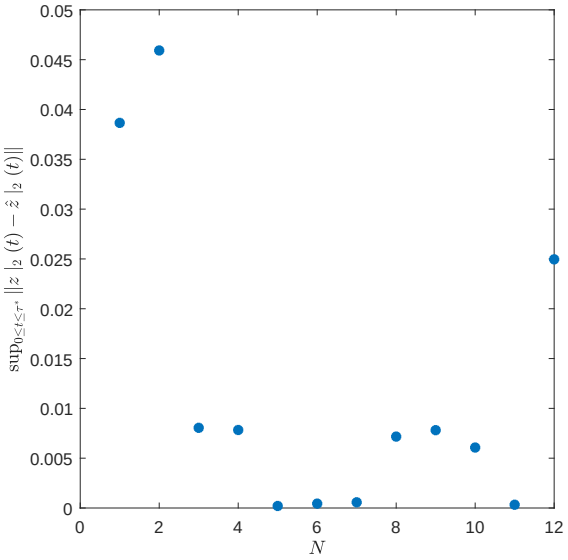

Figure 14: Maximum error norm of the first two states between the linearized system and the identified system using $N = 1, 2, \ldots, 12$

4.7 Conclusion and Future Work

A system identification method using Carleman linearization is developed to identify a lifted linear system that produces trajectories, which remain within a guaranteed error bound from the trajectory of a nonlinear system under mild assumptions. Analytical estimation of truncated state errors and the bound in Assumption 4.4.2, investigation of machine precision effects at high truncation orders, and development of a less conservative error bound for the states of interest are part of future work.

CHAPTER V

ON DYNAMIC MODE DECOMPOSITION OF CONTROL-AFFINE SYSTEMS

5.1 Introduction

This chapter builds on the theoretical foundations for dynamic mode decomposition (DMD) of control-affine dynamical systems by leveraging the theory of vector-valued reproducing kernel Hilbert spaces (RKHSs). Specifically, control Liouville operators and control occupation kernels are used to separate the drift dynamics from the input dynamics. A provably convergent finite-rank estimation of a compact control Liouville operator is obtained, provided sufficiently rich data are available. A matrix representation of the finite-rank operator is used to construct a data-driven representation of its singular values, left singular functions, and right singular functions. The singular value decomposition is used to generate a data-driven model of the control-affine nonlinear system. The developed method generates a model that can be used to predict the trajectories of the system in response to any admissible control input.

5.2 Problem Statement

The objective in this chapter is to learn an unknown control affine system from observed control signals and controlled trajectories, $\{u_j : [0, T_j] \to \mathbb{R}^m\}_{j=1}^M$ and $\{\gamma_{u_j} : [0, T_j] \to \mathbb{R}^n\}_{j=1}^M$, respectively, satisfying

$$\dot{x} = F(x, u) = f(x) + g(x)u, \tag{5.2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is the drift function and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the control effectiveness matrix.

Systems of the form (5.2.1) encompass linear systems and Euler-Lagrange models with invertible inertia matrices, and hence, represent a wide class of physical plants, including but not limited to robotic manipulators and autonomous ground, aerial, and underwater vehicles.

In order to facilitate the description of the controlled dynamical system in terms of operators, the vvRKHS framework from section 2.5 is utilized in this chapter.

Definition 5.2.1 *Given a compact control Liouville operator $A_{f,g} : \tilde{H}_d \to H$, the tuples $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^{\infty}$, with $\sigma_i \in \mathbb{R}^n$, $\phi_i \in \tilde{H}_d$, and $\psi_i \in H$, are singular values, left singular vectors, and right singular vectors of $A_{f,g}$, respectively, if $\forall h \in \operatorname{span} d$,*

$$A_{f,g}h = \sum_{i=1}^{\infty} \sigma_i \psi_i \left\langle h, \phi_i \right\rangle_{\tilde{H}_d}. \tag{5.2.2}$$

Let $h_{\mathrm{id}} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity function. Given singular values, left singular vectors, and right singular vectors of $A_{f,g}$ and a control signal u, the dynamics of the system can be expressed as

$$\dot{x} = \begin{pmatrix} A_{f,g}(h_{\mathrm{id}})_1 \\ \vdots \\ A_{f,g}(h_{\mathrm{id}})_n \end{pmatrix} \begin{pmatrix} 1 \\ u(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{\infty} \sigma_i \psi_i \left\langle (h_{\mathrm{id}})_1, \phi_i \right\rangle_{\tilde{H}_d} \\ \vdots \\ \sum_{i=1}^{\infty} \sigma_i \psi_i \left\langle (h_{\mathrm{id}})_n, \phi_i \right\rangle_{\tilde{H}_d} \end{pmatrix} \begin{pmatrix} 1 \\ u(t) \end{pmatrix} \tag{5.2.3}$$

The objective of this work is to generate a provably convergent finite truncation of the above model. To that end, we derive a finite rank representation of the control Liouville operator.

5.3 Finite-rank Representation of the Control Liouville Operator

To facilitate computation, an explicit finite-rank representation of $A_{f,g}$ is needed to determine the dynamic modes of the resultant system. In the following, finite collections of linearly independent vectors, d^M and β^M are selected to establish the needed finite-rank representation. Since the adjoint of $A_{f,g}$ maps control occupation kernels to kernel differences

([68, Proposition 3]), the span of the collection of kernel differences

$$d^M = \{K_d(\cdot, \gamma_{u_i}(T_i)) - K_d(\cdot, \gamma_{u_i}(0))\}_{i=1}^M \subset \tilde{H}_d \tag{5.3.1}$$

is selected to be the domain of $A_{f,g}$. The corresponding Gram matrix is denoted by $G_{d^M} = \left(\langle d_i, d_j \rangle_{\tilde{H}_d}\right)_{i,j=1}^M$. The output of $A_{f,g}$ is projected onto the span of the control occupation kernels

$$\beta^M = \left\{\Gamma_{\gamma_{u_i}, u_i}\right\}_{i=1}^M \subset H. \tag{5.3.2}$$

The corresponding Gram matrix is denoted by $G_{\beta^M} = \left(\langle \beta_i, \beta_j \rangle_H\right)_{i,j=1}^M$.

A rank-M (or less) representation of the operator $A_{f,g}$ is then given by $P_{\beta^M} A_{f,g} P_{d^M} : \tilde{H}_d \to \operatorname{span} \beta^M$, where P_{d^M} and P_{β^M} denote projection operators onto $\operatorname{span} d^M$ and $\operatorname{span} \beta^M$, respectively.

Under the compactness assumptions and given rich enough data so that the spans of $\{d_i\}_{i=1}^\infty$ and $\{\beta_i\}_{i=1}^\infty$ are dense in $\tilde{H}_d$ and H, respectively, the sequence of finite-rank operators $\{P_{\beta^M} A_{f,g} P_{d^M}\}_{M=1}^\infty$ can be shown to converge, in norm, to $A_{f,g}$. To facilitate the proof of convergence, we introduces the following lemma.

Lemma 5.3.1 *Let H and G be RKHSs defined on $X \subset \mathbb{R}^n$ and let $A_N : H \to G$ be a finite-rank operator with rank N. If the spans of $\{d_i\}_{i=1}^\infty$ and $\{\beta_i\}_{i=1}^\infty$ are dense in H and G, respectively, then for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$ and $h \in H$, $\|A_N h - A_N P_{d^i} h\|_G \leq \epsilon \|h\|_H$ and $\left\|A_N h - P_{\beta^i} A_N h\right\|_G \leq \epsilon \|h\|_H$.*

Proof. See the proof of [33, Theorem 2]. ■

The convergence result for Liouville operators on Bargmann-Fock spaces restricted to the set of real numbers, which will be used for H and $\tilde{H}_d$, follows from the following more general result.

Proposition 5.3.1 *If $A : \tilde{H}_d \to H$ is a compact operator and the spans of $\{d_i\}_{i=1}^\infty$ and $\{\beta_i\}_{i=1}^\infty$ are dense in $\tilde{H}_d$ and H, respectively, then $\lim_{M\to\infty} \left\|A - P_{\beta^M} A P_{d^M}\right\|_{\tilde{H}_d}^H = 0$, where $\|\cdot\|_{\tilde{H}_d}^H$ denotes the operator norm of operators from $\tilde{H}_d$ to H.*

Proof. Let $\{A_N\}_{N=1}^{\infty}$ be a sequence of rank-N operators converging, in norm, to A. For an arbitrary $h \in \tilde{H}_d$,

$$\begin{aligned}\left\|Ah - P_{\beta^M} A P_{d^M} h\right\|_H \leq \|Ah - A_N h\|_H + \|A_N h - A_N P_{d^M} h\|_H \\ + \left\|A_N P_{d^M} h - P_{\beta^M} A_N P_{d^M} h\right\|_H + \left\|P_{\beta^M} A_N P_{d^M} h - P_{\beta^M} A P_{d^M} h\right\|_H .\end{aligned}$$

Using the fact that A_N and $P_{\beta^M} A_N P_{d^M}$ are finite-rank operators, Lemma 5.3.1, can be used to conclude that for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$

$$\left\|Ah - P_{\beta^i} A P_{d^i} h\right\|_H \leq \|Ah - A_N h\|_H + 2\epsilon \|h\|_{\tilde{H}_d} + \|A_N P_{d^i} h - A P_{d^i} h\|_H .$$

Since A_N converges to A in norm, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $j \geq N$, and $g \in \tilde{H}_d$ $\|Ag - A_j g\|_H \leq \epsilon \|g\|_{\tilde{H}_d}$. Thus, for all $j \geq N$ and $i \geq M(j)$, $\left\|Ah - P_{\beta^i} A P_{d^i} h\right\|_H \leq 4\epsilon \|h\|_{\tilde{H}_d}$. ■

The convergence result can then be stated as follows.

Theorem 5.3.1 *Let $\rho_d \in \mathbb{R}$, $\varrho_d \in \mathbb{R}$, and $\rho = \begin{bmatrix} \rho_1 & \dots & \rho_{m+1} \end{bmatrix}^\top \in \mathbb{R}^{m+1}$ be parameters such that $\rho_i < \varrho_d$, and $\varrho_d < \rho_d$ for $i = 1, \dots, m+1$. Let $\tilde{H}_d = F^2_{\rho_d}(\mathbb{R}^n)$, $\tilde{G}_d = F^2_{\varrho_d}(\mathbb{R}^n)$, and $H = F^2_{\rho}(\mathbb{R}^n)$. If f and g are component-wise polynomial, and if the spans of the collections $\{d_i\}_{i=1}^{\infty}$ and $\{\beta_i\}_{i=1}^{\infty}$ are dense in $\tilde{H}_d$ and H, respectively, then $\lim_{M\to\infty} \left\|A_{f,g} - P_{\beta^M} A_{f,g} P_{d^M}\right\|_{\tilde{H}_d}^H = 0$.*

Proof. [68, Propositions 7, 8, and 9] imply that $A_{f,g}$ is compact and hence, the theorem follows from Proposition 5.3.1. ■

5.4 Matrix Representation of the Finite-rank Operator

To formulate a matrix representation of the finite-rank operator $P_{\beta^M} A_{f,g} P_{d^M}$, the operator is restricted to $\operatorname{span} d^M$ to yield the operator $P_{\beta^M} A_{f,g}|_{d^M} : \operatorname{span} d^M \to \operatorname{span} \beta^M$. For brevity of exposition, the superscript M is suppressed hereafter and d and β are interpreted as $M-$dimensional vectors.

Proposition 5.4.1 *If $h = \delta^\top d \in \operatorname{span} d$ is a function with coefficients $\delta \in \mathbb{R}^M$ and if $g = P_\beta A_{f,g} h$, then $g = a^\top \beta$, where $a = G_\beta^+ G_d \delta$ and $(\cdot)^+$ denotes the Moore-Penrose pseudoinverse.*

Proof. [68, Propositions 3 and 6] imply that that for all $j = 1, \cdots, M$, $A_{f,g}^* \beta_j = d_j$. Note that since g is a projection of $A_{f,g}h$ onto $\operatorname{span} \beta$, $g = a^\top \beta$ for any a that solves

$$G_\beta a = \begin{bmatrix} \langle A_{f,g} h, \beta_1 \rangle_H \\ \vdots \\ \langle A_{f,g} h, \beta_M \rangle_H \end{bmatrix} = \begin{bmatrix} \left\langle h, A_{f,g}^* \beta_1 \right\rangle_{\tilde{H}_d} \\ \vdots \\ \left\langle h, A_{f,g}^* \beta_M \right\rangle_{\tilde{H}_d} \end{bmatrix} \tag{5.4.1}$$

Using the adjoint relationship,

$$G_\beta a = \begin{bmatrix} \langle h, d_1 \rangle_{\tilde{H}_d} \\ \vdots \\ \langle h, d_M \rangle_{\tilde{H}_d} \end{bmatrix} = \begin{bmatrix} \left\langle \delta^\top d, d_1 \right\rangle_{\tilde{H}_d} \\ \vdots \\ \left\langle \delta^\top d, d_M \right\rangle_{\tilde{H}_d} \end{bmatrix} = G_d \delta \tag{5.4.2}$$

Selecting the solution

$$a = G_\beta^+ G_d \delta \tag{5.4.3}$$

of (7.4.2), a matrix representation $[P_\beta A_{f,g}]_d^\beta$ of the operator $P_\beta A_{f,g}|_d$ is obtained as $G_\beta^+ G_d$. ■

Note that matrix representations are generally not unique. Different representations may be obtained by selecting different solutions of (7.4.1) and (7.4.3). In the case where the Gram matrix G_β is nonsingular, equation (7.4.1) has a unique solutions, resulting in the unique matrix representation $G_\beta^{-1} G_d$.

In the following section, the matrix representation $[P_\beta A_{f,g}]_d^\beta$ is used to construct a data-driven representation of the singular values and the left and right singular functions of $P_\beta A_{f,g}|_d$.

5.5 Singular Functions of the Finite-rank Operator

The following proposition states that the SVD of $P_\beta A_{f,g}|_d$ can be computed using matrices in the matrix representation $[P_\beta A_{f,g}]_d^\beta$ derived in the previous section.

Proposition 5.5.1 *If* (W, Σ, V) *is the SVD of* G_β^+ *with* $W = \begin{bmatrix} w_1, & \dots, & w_M \end{bmatrix}$, $V = \begin{bmatrix} v_1, & \dots, & v_M \end{bmatrix}$, *and* $\Sigma = \operatorname{diag}\left(\begin{bmatrix} \sigma_1, & \dots, & \sigma_M \end{bmatrix}\right)$, *then for all* $i = 1, \dots, M$, σ_i *are singular values of* $P_\beta A_{f,g}|_d$ *with left singular functions* $\phi_i := v_i^\top d$ *and right singular functions* $\psi_i := w_i^\top \beta$.

Proof. Let $\phi_i = v_i^\top d$ and $\psi_i = w_i^\top \beta$ and $h = \delta^\top d$. Then,

$$P_\beta A_{f,g} h = \sum_{i=1}^{M} \sigma_i \psi_i \left\langle h, \phi_i \right\rangle_{\tilde{H}_d} \iff P_\beta A_{f,g} \delta^\top d = \sum_{i=1}^{M} \sigma_i w_i^\top \beta \left\langle \delta^\top d, v_i^\top d \right\rangle_{\tilde{H}_d}$$

Using the finite-rank representation, the collection $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^M$, is an SVD of $P_\beta A_{f,g}|_d$, if for all $\delta \in \mathbb{R}^M$,

$$\left(G_\beta^+ G_d \delta\right)^\top \beta = \left(\sum_{i=1}^{M} \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{\tilde{H}_d} w_i^\top \right) \beta. \tag{5.5.1}$$

Simple matrix manipulations yield the chain of implications

$$\begin{aligned}
(5.5.1) &\impliedby \forall \delta \in \mathbb{R}^M, G_\beta^+ G_d \delta = \sum_{i=1}^{M} \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{\tilde{H}_d} w_i \\
&\iff \forall \delta \in \mathbb{R}^M, G_\beta^+ G_d \delta = \sum_{i=1}^{M} \sigma_i \left(w_i v_i^\top G_d \right) \delta \\
&\impliedby G_\beta^+ G_d = \sum_{i=1}^{M} \sigma_i \left(w_i v_i^\top \right) G_d \\
&\impliedby G_\beta^+ = \sum_{i=1}^{M} \sigma_i w_i v_i^\top = W \Sigma V^\top,
\end{aligned}$$

which proves the proposition. ∎

In the following section, the singular values and the left and right singular vectors are used, along with a finite truncation of (7.3.1) to generate a data-driven model.

5.6 The SCLDMD Algorithm

Motivated by (7.3.1), assuming that $h_{\mathrm{id},j} \in \tilde{H}_d$ for $j = 1, \cdots, n$, the system dynamics are approximated using the finite-rank representation, with rank at most M, as

$$\dot{x} \approx \hat{F}_M(x,u) := [P_\beta A_{f,g} P_d h_{\mathrm{id}}](x) \begin{pmatrix} 1 \\ u \end{pmatrix},$$

where $P_\beta A_{f,g} P_d h_{\mathrm{id}}$ denotes row-wise operation of the operator $P_\beta A_{f,g}$ on the function $P_d h_{\mathrm{id}}$. Using the definition of singular values and singular functions of $P_\beta A_{f,g} \mid_d$,

$$\dot{x} \approx \sum_{i=1}^{M} \sigma_i \xi_i w_i^\top \beta(x) \begin{pmatrix} 1 \\ u \end{pmatrix} = \xi \Sigma W^\top \beta(x) \begin{pmatrix} 1 \\ u \end{pmatrix}, \tag{5.6.1}$$

where $\xi_i := \langle P_d h_{\mathrm{id}}, \phi_i \rangle_{\tilde{H}_d}$ and $\xi := \begin{bmatrix} \xi_1, & \ldots, & \xi_M \end{bmatrix}$.

The modes ξ can be computed using $\phi_i = v_i^\top d$ as

$$\begin{aligned} \xi &= \begin{bmatrix} \left\langle P_d h_{\mathrm{id},1}, v_1^\top d \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle P_d h_{\mathrm{id},1}, v_M^\top d \right\rangle_{\tilde{H}_d} \\ \vdots & \ddots & \vdots \\ \left\langle P_d h_{\mathrm{id},n}, v_1^\top d \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle P_d h_{\mathrm{id},n}, v_M^\top d \right\rangle_{\tilde{H}_d} \end{bmatrix} \\ &= \begin{bmatrix} \left\langle \delta_1^\top d, d_1 \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle \delta_1^\top d, d_M \right\rangle_{\tilde{H}_d} \\ \vdots & \ddots & \vdots \\ \left\langle \delta_n^\top d, d_1 \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle \delta_n^\top d, d_M \right\rangle_{\tilde{H}_d} \end{bmatrix} V = \delta^\top G_d V, \end{aligned}$$

where $\delta := \begin{bmatrix} \delta_1, & \ldots, & \delta_n \end{bmatrix}$. Using the reproducing property of the reproducing kernel of $\tilde{H}_d$, the coefficients δ_i in the projection of $h_{\mathrm{id},i}$ onto d satisfy

$$G_d \delta_i = \begin{bmatrix} \langle (h_{\mathrm{id}})_i, d_1 \rangle_{\tilde{H}_d} \\ \vdots \\ \langle (h_{\mathrm{id}})_i, d_M \rangle_{\tilde{H}_d} \end{bmatrix} = \begin{bmatrix} (\gamma_{u_1}(T_1))_i - (\gamma_{u_1}(0))_i \\ \vdots \\ (\gamma_{u_M}(T_M))_i - (\gamma_{u_M}(0))_i \end{bmatrix}.$$

Letting $D := \left(\left(\gamma_{u_j}(T_j) \right)_i - \left(\gamma_{u_j}(0) \right)_i \right)_{i,j=1}^{n,M}$ it can be concluded that $\delta^\top G_d = D$. Finally, the

modes ξ are given by $\xi = DV$ and the estimated open-loop model is given by

$$\dot{x} \approx \hat{F}_M(x,u) = DV\Sigma W^\top \beta(x) \begin{pmatrix} 1 \\ u \end{pmatrix} = DG_\beta^+ \beta(x) \begin{pmatrix} 1 \\ u \end{pmatrix}. \tag{5.6.2}$$

the approximation $\hat{f}_M(x)$, an approximation of the drift dynamics, $f(x)$, is given by the first column of $DG_\beta^+\beta(x)$ and $\hat{g}_M(x)$, an approximation of the control-effectiveness matrix, $g(x)$, is given by the last m columns of $DG_\beta^+\beta(x)$.

Since $P_\beta A_{f,g} P_d$ converges to $A_{f,g}$ in norm as $M \to \infty$, and since the space $F^2_{\tilde{\rho}_d}(\mathbb{R}^n)$ contains $h_{\mathrm{id},j}$ for $j = 1, \cdots, n$, the following result is immediate.

Corollary 5.6.1 *Under the hypothesis of Theorem 5.3.1, for all $u \in \mathcal{U}$, where for all $t \in [0,T]$, $\| \begin{pmatrix} 1 & u^\top(t) \end{pmatrix} \| < \overline{U}$, $\lim_{M\to\infty} \left(\sup_{x\in X} \left\| \hat{F}_M(x,u) - F(x,u) \right\|_2 \right) = 0$.*

Proof. Since the space $F^2_{\tilde{\rho}_d}(\mathbb{R}^n)$ contains $h_{\mathrm{id},j}$ for $j = 1, \cdots, n$, the functions $P_{\beta^j} A_{f,g} P_{d^j} h_{\mathrm{id},j}$ and $A_{f,g} h_{\mathrm{id},j}$ exist as members of H. Since $x \mapsto K(x,x) = \mathrm{diag}\left(\exp\left(\frac{x^\top x}{\tilde{\rho}_1}\right) \quad \dots \quad \exp\left(\frac{x^\top x}{\tilde{\rho}_{m+1}}\right) \right)$ is continuous and X is compact, there exists a real number $\overline{K}$ such that $\sup_{x\in X} \|K(x,x)\|_{\mathcal{Y}}^{\mathcal{Y}} = \overline{K}$.

Let $Y(x) = \begin{pmatrix} f(x) & g(x) \end{pmatrix} \in \mathbb{R}^{n\times(m+1)}$, and $\hat{Y}^M(x) = \begin{pmatrix} \hat{f}^M(x) & \hat{g}^M(x) \end{pmatrix} = DV\Sigma W^\top \beta(x) \in \mathbb{R}^{n\times(m+1)}$ is the identified system using the finite-rank representation with rank at-most M, and $Y_j(x)$ is the j^{th} row of $Y(x)$.

Theorem 5.3.1 can then be used to conclude that for all $\epsilon > 0$, $j = 1, \ldots, n$, there exists $M(j) \in \mathbb{N}$ such that for all $i \geq M(j)$, $\left\| \hat{Y}_j^i - Y_j \right\|_H^2 \leq \frac{\epsilon^2}{\overline{U}^2 \overline{K}}$. Using the reproducing

property,for$i \geq \overline{M} \coloneqq \max_j M(j)$,

$$\left((F(x,u))_j - (\hat{F}_M(x,u))_j\right)^2 = \left\langle (\hat{Y}_j^i(x) - Y_j(x)), \begin{pmatrix} 1 & u^\top \end{pmatrix} \right\rangle_{\mathcal{Y}}^2 =$$

$$\left\langle (\hat{Y}_j^i - Y_j), k_{x,\begin{pmatrix} 1 & u^\top \end{pmatrix}} \right\rangle_H^2 \leq \left\|\hat{Y}_j^i - Y_j\right\|_H^2 \left\| k_{x,\begin{pmatrix} 1 & u^\top \end{pmatrix}} \right\|_H^2 =$$

$$\left\|\hat{Y}_j^i - Y_j\right\|_H^2 \left\langle k_{x,\begin{pmatrix} 1 & u^\top \end{pmatrix}}, k_{x,\begin{pmatrix} 1 & u^\top \end{pmatrix}} \right\rangle_H = \left\|\hat{Y}_j^i - Y_j\right\|_H^2 \left\langle k_x \begin{pmatrix} 1 & u^\top \end{pmatrix}, k_x \begin{pmatrix} 1 & u^\top \end{pmatrix} \right\rangle_H =$$

$$\left\|\hat{Y}_j^i - Y_j\right\|_H^2 \left\langle \begin{pmatrix} 1 & u^\top \end{pmatrix}, k_x^* k_x \begin{pmatrix} 1 & u^\top \end{pmatrix} \right\rangle_{\mathcal{Y}} = \left\|\hat{Y}_j^i - Y_j\right\|_H^2 \left\langle \begin{pmatrix} 1 & u^\top \end{pmatrix}, K(x,x) \begin{pmatrix} 1 & u^\top \end{pmatrix} \right\rangle_{\mathcal{Y}} =$$

$$\left\|\hat{Y}_j^i - Y_j\right\|_H^2 \begin{pmatrix} 1 & u^\top \end{pmatrix} K(x,x) \begin{pmatrix} 1 \\ u \end{pmatrix} \leq \left\|\hat{Y}_j^i - Y_j\right\|_H^2 \overline{K} \left\| \begin{pmatrix} 1 \\ u \end{pmatrix} \right\|^2 \leq \frac{\epsilon^2}{\overline{U}^2 \overline{K}} \overline{U}^2 \overline{K} = \epsilon^2$$

As a result, for all $\epsilon \geq 0$ there exists $\overline{M}$ such that for all $i \geq \overline{M}$,

$$\sup_{x \in X} \left\| \hat{F}_M(x,u) - F(x,u) \right\|_2 = \sup_{x \in X} \left\| (\hat{Y}^M(x) - Y(x)) \begin{pmatrix} 1 \\ u \end{pmatrix} \right\|_2 \leq \epsilon,$$

which completes the proof.

■

The SCLDMD technique is summarized in Algorithm 2. The characterization $\Gamma_{\gamma_{u_j}} = \int_0^{T_j} \tilde{K}\left(\cdot, \gamma_{u_j}(t)\right) \mathrm{d}t$ of occupation kernels is used on line 11. The inner product of two control occupation kernels is given in [68] as

$$\left\langle \Gamma_{\gamma_{u_i}, u_i}, \Gamma_{\gamma_{u_j}, u_j} \right\rangle_H = \int_0^{T_j} \int_0^{T_i} \begin{bmatrix} 1 & u_i^\top(\tau) \end{bmatrix} K\left(\gamma_{u_j}(t), \gamma_{u_i}(\tau)\right) \begin{bmatrix} 1 \\ u_j(t) \end{bmatrix} \mathrm{d}\tau \mathrm{d}t. \tag{5.6.3}$$

5.7 Numerical Experiments

The purpose of the numerical experiment is to demonstrate the efficacy of the SCLDMD algorithm.

Algorithm 2 The SCLDMD algorithm

Input: Trajectories $\{\gamma_{u_i}\}_{i=1}^M$, a control signal u, a numerical integration procedure, and reproducing kernels $\tilde{K}_d$ and K of $\tilde{H}_d$ and H, respectively.

Output: $\{\xi_j, \sigma_j, \varphi_j, \phi_j\}_{j=1}^M$

$G_\beta \leftarrow \left(\left\langle \Gamma_{\gamma_{u_i},u_i}, \Gamma_{\gamma_{u_j},u_j} \right\rangle_H\right)_{i,j=1}^M$ (see (5.6.3))

$D \leftarrow \left(\left(\gamma_{u_j}(T_j)\right)_i - \left(\gamma_{u_j}(0)\right)_i\right)_{i,j=1}^{n,M}$

$(W, \Sigma, V) \leftarrow$ SVD of G_β^+

$\xi \leftarrow DV$

$\phi_j \leftarrow \sum_{i=1}^M (V)_{i,j} \left(K_d(\cdot, \gamma_{u_i}(T_i)) - K_d(\cdot, \gamma_{u_i}(0))\right)$

$\psi_j \leftarrow \sum_{i=1}^M \int_0^{T_i} (W)_{i,j} \left[\begin{bmatrix} 1 & u_i(t)^\top \end{bmatrix} K_{\gamma_{u_i}(t)}\right](\cdot)\,\mathrm{d}t$

Return $\{\xi_j, \sigma_j, \varphi_j, \phi_j\}_{j=1}^M$

This experiment utilizes the nonlinear model of the Duffing oscillator given by

$$\dot{x} = f(x) + g(x)u = \begin{pmatrix} x_2 \\ x_1 - x_1^3 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 + \sin(x_1) \end{pmatrix} u, \tag{5.7.1}$$

where $f(x) = \begin{pmatrix} x_2 \\ x_1 - x_1^3 \end{pmatrix}$ is the drift function, $g(x) = \begin{pmatrix} 0 \\ 2 + \sin(x_1) \end{pmatrix}$ is the control effectiveness function, and u is the controller. To approximate the system dynamics, 225 trajectories of the system are recorded, along with the corresponding control signals, starting from a grid of initial conditions, under a control signal that is composed of the sum of 15 sinusoidal signals with randomly generated frequencies between 1 and 3 radians per second, randomly generated phases between -1 and 1 radians, and randomly generated coefficients between -1 and 1. The recorded trajectories and control signals, which are stored with a time step of 0.05 second and a duration of 1 second, are then utilized to construct the estimates $\hat{f}_M$ and $\hat{g}_M$ of f and g, respectively. The grid of initial conditions is composed of 15 points in the x_1 coordinate and 15 points in the x_2 coordinate, equally spaced in the interval $[-3, 3]$ in each coordinate. The exponential dot product kernel $\tilde{K}_d(x, y) = \exp(\frac{x^\top y}{\mu})$ is used to define $\tilde{H}_d$

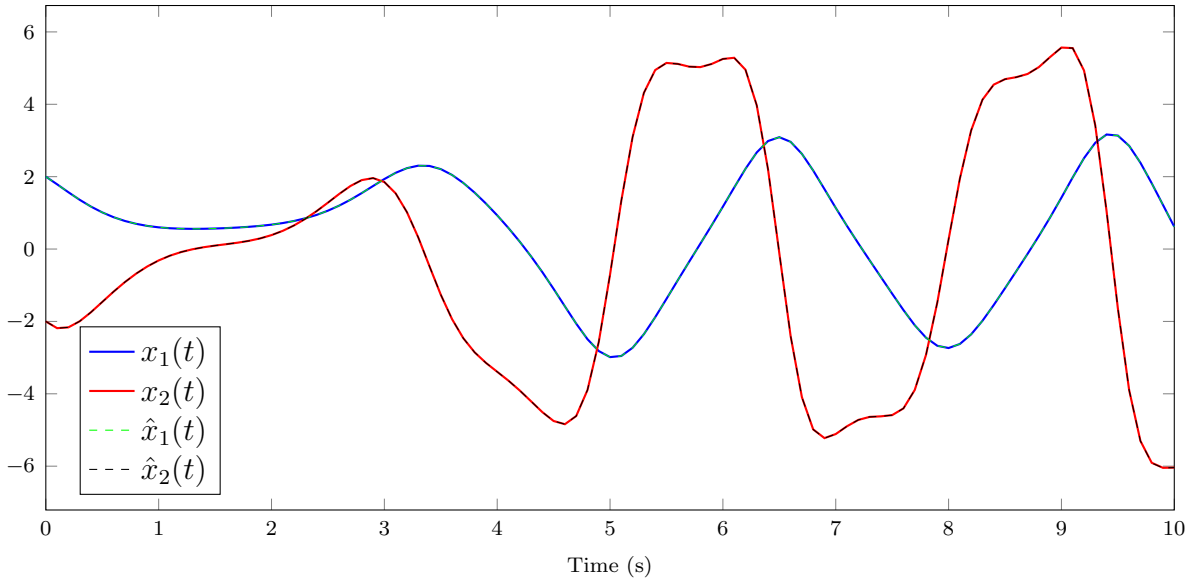

Figure 15: The responses of the true Duffing oscillator and the identified duffing oscillator to the input $u(t) = \sin(t) + \cos(2t)$.

with $\mu = 11$, and the kernel operator K is selected to be a diagonal matrix of exponential dot product kernels with the parameter $\mu_v = 10$. Simpson's 1/3 rule is used for numerical integration.

Figure 15 shows the responses of the true Duffing oscillator from (5.7.1) and of the Duffing oscillator identified using SCLDMD, starting from the initial condition of $[2, -2]$ under the input $u(t) = \sin(t) + \cos(2t)$ for 10 seconds. The states of the identified model, $\hat{x}_1$ and $\hat{x}_2$,

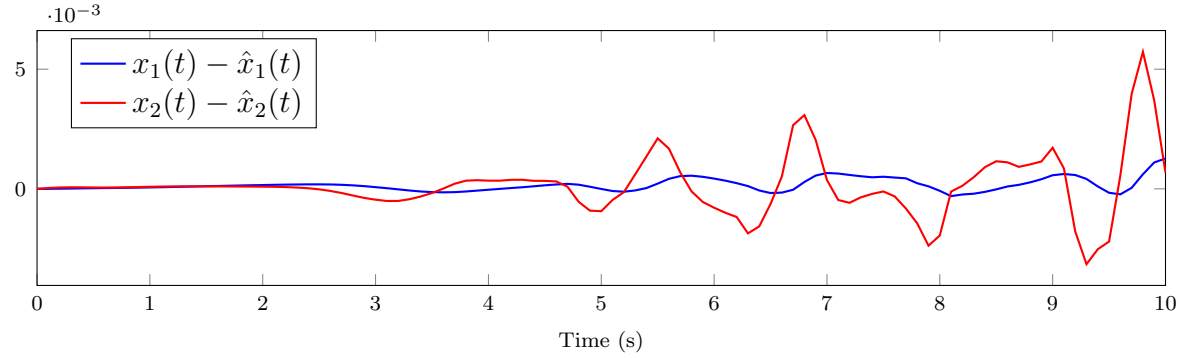

Figure 16: The difference in responses of the true Duffing oscillator and the identified duffing oscillator to the input $u(t) = \sin(t) + \cos(2t)$.

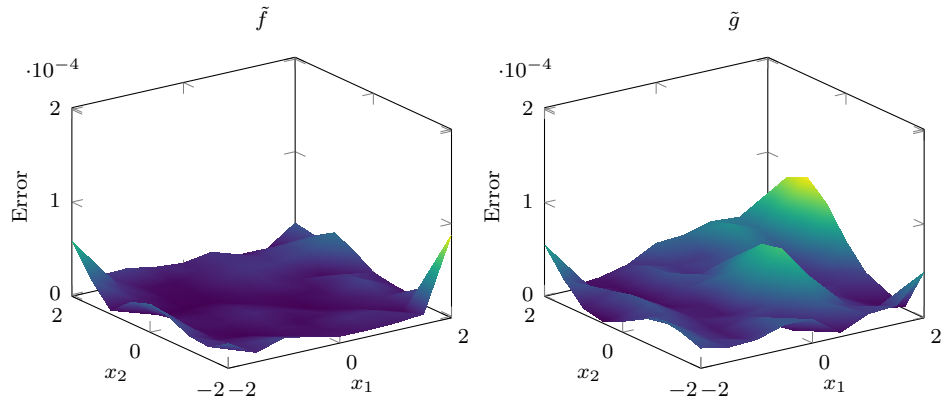

Figure 17: Error $|f(x) - \hat{f}(x)|$ as a function of x (left) and error $|g(x) - \hat{g}(x)|$ as a function of x (right).

track closely the states x_1 and x_2 of (5.7.1). Figure 16 shows the difference between the trajectories of the actual system and the identified model, which is in the order of 10^{-3}. The maximum error percentage for $\hat{x}_2$, which is the higher of the two, is 0.001%.

Figure 17 shows the vector field estimation errors for both the drift dynamics, $\tilde{f}$, and the control effectiveness, $\tilde{g}$, as functions of x_1 and x_2 in the domain $(x_1, x_2) \in [-2, 2] \times [-2, 2]$. Both errors, $\tilde{f}$ and $\tilde{g}$ are of the order 10^{-4} in the above mentioned domain, where the error increases near the edges of the domain. Although g from (5.7.1) is only a function of x_1, in Figure 17 it can be seen that $\hat{g}$ changes with x_2. This is expected since the estimate $\hat{g}$ obtained through SCLDMD is a function of both states, given that the SCLDMD algorithm is not informed of the state dependencies of the control effectiveness. Exploring the incorporation of partial knowledge of the system to improve SCLDMD is part of future work.

5.8 Conclusion

This chapter introduces a novel approach towards the construction of a finite-rank representation of the control Liouville operator. New results on the construction of singular values and functions of the finite rank operator using singular values and vectors of a matrix representation are also obtained. Once the singular values and functions are at hand, the drift

dynamics and the control effectiveness can be approximated, which facilitates the prediction of the states of the dynamical system in response to an open-loop control signal. Moreover, the finite-rank representation of the control Liouville operator is shown to be convergent (in norm) to the true control Liouville operator, provided sufficiently rich data are available.

A numerical experiment using the Duffing oscillator is presented to demonstrate the performance of the developed technique. Analysis of the effects of integration error and measurement noise on the SCLDMD technique, as well as incorporating partial knowledge of the system to improve the identification technique are part of future work.

CHAPTER VI

ON CONVERGENT DYNAMIC MODE DECOMPOSITION AND ITS EQUIVALENCE WITH OCCUPATION KERNEL REGRESSION

6.1 Introduction

This chapter presents a new technique for norm-convergent dynamic mode decomposition of deterministic systems. The developed method utilizes recent results on singular dynamic mode decomposition where it is shown that by appropriate selection of domain and range Hilbert spaces, the Liouville operator (also known as the Koopman generator) can be made to be compact. In this chapter, it is shown that by selecting appropriate collections of finite basis functions in the domain and the range, a novel finite-rank representation of the Liouville operator may be obtained. It is also shown that the model resulting from dynamic mode decomposition of the finite-rank representation is closely related to regularized regression using the so-called occupation kernels as basis functions.

6.2 Reproducing Kernel Hilbert Spaces and Dynamic Mode Decomposition

The objective is to find functions for which

$$|A_f\phi(x) - \lambda\phi(x)| < \epsilon \tag{6.2.1}$$

for some λ and some small positive ϵ and all x within a domain of interest. Since norm convergence in a RKHS of continuous functions yields uniform convergence over compact sets [65,76], it is sufficient to satisfy $\|A_f\phi - \lambda\phi\|_H < \epsilon$. In turn, if a finite-rank approximation of A_f, call it $\tilde{A}_f$, is close enough, it is sufficient to satisfy $\|A_f - \tilde{A}_f\| < \epsilon$, and the rest follows

as

$$|A_f\phi(x) - \lambda\phi(x)| < C\|A_f\phi - \lambda\phi\|_H$$
$$C\|A_f\phi - \tilde{A}_f\phi\|_H < C\|A_f - \tilde{A}_f\|_H < C\epsilon,$$

where C is a positive constant that depends on the domain of interest and the kernel function, and the function ϕ is assumed to be normalized.

A convergent approximation of the spectrum of the transfer operator using the spectrum of finite-rank operators requires compactness and convergence in norm, which motivates the following section.

6.3 Compact Liouville Operators and Occupation Kernels

This section, included here for completeness, recalls the results from [34] pertaining to the existence of compact Liouville operators, given formally as $A_f g(x) = \nabla g(x) f(x)$. Compactness is achieved through the consideration of differing spaces for the domain and the range of the operator. We emphasize that compact Liouville operators are not restricted to these particular pairs of functions spaces. This section only provides examples demonstrating the existence of such operators, thereby validating the approach in the sequel.

6.3.1 Liouville Operators on Real Bergmann-Fock Spaces

The exponential dot product kernel, with parameter $\mu > 0$, is given as $K(x, y) = \exp\left(\frac{x^\top y}{\mu}\right)$. In the single variable case, the native space for this kernel is the restriction of the Bergmann-Fock space to real numbers, denoted by $F^2_\mu(\mathbb{R})$. This space consists of the set of polynomials of the form $h(x) = \sum_{k=0}^\infty a_k x^k$, where the coefficients satisfy $\sum_{k=0}^\infty |a_k|^2 \mu^k k! < \infty$, and the norm is given by $\|h\|^2_\mu = \sum_{k=0}^\infty |a_k|^2 \mu^k k!$. Extension of this definition to the multivariable case yields the space $F^2_\mu(\mathbb{R}^n)$ where the collection of monomials, $x^\alpha \frac{\mu^{|\alpha|}}{\sqrt{\alpha!}}$, with multi-indices $\alpha \in \mathbb{N}^n$ forms an orthonormal basis, where for $\alpha \in \mathbb{N}^n$, $\alpha! = \prod_{i=1}^n \alpha_i!$, $|\alpha| = \sum_{i=1}^n \alpha_i$, and $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$. In this setting, differential operators from $F^2_{\mu_1}(\mathbb{R}^n)$ to $F^2_{\mu_2}(\mathbb{R}^n)$ can be shown

to be compact provided $\mu_2 < \mu_1$.

Proposition 6.3.1 ([34]) *If $\mu_2 < \mu_1$, then the differential operators $\frac{\partial}{\partial x_i} : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_2}(\mathbb{R}^n)$, are compact for $i = 1, \ldots, n$.*

Multiplication operators can be shown to be bounded provided that their symbols are polynomial.

Proposition 6.3.2 ([34]) *If $\mu_2 < \mu_1$, then for any polynomial function $p : \mathbb{R}^n \to \mathbb{R}$, the multiplication operator $M_p : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_2}(\mathbb{R}^n)$, defined as $[M_p h](x) = p(x)h(x)$, is bounded.*

Boundedness of Liouville operators with polynomial symbols then follows trivially from the above two results.

Theorem 6.3.1 ([34]) *If $\mu_2 < \mu_1$ and if $f : \mathbb{R}^n \to \mathbb{R}^n$ is a component-wise polynomial function, then the Liouville operator $A_f : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_2}(\mathbb{R}^n)$ defined as $A_f g = \nabla g \cdot f$ is a compact operator.*

6.3.2 Occupation Kernels and Liouville Operators

Let H be an RKHS over a compact set $X \in \mathbb{R}^n$ consisting of continuous functions and let K be the reproducing kernel of H. Given a continuous signal $\gamma : [0, T] \to X$, the linear functional $g \mapsto \int_0^T g(\gamma(t))\mathrm{d}t$ is bounded. Hence, by the Riesz representation theorem [77, Theorem 13.32], there exists a function $\Gamma_\gamma \in H$ such that $\langle g, \Gamma_\gamma \rangle_H = \int_0^T g(\gamma(t))\mathrm{d}t$ for all $g \in H$. The function Γ_γ is called the occupation kernel corresponding to γ in H. These occupation kernels were first introduced in [40, 67].

Occupation kernels corresponding to trajectories of a dynamical system have a useful relationship with adjoints of Liouville operators and the reproducing kernels of the underlying RKHSs.

Proposition 6.3.3 ([33]) *If $\gamma : [0, T] \to X$ is a solution of $\dot{x} = f(x)$ for some locally Lipschitz continuous function $f : X \to \mathbb{R}^n$, then $A^*_f \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$.*

Proposition 6.3.4 ([78]) *Given any continuous function $\gamma : [0, T] \to X$, the occupation kernel corresponding to γ in H may be expressed as $\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t))\mathrm{d}t$. As a consequence, given two functions $\gamma_i : [0, T_i] \to X$ and $\gamma_j : [0, T_j] \to X$*

$$\left\langle \Gamma_{\gamma_j}, \Gamma_{\gamma_i} \right\rangle_H = \int_0^{T_i} \int_0^{T_j} K\left(\gamma_i(\tau), \gamma_j(t)\right) \mathrm{d}t\mathrm{d}\tau. \tag{6.3.1}$$

6.4 Finite-rank Representation of the Liouville Operator

This section provides a novel finite-rank representation of the Liouville operator A_f and subsequently, a novel approach to obtain the dynamic modes of the underlying dynamical system. Let H_d and H_r be RKHSs with reproducing kernels K_d and K_r, respectively. The RKHS H_d is used as the domain of the Liouville operator and the RKHS H_r is used as the range. In what follows, finite collections of vectors $d^M \subset H_d$ and $r^M \subset H_r$ are selected to establish the needed finite-rank representation of the Liouville operator.

Since the adjoint of A_f maps occupation kernels to kernel differences (Proposition 6.3.3), the span of the collection of kernel differences

$$d^M = \{K_d(\cdot, \gamma_i(T_i)) - K_d(\cdot, \gamma_i(0))\}_{i=1}^M \subset H_d \tag{6.4.1}$$

are selected to be the domain of A_f. The corresponding Gram matrix is denoted by $G_{d^M} = \left(\langle d_i, d_j \rangle_{H_d}\right)_{i,j=1}^M$. The result of A_f operating on functions in $\operatorname{span} d^M$ is projected onto the span of the collection of occupation kernels

$$\operatorname{span} r^M = \operatorname{span} \{\Gamma_{\gamma_i}\}_{i=1}^M \subset H_r. \tag{6.4.2}$$

The corresponding Gram matrix is denoted by $G_{r^M} = \left(\langle r_i, r_j \rangle_{H_r}\right)_{i,j=1}^M$. A rank-$M$ representation of the operator A_f is then given by $P_{r^M} A_f P_{d^M} : H_d \to \operatorname{span} r^M$, where P_{r^M} and P_{d^M} denote projection operators onto $\operatorname{span} r^M$ and $\operatorname{span} d^M$, respectively. Under the compactness assumptions and given rich enough data so that the spans of $\{d_i\}_{i=1}^\infty$ and $\{r_i\}_{i=1}^\infty$ are dense in H_d and H_r, respectively, the sequence of finite-rank operators $\{P_{r^M} A_f P_{d^M}\}_{M=1}^\infty$ can be shown to converge, in the norm topology, to A_f. To facilitate the proof of convergence, we recall the following result from [33].

Proposition 6.4.1 ([33]) *Let H_d and H_r be RKHSs defined on $X \subset \mathbb{R}^n$ and let $A_N : H_d \to H_r$ be a finite-rank operator with rank N. If $\operatorname{span}\{d_i\}_{i=1}^{\infty} \subset H_d$ is dense in H_d and $\operatorname{span}\{r_i\}_{i=1}^{\infty} \subset H_r$ is dense in H_r, then for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$ and $h \in H_d$, $\|A_N h - A_N P_{d^i} h\|_{H_r} \leq \epsilon \|h\|_{H_d}$ and $\|A_N h - P_{r^i} A_N h\|_{H_r} \leq \epsilon \|h\|_{H_d}$.*

The convergence result for Liouville operators on Bergmann-Fock spaces restricted to the set of real numbers follows from the following more general result.

Proposition 6.4.2 *If $A : H_d \to H_r$ is a compact operator, and spans of the collections $\{d_i\}_{i=1}^{\infty}$ and $\{r_i\}_{i=1}^{\infty}$ are dense in H_d and H_r, respectively, then $\lim_{M\to\infty} \|A - P_{r^M} A P_{d^M}\|_{H_d}^{H_r} = 0$, where $\|\cdot\|_{H_d}^{H_r}$ denotes the operator norm of operators from H_d to H_r.*

Proof. Let $\{A_N\}_{N=1}^{\infty}$ be a sequence of rank-N operators converging, in norm, to A. For an arbitrary $h \in H_d$,

$$\begin{aligned}\|Ah - P_{r^M} A P_{d^M} h\|_{H_r} \leq \|Ah - A_N h\|_{H_r} \\ + \|A_N h - A_N P_{d^M} h\|_{H_r} + \|A_N P_{d^M} h - P_{r^M} A_N P_{d^M} h\|_{H_r} \\ + \|P_{r^M} A_N P_{d^M} h - P_{r^M} A P_{d^M} h\|_{H_r}.\end{aligned}$$

Using the fact that A_N, $P_{r^M} A_N$, and $A_N P_{d^M}$ are all finite-rank operators and the fact that the projection operator is bounded with norm bound 1, Proposition 6.4.1, can be used to conclude that for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$

$$\|Ah - P_{r^i} A P_{d^i} h\|_{H_r} \quad \leq \quad \|(A - A_N)h\|_{H_r} \quad + \quad 2\epsilon \|h\|_{H_d} \quad + \quad \|(A_N - A) P_{d^i} h\|_{H_r}.$$

Since A_N converges to A in norm, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $j \geq N$, and $g \in H_d$ $\|Ag - A_j g\|_H \leq \epsilon \|g\|_{H_d}$. Thus, for all $j \geq N$ and $i \geq M(j)$, $\|Ah - P_{r^i} A P_{d^i} h\|_{H_r} \leq 4\epsilon \|h\|_{H_d}$. ∎

6.4.1 Matrix Representation of the Finite-rank Operator

To formulate a matrix representation of the finite-rank operator $P_{r^M} A_f P_{d^M}$, the operator is restricted to $\operatorname{span} d^M$ to yield the finite-rank operator $P_{r^M} A_f|_{d^M} : \operatorname{span} d^M \to \operatorname{span} r^M$. The

resulting matrix representation is denoted by $[A_f]_d^r$. For brevity of exposition, the superscript M is suppressed hereafter and d and r are interpreted as $M-$dimensional vectors.

Theorem 6.4.1 *Let $g = a^\top r \in \operatorname{span} r$ and $h = \delta^\top d \in \operatorname{span} d$ be functions with coefficients $a \in \mathbb{R}^M$ and $\delta \in \mathbb{R}^M$, respectively. If $g = P_r A_f|_d h$, then $a = G_r^+ G_d \delta$. That is, $[A_f]_d^r \coloneqq G_r^+ G_d$, where G_r^+ denotes the Moore-Penrose pseudoinverse of G_r, is a matrix representation of $P_r A_f|_d$.*

Proof. Since $g = P_r A_f|_d h = a^\top r$, the coefficients a solve

$$G_r a = \begin{bmatrix} \left\langle A_f \delta^\top d, r_1 \right\rangle_{H_r} \\ \vdots \\ \left\langle A_f \delta^\top d, r_M \right\rangle_{H_r} \end{bmatrix} = \begin{bmatrix} \left\langle \delta^\top d, A_f^* r_1 \right\rangle_{H_d} \\ \vdots \\ \left\langle \delta^\top d, A_f^* r_M \right\rangle_{H_d} \end{bmatrix}$$

Proposition 6.3.3 implies that $A_f^* r_i = d_i$ for all $i = 1, \cdots, M$, and as such, $G_r a = G_d \delta$. Since $a = G_r^+ G_d \delta$ is a solution of $G_r a = G_d \delta$, we have, $a = G_r^+ G_d \delta$, and as a result, $G_r^+ G_d$ is a matrix representation of $P_r A_f|_d$. ■

In the following section, the matrix representation $[A_f]_d^r$ is used to construct a data-driven representation of the singular values and the left and right singular functions of the operator $P_r A_f|_d$.

6.4.2 Singular Functions of the Finite-rank Operator

The tuples $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^M$, with $\sigma_i \in \mathbb{R}$, $\phi_i \in H_d$, and $\psi_i \in H_r$, are singular values, left singular vectors, and right singular vectors of $P_r A_f|_d$, respectively, if $\forall h \in \operatorname{span} d$, $P_r A_f h = \sum_{i=1}^M \sigma_i \psi_i \langle h, \phi_i \rangle_{H_d}$. The following proposition states that the SVD of $P_r A_f|_d$ can be computed using matrices in the matrix representation $[A_f]_d^r$ developed in the previous section.

Theorem 6.4.2 *If (W, Σ, V) is the SVD of G_r^+ with $W = \begin{bmatrix} w_1, & \ldots, & w_M \end{bmatrix}$, $V = \begin{bmatrix} v_1, & \ldots, & v_M \end{bmatrix}$, and $\Sigma = \operatorname{diag}\left(\begin{bmatrix} \sigma_1, & \ldots, & \sigma_M \end{bmatrix}\right)$, then for all $i = 1, \ldots, M$, σ_i are singular values of of*

$P_r A_f|_d$ with left singular functions $\phi_i := v_i^\top d$ and right singular functions $\psi_i := w_i^\top r$, respectively.

Proof. Let $\phi_i = v_i^\top d$ and $\psi_i = w_i^\top r$ and $h = \delta^\top d$. Then,

$$P_r A_f h = \sum_{i=1}^{M} \sigma_i \psi_i \langle h, \phi_i \rangle_{H_d} \iff P_r A_f \delta^\top d = \sum_{i=1}^{M} \sigma_i w_i^\top r \left\langle \delta^\top d, v_i^\top d \right\rangle_{H_d}$$

Using the finite-rank representation, the collection $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^{M}$, is a SVD of $P_r A_f|_d$, if for all $\delta \in \mathbb{R}^M$,

$$\left(G_r^+ G_d \delta\right)^\top r = \left(\sum_{i=1}^{M} \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{H_d} w_i^\top \right) r. \tag{6.4.3}$$

Simple matrix manipulations yield the chain of implications

$$\begin{aligned}
(7.5.1) \impliedby \ & \forall \delta \in \mathbb{R}^M, G_r^+ G_d \delta = \sum_{i=1}^{M} \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{H_d} w_i \\
\iff \ & \forall \delta \in \mathbb{R}^M, G_r^+ G_d \delta = \sum_{i=1}^{M} \sigma_i \left(w_i v_i^\top G_d \right) \delta \\
\impliedby \ & G_r^+ G_d = \sum_{i=1}^{M} \sigma_i \left(w_i v_i^\top \right) G_d \impliedby G_r^+ = \sum_{i=1}^{M} \sigma_i w_i v_i^\top = W \Sigma V^\top,
\end{aligned}$$

which proves the proposition. ∎

In the following section, the singular values and the left and right singular vectors are used to generate a data-driven model via a method termed singular Liouville dynamic mode decomposition (SLDMD).

6.5 The SLDMD Algorithm

Let the identity observable $(h_{\mathrm{id}})_j$ be defined as $(h_{\mathrm{id}})_j(x) = x_j$, where x_j is the $j-$th component of x.

Theorem 6.5.1 *If $(h_{\mathrm{id}})_j \in H_d$ for $j = 1, \cdots, n$, then*

$$\hat{f}_M(x) := \begin{bmatrix} [P_r A_f P_d (h_{\mathrm{id}})_1](x) \\ \vdots \\ [P_r A_f P_d (h_{\mathrm{id}})_n](x) \end{bmatrix} = D G_r^+ r(x) \tag{6.5.1}$$

where $D \coloneqq \left((\gamma_j(T_j))_i - (\gamma_j(0))_i\right)_{i,j=1}^{n,M}$ *and* $(\gamma_j(\cdot))_i$ *denotes the* $i-$*th component of* $\gamma_j(\cdot)$.

Proof. Using Theorem 6.4.2 and the definition of singular values and singular functions of $P_r A_f|_d$,

$$\hat{f}_M(x) = \sum_{i=1}^{M} \sigma_i \xi_i w_i^\top r(x) = \xi \Sigma W^\top r(x), \tag{6.5.2}$$

where $\xi_i \coloneqq \begin{bmatrix} \langle P_d(h_{\mathrm{id}})_1, \phi_i \rangle_{H_d} & \dots, \langle P_d(h_{\mathrm{id}})_n, \phi_i \rangle_{H_d} \end{bmatrix}^\top$ and $\xi \coloneqq \begin{bmatrix} \xi_1 & \dots & \xi_M \end{bmatrix}$.

The modes ξ can be computed using $\phi_i = v_i^\top d$ as

$$\begin{aligned} \xi &= \begin{bmatrix} \left\langle P_d(h_{\mathrm{id}})_1, v_1^\top d \right\rangle_{H_d}, & \dots, & \left\langle P_d(h_{\mathrm{id}})_1, v_M^\top d \right\rangle_{H_d} \\ \vdots & \ddots & \vdots \\ \left\langle P_d(h_{\mathrm{id}})_n, v_1^\top d \right\rangle_{H_d}, & \dots, & \left\langle P_d(h_{\mathrm{id}})_n, v_M^\top d \right\rangle_{H_d} \end{bmatrix} \\ &= \begin{bmatrix} \left\langle \delta_1^\top d, d_1 \right\rangle_{H_d}, & \dots, & \left\langle \delta_1^\top d, d_M \right\rangle_{H_d} \\ \vdots & \ddots & \vdots \\ \left\langle \delta_n^\top d, d_1 \right\rangle_{H_d}, & \dots, & \left\langle \delta_n^\top d, d_M \right\rangle_{H_d} \end{bmatrix} V = \delta^\top G_d V, \end{aligned}$$

where $\delta \coloneqq \begin{bmatrix} \delta_1, & \dots, & \delta_n \end{bmatrix}$. Using the reproducing property of the reproducing kernel of H_d, the coefficients δ_i in the projection of $(h_{\mathrm{id}})_i$ onto d solve the system of linear equations

$$G_d \delta_i = \begin{bmatrix} \langle (h_{\mathrm{id}})_i, d_1 \rangle_{H_d} \\ \vdots \\ \langle (h_{\mathrm{id}})_i, d_M \rangle_{H_d} \end{bmatrix} = \begin{bmatrix} (\gamma_1(T_1))_i - (\gamma_1(0))_i \\ \vdots \\ (\gamma_M(T_M))_i - (\gamma_M(0))_i \end{bmatrix}.$$

Letting $D \coloneqq \left((\gamma_j(T_j))_i - (\gamma_j(0))_i\right)_{i,j=1}^{n,M}$ it can be concluded that $\delta^\top G_d = D$. Finally, the modes ξ are given by $\xi = DV$ and

$$\hat{f}_M(x) = DV\Sigma W^\top r(x) = DG_r^+ r(x).$$

■

Using $\hat{f}_M$ as a rank$-M$ estimate of f, the estimated system model is of the form $\dot{x} \approx \hat{f}_M(x) = Ar(x)$, where $A \in \mathbb{R}^{n \times M}$ is a solution of

$$AG_r = D. \tag{6.5.3}$$

The use of $\hat{f}_M$ as an estimate of f is justified by the following result, which follows from the fact that $P_r A_f P_d$ converges to A_f in norm as $M \to \infty$.

Theorem 6.5.2 *If $\mu_r < \mu_d$, $H_d = F^2_{\mu_d}(\mathbb{R}^n)$, $H_r = F^2_{\mu_r}(\mathbb{R}^n)$, $f : \mathbb{R}^n \to \mathbb{R}^n$ is a componentwise polynomial function, and the spans of the collections $\{d_i\}_{i=1}^{\infty}$ and $\{r_i\}_{i=1}^{\infty}$ are dense in H_d and H_r, respectively, then $\lim_{M\to\infty}\left(\sup_{x\in X}\left\|\hat{f}_M(x) - f(x)\right\|_2\right) = 0$.*

Proof. Since the space $F^2_{\mu_d}(\mathbb{R}^n)$ contains $(h_{\mathrm{id}})_j$ for $j = 1, \cdots, n$, the functions $\hat{f}_{M,j} \coloneqq P_r A_f P_d (h_{\mathrm{id}})_j$ and $f_j \coloneqq A_f (h_{\mathrm{id}})_j$ that denote the $j-$th row of $\hat{f}_M$ and f, respectively, exist as members of H_r. Since $x \mapsto K_r(x,x) = \exp\left(\frac{x^\top x}{\mu_r}\right)$ is continuous and X is compact, there exists a real number $\overline{K}$ such that $\sup_{x\in X} K_r(x,x) = \overline{K}$. Since $\mu_r < \mu_d$, Theorem 6.3.1 implies that $A_f : F^2_{\mu_d}(\mathbb{R}^n) \to F^2_{\mu_r}(\mathbb{R}^n)$ is compact. Proposition 6.4.2 can then be used to conclude that for all $\epsilon > 0$ and $j = 1, \ldots, n$, there exists $M(j) \in \mathbb{N}$ such that for all $i \geq M(j)$, $\left\|\hat{f}_{i,j} - f_j\right\|^2_{H_r} \leq \frac{\epsilon^2}{n\overline{K}^2}$. Using the reproducing property, for $i \geq \overline{M} \coloneqq \max_j M(j)$,

$$\begin{aligned}
\left\|\hat{f}_i(x) - f(x)\right\|_2^2 &= \sum_{j=1}^{n} \left\langle \left(\hat{f}_{i,j} - f_j\right), K_r(\cdot, x)\right\rangle_{H_r}^2 \\
&\leq \sum_{j=1}^{n} \left\|\hat{f}_{i,j} - f_j\right\|^2_{H_r} \|K_r(\cdot,x)\|^2_{H_r} \\
&\leq \sum_{j=1}^{n} \frac{\epsilon^2}{n\overline{K}^2} \langle K_r(\cdot,x), K_r(\cdot,x)\rangle^2_{H_r} = \frac{\epsilon^2}{\overline{K}^2} K_r(x,x)^2.
\end{aligned}$$

As a result, for all $\epsilon \geq 0$ there exists $\overline{M}$ such that for all $i \geq \overline{M}$,

$$\sup_{x\in X}\left\|\hat{f}_i(x) - f(x)\right\|_2 \leq \sqrt{\frac{\epsilon^2}{\overline{K}^2}\sup_{x\in X} K_r(x,x)^2} = \epsilon,$$

which completes the proof. ∎

This convergence result, typically unattainable for Koopman-based dynamic mode decomposition, is a key feature of this algorithm.

6.6 Occupation kernel regression

Another approach to modeling is to directly use the occupation kernels as basis functions for regression.

If the components of the vector field f reside in the RKHS, i.e., $f = \begin{bmatrix} f_1 & \dots & f_n \end{bmatrix}^\top$, with $f_i \in H$ for $i = 1, \dots, n$, then using the defining characteristic of the occupation kernel, the inner product $\langle f_i, \Gamma_\gamma \rangle_H$ may be expressed as

$$\langle f_i, \Gamma_\gamma \rangle_H = \int_0^T f_i(\gamma(t)) \mathrm{d}t = (\gamma(T))_i - (\gamma(0))_i$$

for each $i = 1, \dots, n$. As such, given a set of solutions $\{\gamma_j : [0, T_j] \to \mathbb{R}^n\}_{j=1}^M$, of $\dot{x} = f(x)$, components of the function f can be estimated by minimizing the error between the inner products $\langle f_i, \Gamma_\gamma \rangle_H$ and the displacement $(\gamma(T))_i - (\gamma(0))_i$ of the $i-$th component of the trajectory.

A regularized regression problem to determine an approximation $\hat{f}_i$ of the i-th row of f can thus be formulated as

$$\min_{\hat{f}_i \in H} \sum_{j=1}^{M} \left(\langle \hat{f}_i, \Gamma_{\gamma_j} \rangle_H - \left(\gamma_j(T_j)\right)_i - \left(\gamma_j(0)\right)_i \right)^2 + \lambda \|\hat{f}_i\|_H^2, \tag{6.6.1}$$

where $\lambda > 0$ is a regularization parameter. Using the Representer theorem for occupation kernels [46, Proposition 1], the minimizer of (6.6.1) may be expressed as a linear combination of occupation kernels

$$\hat{f}_i = A_{1,i} \Gamma_{\gamma_1} + \cdots + A_{M,i} \Gamma_{\gamma_M} = A_i r(x),$$

where $A_i \coloneqq \begin{bmatrix} A_{1,i} & \dots & A_{M,i} \end{bmatrix}$. Using this representation of the minimizer, the inner products and the norm in the optimization problem can be computed as

$$\langle \hat{f}_i, \Gamma_{\gamma_j} \rangle_H = \left\langle \sum_{k=1}^{M} A_{k,i} \Gamma_{\gamma_k}, \Gamma_{\gamma_j} \right\rangle_H = (G_r^j)^\top A_i^\top$$

and

$$\|\hat{f}_i\|_H^2 = A_i G_r A_i^\top,$$

where G_r^j denotes the $j-$th column of G_r.

Hence, the resolution (6.6.1) reduces to the finite dimensional convex optimization problem

$$\min_{A_i \in \mathbb{R}^{1\times M}} \sum_{j=1}^{M} \left((G_r^j)^\top A_i^\top - (\gamma_j(T_j) - \gamma_j(0))_i \right)^2 + \lambda A_i G_r A_i^\top. \tag{6.6.2}$$

Solutions of the finite dimensional optimization problem coincide with solutions of the linear system $A_i(G_r + \lambda I_M)G_r = D_i G_r$, where D_i is the $i-$th row of D and I_M denotes the $M \times M$ identity matrix.

Using the fact that the solution of $A_i(G_r + \lambda I_M) = D_i$ is one of the solutions of $A_i(G_r + \lambda I_M)G_r = D_i G_r$ (the only one if G_r is nonsingular), the estimated model can be expressed as $\dot{x} \approx Ar(x)$, where $A \in \mathbb{R}^{n\times M}$ is the solution of

$$A(G_r + \lambda I_M) = D. \tag{6.6.3}$$

Interestingly, the model obtained via OKR coincides with that obtained using DMD if the regularization parameter is set to zero. As such, if occupation kernels are used for regression, the resulting model encodes additional structure, i.e., the singular functions and the singular values of the underlying Liouville operator, as provided by SLDMD. Furthermore, theorem 6.5.2, also applicable to the regression model without regularization, provides an alternative to the cumulative prediction error convergence guarantees (see, for example [79] and Chapter 6 of [65]) that are typically available for regression problems.

6.7 Numerical Experiments

The purpose of the numerical experiment is to demonstrate the effectiveness of the SLDMD algorithm, the OKR algorithm, and to compare the two methods. Additionally, the numerical experiment demonstrates the effects of the regularization constant λ on the OKR method.

This experiment utilizes the nonlinear model of the Duffing oscillator given by

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= (x_1 - x_1^3). \end{aligned} \tag{6.7.1}$$

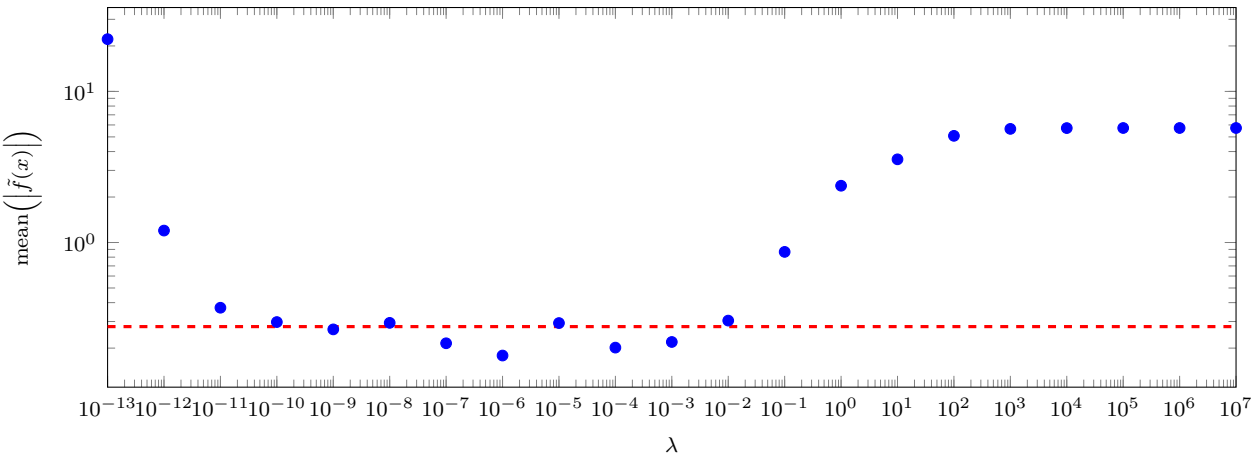

Figure 18: The blue marks are the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3, 3]$ for different values of λ using OKR trained with noisy trajectories. the dashed red line is the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3, 3]$ calculated using the SLDMD method trained with noisy trajectories.

To approximate the system dynamics for any given value of $\lambda > 0$, 169 trajectories of the system are recorded starting from a grid of initial conditions. In the first trial, each trajectory is corrupted by Gaussian measurement noise with standard deviation 0.001. The initial velocities are obtained by numerically differentiating the measured noisy trajectories. In the second trial, the trajectories are not corrupted by noise, and the initial velocities are similarly obtained by numerically differentiating the measured trajectories. The recorded trajectories are then utilized to approximate f. The SLDMD algorithm from Section 6.5 is implemented using the Moore-Penrose pseudoinverse. The OKR algorithm from Section 6.6 is implemented using the usual matrix inverse. Both methods use the kernel $K(x, y) = \exp\left(\frac{x^\top y}{\mu}\right)$ with $\mu = 5$.

Figures 18 and 19 show the mean of the difference between the actual values $f_2(x_1)$ and the estimated values $\hat{f}_2(x_1)$ over $x_2 \in [-3, 3]$ using different values of λ. Figure 18 shows the results for the system identification methods using noisy trajectories, where OKR clearly outperforms SLDMD for specific values of λ. Whereas, in Figure 19, which shows the results of the two system identification methods using noise free trajectories, the OKR technique

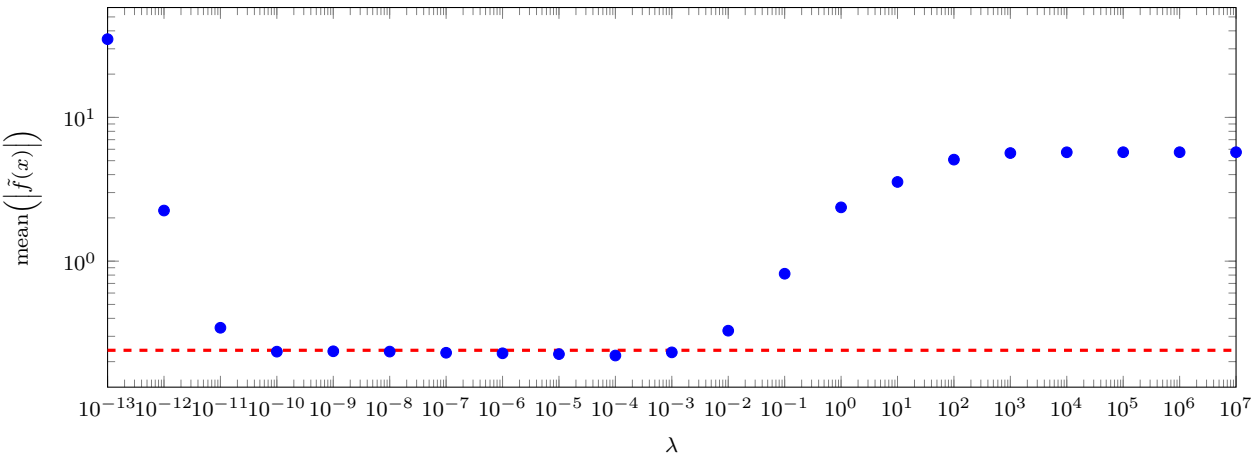

Figure 19: The blue marks are the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3, 3]$ for different values of λ using OKR trained with noise free trajectories. the dashed red line is the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3, 3]$ calculated using the SLDMD method trained with noise free trajectories.

performs almost identically to the SLDMD technique for λ values between 10^{-10} and 10^{-3}.

If the Gram matrix is not full rank, which is the case for both trials of the numerical experiments, then inversion of $G_r + \lambda I_M$ is numerically unstable for small values of λ. As such, consistent with Figures 18 and 19, it is expected that the approximation $\hat{f}$ computed using OKR with a small value of λ would be poor. On the other hand, the regularized inverse $(G_r + \lambda I_M)^{-1}$ converges to the zero matrix as λ increases, which explains the error plateau seen in Figures 18 and 19 for $\lambda > 10^2$.

In theory, in the case where the trajectories are noise free and the inner products in the Gram matrix G_r are evaluated exactly, the SLDMD algorithm should outperform OKR for any value of λ, since the regularization introduces a bias. The authors speculate that the slight improvement, seen in Figures 18 and 19 for $10^{-10} < \lambda < 10^{-3}$ is due to the integration errors introduced when computing the entries of the Gram matrix using Simpson's rule. Since regularization can prevent over-fitting when the underlying data are noisy, a more accurate approximation may be obtained for some values of λ. In the case where the trajectories are corrupted with measurement noise, there are errors in both G_r and D of (6.6.3), which make

the effects of λ even more significant, consistent with Figure 18.

6.8 Conclusion

This chapter introduces a novel approach towards the construction of a finite rank representation of the Liouville operator. New results on the construction of singular values and functions of the finite rank operator using singular values and vectors of a matrix representation are also obtained. The singular values and functions give rise to a new dynamic mode decomposition model that is shown to be equivalent to regression using occupation kernels.

Numerical experiments that study the effect of the regularization parameter indicate that regularization may yield better results when the data are corrupted by integration errors and measurement noise. In order to support this conjecture a detailed Monte Carlo simulation is needed, which is a part of future work.

CHAPTER VII

A DYNAMIC MODE DECOMPOSITION APPROACH TO PARAMETER IDENTIFICATION

7.1 Introduction

This chapter presents a data-driven algorithm for simultaneous system identification and parameter estimation in control-affine nonlinear systems. Parameter estimation is achieved by training a data-driven predictive model using state-action measurements and various known values at the parameters of interest. The predictive model is then used in conjunction with state-action data corresponding to unknown values of the parameters to estimate the said unknown value.

7.2 Overview

7.2.1 Problem Statement

Consider a nonlinear control-affine discrete-time dynamical system of the form

$$z_{k+1} = f_0(z_k, \theta) + g_0(z_k, \theta)u_k,$$

where f_0 and g_0 are unknown continuous functions and θ denotes a vector of parameters of interest. Unlike traditional system identification and parameter estimation problems, where the uncertain dynamics are expressed as linear or nonlinear combinations of known basis functions and unknown parameters, we do not make any assumptions on the structure of the unknown functions beyond continuity.

The objective in this chapter is to learn the current set of parameters $\pi \in \mathbb{R}^p$ using an observer trajectory of the system, i.e., measurements of the system state and the input expressed in the form of a *query dataset* $\{u_i, z_i, w_i\}_{i=1}^N$, where

$$w_i \coloneqq z_{i+1} = F(z_i, u_i, \pi) = f_0(z_i, \pi) + g_0(z_i, \pi)u_i. \tag{7.2.1}$$

For example, in the case of a package delivery drone, the parameter of interest θ could correspond to mass of a package, and the current parameter π could be the mass of the package currently being delivered. The objective is then to estimate π, the current mass.

7.2.2 Approach

In this chapter, we solve the problem above by training a model that encodes the structural relationship between the parameters and the dynamical system. To train such a model, we need a *training dataset* comprised of system trajectories $\{u_i^j, z_i^j, w_i^j, \pi^j\}_{j=1,i=1}^{P,M}$, where π^j are vectors of known values of the parameters of interest, satisfying

$$y_i^j \coloneqq x_{i+1}^j = F(x_i^j, u_i^j, \pi^j) = f(x_i^j, \pi^j) + g(x_i^j, \pi^j)u_i^j. \tag{7.2.2}$$

For example, in the case of the package delivery drone, we can record a training dataset by flying the drone with packages of varying known masses π^j.

We then train a model $\hat{F} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n$ to estimate F using the training dataset and use it to infer a new, unknown mass that corresponds to the query dataset described in the previous section.

In order to learn the parameters π, we need to solve the following optimization problem

$$\hat{\pi} = \arg\min_{\eta \in \Pi} \frac{1}{N} \sum_{i=1}^{N} (F(z_i, u_i, \eta) - F(z_i, u_i, \pi))^2, \tag{7.2.3}$$

where $\Pi \subset \mathbb{R}^p$ is a known set containing π. Since F is unknown a data-driven optimization problem is formulated as

$$\hat{\pi} = \arg\min_{\eta \in \Pi} \frac{1}{N} \sum_{i=1}^{N} (\hat{F}(z_i, u_i, \eta) - w_i)^2. \tag{7.2.4}$$

In this chapter, we adapt the DCLDMD technique developed in [64], to generate the model $\hat{F}$. In particular, [64] considered modeling of closed dynamical systems under feedback control laws, whereas in this chapter, we adapt the techniques in [64] to model open dynamical systems.

7.3 The Control Liouville Operator and its Finite-rank Representation

In order to facilitate the description of the controlled dynamical system in terms of Koopman-like operators, the vvRKHS framework is utilized .

Definition 7.3.1 *Given an operator $A_{f,g} : \tilde{H}_d \to H$, the tuples $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^{\infty}$, with $\sigma_i \in \mathbb{R}^n$, $\phi_i \in \tilde{H}_d$, and $\psi_i \in H$, are singular values, left singular vectors, and right singular vectors of $A_{f,g}$, respectively, if $\forall h \in \operatorname{span} d$,*

$$A_{f,g}h = \sum_{i=1}^{\infty} \sigma_i \psi_i \langle h, \phi_i \rangle_{\tilde{H}_d} . \tag{7.3.1}$$

Let $h_{\mathrm{id}} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity function. Given singular values, left singular vectors, and right singular vectors of $A_{f,g}$ and a control input u, the dynamics of the system can be expressed as

$$x_{k+1} = y_k = \begin{pmatrix} A_{f,g}(h_{\mathrm{id}})_1(x_k) \\ \vdots \\ A_{f,g}(h_{\mathrm{id}})_n(x_k) \end{pmatrix} \begin{pmatrix} 1 \\ u_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{\infty} \sigma_i \psi_i(x_k) \langle (h_{\mathrm{id}})_1, \phi_i \rangle_{\tilde{H}_d} \\ \vdots \\ \sum_{i=1}^{\infty} \sigma_i \psi_i(x_k) \langle (h_{\mathrm{id}})_M, \phi_i \rangle_{\tilde{H}_d} \end{pmatrix} \begin{pmatrix} 1 \\ u_k \end{pmatrix}, \tag{7.3.2}$$

assuming $h_{\mathrm{id}} \in \operatorname{span} d$.

To facilitate computation, an explicit finite-rank representation of $A_{f,g}$ is needed to determine the dynamic modes of the resultant system. In the following, finite collections of linearly independent vectors, d^M and β^M are selected to establish the needed finite-rank representation.

$$d^M = \{K_d(\cdot, x_i)\}_{i=1}^{M} \subset \tilde{H}_d \tag{7.3.3}$$

is selected to be the domain of $A_{f,g}$. The corresponding Gram matrix is denoted by $G_{d^M} = \left(\langle d_i, d_j\rangle_{\tilde{H}_d}\right)_{i,j=1}^{M}$. The output of $A_{f,g}$ is projected onto the span of vector valued kernels

$$\beta^M = \{K_{x_i,\overline{u}_i}\}_{i=1}^{M} = \{\mathrm{diag}(K(\cdot, x_i)\cdot \overline{u}_i\}_{i=1}^{M} \subset H. \tag{7.3.4}$$

The corresponding Gram matrix is denoted by $G_{\beta^M} = \left(\langle \beta_i, \beta_j\rangle_H\right)_{i,j=1}^{M}$.

A rank-M (or less) representation of the operator $A_{f,g}$ is then given by $P_{\beta^M} A_{f,g} P_{d^M} : \tilde{H}_d \to \mathrm{span}\, \beta^M$, where P_{d^M} and P_{β^M} denote projection operators onto $\mathrm{span}\, d^M$ and $\mathrm{span}\, \beta^M$, respectively.

7.4 Matrix Representation of the Finite-rank Operator

To formulate a matrix representation of the finite-rank operator $P_{\beta^M} A_{f,g} P_{d^M}$, the operator is restricted to $\mathrm{span}\, d^M$ to yield the operator $P_{\beta^M} A_{f,g}|_{d^M} : \mathrm{span}\, d^M \to \mathrm{span}\, \beta^M$. For brevity of exposition, the superscript M is suppressed hereafter and d and β are interpreted as $M-$dimensional vectors.

The adjoint relationship can be derived as follows, $\langle d_i, A^*_{f,g}\beta_k\rangle_{\tilde{H}_d} = \langle A_{f,g} d_i, \beta_k\rangle_H = \langle [A_{f,g} d_i](x_k), \begin{pmatrix} 1 & u_k^\top \end{pmatrix}\rangle_{\mathbb{R}^{m+1}} = K_d(x_i, x_{k+1})$ as shown [64]. Therefore, $A^*_{f,g}\beta_i = A^*_{f,g} K_{x_i,\overline{u}_i} = K_d(\cdot, y_i)$, where $y_i = F(x_i, u_i)$.

Proposition 7.4.1 *If $h = \delta^\top d \in \mathrm{span}\, d$ is a function with coefficients $\delta \in \mathbb{R}^M$ and if $g = P_\beta A_{f,g} h$, then $g = a^\top \beta$, where $a = G_\beta^+ I\delta$ and $(\cdot)^+$ denotes the Moore-Penrose pseudoinverse.*

Proof. Since $A^*_{f,g}\beta_j = A^*_{f,g} K_{x_i,\overline{u}_i} = K_d(\cdot, y_i)$, where $y_i = F(x_i, u_i)$. Note that since g is a projection of $A_{f,g}h$ onto $\mathrm{span}\, \beta$, $g = a^\top \beta$ for any a that solves

$$G_\beta a = \begin{bmatrix} \langle A_{f,g}h, \beta_1\rangle_H \\ \vdots \\ \langle A_{f,g}h, \beta_M\rangle_H \end{bmatrix} = \begin{bmatrix} \left\langle h, A^*_{f,g}\beta_1\right\rangle_H \\ \vdots \\ \left\langle h, A^*_{f,g}\beta_M\right\rangle_H \end{bmatrix} = \begin{bmatrix} \left\langle h, A^*_{f,g}K_{x_1,\overline{u}_1}\right\rangle_H \\ \vdots \\ \left\langle h, A^*_{f,g}K_{x_M,\overline{u}_M}\right\rangle_H \end{bmatrix} \tag{7.4.1}$$

Using the adjoint relationship,

$$G_\beta a = \begin{bmatrix} \langle h, K_d(\cdot, y_1)\rangle_H \\ \vdots \\ \langle h, K_d(\cdot, y_M)\rangle_H \end{bmatrix} = \begin{bmatrix} \left\langle \delta^\top d, K_d(\cdot, y_1)\right\rangle_H \\ \vdots \\ \left\langle \delta^\top d, K_d(\cdot, y_M)\right\rangle_H \end{bmatrix} = I\delta, \tag{7.4.2}$$

where $I = \left(\langle K_d(\cdot, x_i), K_d(\cdot, y_j)\rangle_{\tilde{H}_d}\right)_{i,j=1}^{M}$ Selecting the solution

$$a = G_\beta^+ I^\top \delta \tag{7.4.3}$$

of (7.4.2), a matrix representation $[P_\beta A_{f,g}]_d^\beta$ of the operator $P_\beta A_{f,g}|_d$ is obtained as $G_\beta^+ I^\top$. ■

Note that matrix representations are generally not unique. Different representations may be obtained by selecting different solutions of (7.4.1) and (7.4.3). In the case where the Gram matrix G_β is nonsingular, equation (7.4.1) has a unique solutions, resulting in the unique matrix representation $G_\beta^{-1} I^\top$.

In the following section, the matrix representation $[P_\beta A_{f,g}]_d^\beta$ is used to construct a data-driven representation of the singular values and the left and right singular functions of $P_\beta A_{f,g}|_d$.

7.5 Pseudo-Singular Functions of the Finite-rank Operator

The following proposition states that the SVD-like decomposition of $P_\beta A_{f,g}|_d$ can be computed using matrices in the matrix representation $[P_\beta A_{f,g}]_d^\beta$ derived in the previous section.

Proposition 7.5.1 *If (W, Σ, V) is the SVD of $G_\beta^+ I^\top G_d^+$ with $W = \begin{bmatrix} w_1, & \ldots, & w_M \end{bmatrix}$, $V = \begin{bmatrix} v_1, & \ldots, & v_M \end{bmatrix}$, and $\Sigma = \operatorname{diag}\left(\begin{bmatrix} \sigma_1, & \ldots, & \sigma_M \end{bmatrix}\right)$, then for all $i = 1, \ldots, M$, σ_i are pseudo-singular values of $P_\beta A_{f,g}|_d$ with left pseudo-singular functions $\phi_i := v_i^\top d$ and right pseudo-singular functions $\psi_i := w_i^\top \beta$.*

Proof. Let $\phi_i = v_i^\top d$ and $\psi_i = w_i^\top \beta$ and $h = \delta^\top d$. Then,

$$P_\beta A_{f,g} h = \sum_{i=1}^{M} \sigma_i \psi_i \langle h, \phi_i\rangle_{\tilde{H}_d} \iff P_\beta A_{f,g} \delta^\top d = \sum_{i=1}^{M} \sigma_i w_i^\top \beta \left\langle \delta^\top d, v_i^\top d\right\rangle_{\tilde{H}_d}$$

Using the finite-rank representation, the collection $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^M$, is an SVD of $P_\beta A_{f,g}|_d$, if for all $\delta \in \mathbb{R}^M$,

$$\left(G_\beta^+ I^\top \delta\right)^\top \beta = \left(\sum_{i=1}^M \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{\tilde{H}_d} w_i^\top\right) \beta. \tag{7.5.1}$$

Simple matrix manipulations yield the chain of implications

$$\begin{aligned}
(7.5.1) &\Longleftarrow \forall \delta \in \mathbb{R}^M, G_\beta^+ I^\top \delta = \sum_{i=1}^M \sigma_i \left\langle \delta^\top d, v_i^\top d \right\rangle_{\tilde{H}_d} w_i \\
&\iff \forall \delta \in \mathbb{R}^M, G_\beta^+ I^\top \delta = \sum_{i=1}^M \sigma_i \left(w_i v_i^\top G_d\right) \delta \\
&\Longleftarrow G_\beta^+ I^\top = \sum_{i=1}^M \sigma_i \left(w_i v_i^\top\right) G_d \\
&\Longleftarrow G_\beta^+ I^\top G_d^+ = \sum_{i=1}^M \sigma_i w_i v_i^\top = W \Sigma V^\top,
\end{aligned}$$

which proves the proposition. ■

Remark 7.5.1 *The standard usage of the term SVD refers to a decomposition using orthonormal left and right singular vectors. The left singular functions $\{\phi_i\}_{i=1}^M$ and right singular functions $\{\psi_i\}_{i=1}^M$ are not necessarily orthonormal. Therefore, the decomposition in the previous proposition is not an SVD of the finite-rank operator $P_\beta A_{f,g}|_d$.*

7.6 The DCLDMDPID Algorithm

7.6.1 DCLDMD

In order to identify the parameters of the system

$$z_{k+1} = f_0(z_k, \pi) + g_0(z_k, \pi) u_k,$$

we first consider a modified system of the form

$$y_k = x_{k+1} = \begin{pmatrix} z_{k+1} \\ \pi_{k+1} \end{pmatrix} = \begin{pmatrix} f_0(z_k) \\ \pi_k \end{pmatrix} + \begin{pmatrix} g_0(z_k) \\ 0 \end{pmatrix} u_k, \tag{7.6.1}$$

where $\pi_0 = \pi$. Then we can use the training dataset $\{u_i^j, z_i^j, y_i^j, \pi^j\}_{j=1,i=1}^{P,M}$ to learn a control affine discrete time nonlinear system of the form

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

Motivated by (7.3.1), assuming that $h_{\mathrm{id},j} \in \tilde{H}_d$ for $j = 1, \cdots, n$, the system dynamics are approximated using the finite-rank representation, with rank at most M, as

$$x_{k+1} \approx \hat{F}_M(x_k, u_k) := [P_\beta A_{f,g} P_d h_{\mathrm{id}}](x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix},$$

where $P_\beta A_{f,g} P_d h_{\mathrm{id}}$ denotes row-wise operation of the operator $P_\beta A_{f,g}$ on the function $P_d h_{\mathrm{id}}$. Using the definition of singular values and singular functions of $P_\beta A_{f,g} \mid_d$,

$$x_{k+1} \approx \sum_{i=1}^{M} \sigma_i \xi_i w_i^\top \beta(x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix} = \xi \Sigma W^\top \beta(x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix}, \tag{7.6.2}$$

where $\xi_i := \langle P_d h_{\mathrm{id}}, \phi_i \rangle_{\tilde{H}_d}$ and $\xi := \begin{bmatrix} \xi_1, & \ldots, & \xi_M \end{bmatrix}$.

The modes ξ can be computed using $\phi_i = v_i^\top d$ as

$$\begin{aligned} \xi &= \begin{bmatrix} \left\langle P_d h_{\mathrm{id},1}, v_1^\top d \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle P_d h_{\mathrm{id},1}, v_M^\top d \right\rangle_{\tilde{H}_d} \\ \vdots & \ddots & \vdots \\ \left\langle P_d h_{\mathrm{id},n}, v_1^\top d \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle P_d h_{\mathrm{id},n}, v_M^\top d \right\rangle_{\tilde{H}_d} \end{bmatrix} \\ &= \begin{bmatrix} \left\langle \delta_1^\top d, d_1 \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle \delta_1^\top d, d_M \right\rangle_{\tilde{H}_d} \\ \vdots & \ddots & \vdots \\ \left\langle \delta_n^\top d, d_1 \right\rangle_{\tilde{H}_d}, & \ldots, & \left\langle \delta_n^\top d, d_M \right\rangle_{\tilde{H}_d} \end{bmatrix} V = \delta^\top G_d V, \end{aligned}$$

where $\delta := \begin{bmatrix} \delta_1, & \ldots, & \delta_n \end{bmatrix}$. Using the reproducing property of the reproducing kernel of $\tilde{H}_d$, the coefficients δ_i in the projection of $h_{\mathrm{id},i}$ onto d satisfy

$$G_d \delta_i = \begin{bmatrix} \langle (h_{\mathrm{id}})_i, d_1 \rangle_{\tilde{H}_d} \\ \vdots \\ \langle (h_{\mathrm{id}})_i, d_M \rangle_{\tilde{H}_d} \end{bmatrix} = \begin{bmatrix} (x_1)_i \\ \vdots \\ (x_M)_i \end{bmatrix}.$$

Letting $D \coloneqq \left((x_j)_i\right)_{i,j=1}^{n,M}$ it can be concluded that $\delta^\top G_d = D$. Finally, the modes ξ are given by $\xi = DV$ and the estimated open-loop model is given by

$$x_{k+1} \approx \hat{F}_M(x_k, u_k) = DV\Sigma W^\top \beta(x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix} = D(G_\beta^+ I^\top G_d^+)^\top \beta(x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix}. \tag{7.6.3}$$

the approximation $\hat{f}_M(x)$, an approximation of the drift dynamics, $f(x)$, is given by the first column of $D(G_\beta^+ I^\top G_d^+)^\top \beta(x_k)$ and $\hat{g}_M(x)$, an approximation of the control-effectiveness matrix, $g(x)$, is given by the last m columns of $D(G_\beta^+ I^\top G_d^+)^\top \beta(x_k)$.

7.6.2 Parameter Identification

Once the a system $\hat{F}$ is obtained, we solve for $\hat{\pi}$ as in (7.2.4). The specific approach used in this chapter uses a grid of candidate parameters in the set Π. Every point in this grid is used as an initial condition for the last p number of states, where p is the dimension of π. Then, the prediction of $\hat{F}$ is compared with the query dataset as in (7.2.4), and the grid point that minimizes the error metric is $\hat{\pi}$ (see Algorithm 3).

7.7 Numerical Experiments

As a demonstration of the efficacy of the developed algorithm, we apply the method to the controlled Duffing oscillator.

The controlled Duffing oscillator is a nonlinear dynamical system with state-space form

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -\delta x_2 - \beta x_1 - \alpha x_1^3 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 + \sin(x_1) \end{pmatrix} u \tag{7.7.1}$$

where α, β, δ are coefficients in $\mathbb{R}$, $[x_1, x_2]^\top \in \mathbb{R}^2$ is the state, and $u \in \mathbb{R}$ is the control input. For the experiments the parameters are selected to be: $\delta = 0$, $\alpha = 1$, and $\beta = -1$.

We modify the system to include the parameters as states

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{\alpha} \\ \dot{\beta} \\ \dot{\delta} \end{pmatrix} = \begin{pmatrix} x_2 \\ -\delta x_2 - \beta x_1 - \alpha x_1^3 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 + \sin(x_1) \\ 0 \\ 0 \\ 0 \end{pmatrix} u, \tag{7.7.2}$$

then descretize (7.7.2) using a time step of 0.1 seconds to yield a discrete-time, control-affine dynamical system of the form $x_{k+1} = F(x_k) + G(x_k)u_k$. Using the tuples $\{(x_k, x_{k+1}, u_k)\}_{k=1}^n$ generated by the dynamical system, we aim to identfiy the system along with parameters α, β, and δ, where $\alpha_0 = \alpha$,$\beta_0 = \beta$ and $\delta_0 = \delta$.

In the implementation of DCLDMDPID, we select 20000 initial conditions sampled randomly using a uniform distribution from the set $[-3, 3] \times [-3, 3] \times [-3, 3] \times [-3, 3] \times [-3, 3] \subset \mathbb{R}^5$ and the control inputs are sampled uniformly from the interval $[-2, 2] \subset \mathbb{R}$.

The Gaussian radial basis function kernel $\tilde{K}(x, y) = \mathrm{e}^{\frac{-\|x-y\|_2^2}{\sigma}}$ is used for $\tilde{H}$, with kernel width $\sigma = 20$. For H, we associate to each pair $\{(x_k, u_k)\}_{k=1}^n$ a kernel $K_{x_k,\bar{u}_k} \coloneqq \begin{pmatrix} 1 & u_k^\top \end{pmatrix} K_{x_k} \in H$. Here we use the kernel operator $K_{x_i} \coloneqq \operatorname{diag}\begin{pmatrix} \tilde{K}_{x_1} & \cdots & \tilde{K}_{x_{m+1}} \end{pmatrix}$ where $\tilde{K}_{x_j}(y) = \mathrm{e}^{\frac{-\left\|x_j - y\right\|_2^2}{\sigma}}$ for $j = 1, \ldots, m+1$, with $\sigma = 20$. Lastly, we select $\varepsilon = 10^{-6}$ for regularization of the Gram matrices in order to ensure invertibility of both $\tilde{G}$ and G, instead of the pseudoinverse, in the finite-rank representation (see Algorithm 3).

Once a system $\hat{F}$ is identified, the process of identifying the parameters is conducted by comparing the trajectories of the identified system using permutations of the possible parameters, i.e. different initial conditions for the added states, with the trajectories of the true system in (7.7.1). the permutation of the possible parameters are the nodes of an equally spaced mesh in $[-3, 3] \times [-3, 3] \times [-3, 3] \subset \mathbb{R}^3$ with a spacing of 0.5 corresponding to the three parameters α, β, and δ. The permutation that produces the smallest mean squared error (MSE) $\frac{1}{50}\sum_{i=1}^{50}([x_1; x_2] - \hat{F}_{(1,2)}(x_i, u_i))^2$ is determined to be the parameter estimations of $\pi = [\alpha, \beta, \delta]$, where $\hat{F}_{(1,2)}(x_i, u_i)$ is the first and second state output of $\hat{F}$.

Algorithm 3 The DCLDMDPID algorithm

Input: Training dataset composed of state snapshots $\{z_i^j\}_{i=1}^M$, parameter values $\{\pi^j\}_{i=1}^M$, control set $\{u_i^j\}_{i=1}^M$, a label set $\{w_i^j\}_{i=1}^M$ such that $w_i^j = F(z_i^j, u_i^j, \pi^j)$, a query dataset composed of state snapshots $\{z_i\}_{i=1}^N$, control set $\{u_i\}_{i=1}^N$, a label set $\{w_i\}_{i=1}^N$ such that $w_i = F(z_i, u_i, \pi)$, and reproducing kernels $\tilde{K}_d$ and K of $\tilde{H}_d$ and H, respectively, and a mesh of of parameter space Π_{mesh}.

Output: $\{\xi_j, \sigma_j, \varphi_j, \phi_j\}_{j=1}^M$, $\hat{F}$ and $\hat{\pi}$.

$x_i = [z_i; \pi_i]$

$y_i = [w_i; \pi_i]$

$\beta \leftarrow [K_{x_i, \overline{u}_i}]_{i=1}^M$

$G_\beta \leftarrow \left(\left\langle K_{x_i,\overline{u}_i}, K_{x_j,\overline{u}_j}\right\rangle_H\right)_{i,j=1}^M$

$G_{d^M} \leftarrow \left(\langle K_d(\cdot, x_i), K_d(\cdot, x_j)\rangle_{\tilde{H}_d}\right)_{i,j=1}^M$

$I = \left(\langle K_d(\cdot, x_i), K_d(\cdot, y_j)\rangle_{\tilde{H}_d}\right)_{i,j=1}^M$

$D \leftarrow \left((x_j)_i\right)_{i,j=1}^{n,M}$

$(W, \Sigma, V) \leftarrow$ SVD of $G_\beta^+ I^\top G_d^+$

$\xi \leftarrow DV$

$\phi_j \leftarrow \sum_{i=1}^M \leftarrow (V)_{i,j} \left(K_d(\cdot, x_i)\right)$

$\psi_j \leftarrow \sum_{i=1}^M (W)_{i,j} \left[\begin{bmatrix} 1 & u_i(t)^\top \end{bmatrix} K_{x_i,\overline{u}_i}\right](\cdot)$

$\hat{F}(x_k, u_k) \leftarrow DV\Sigma W^\top \beta(x_k) \begin{pmatrix} 1 \\ u_k \end{pmatrix}$

$\hat{\pi} \leftarrow \arg\min_{\eta \in \Pi_{mesh}} \sum_{k=1}^N (w_k - \hat{F}([z_k; \eta], u_k))^2$

Return $\{\xi_j, \sigma_j, \varphi_j, \phi_j\}_{j=1}^M$, $\hat{F}$ and $\hat{\pi}$.

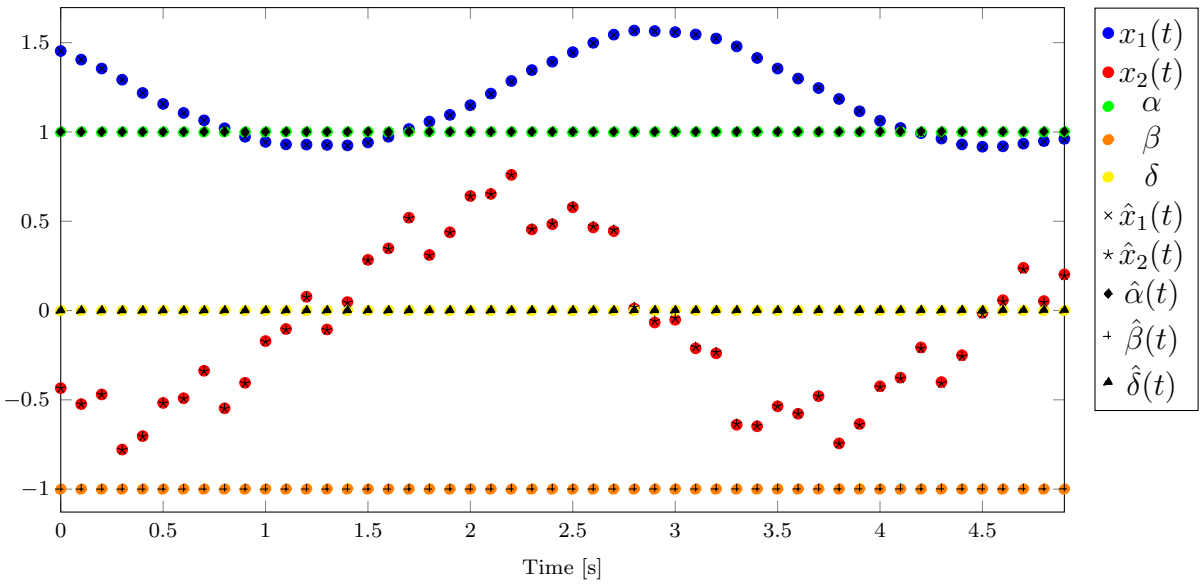

Figure 20: A comparison between the true system trajectories x_1 and x_2 and the trajectories of the identified system $\hat{x}_1$ and $\hat{x}_2$. In addition to the comparison of the true system parameters α, β, and δ and their corresponding parameters as states $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$.

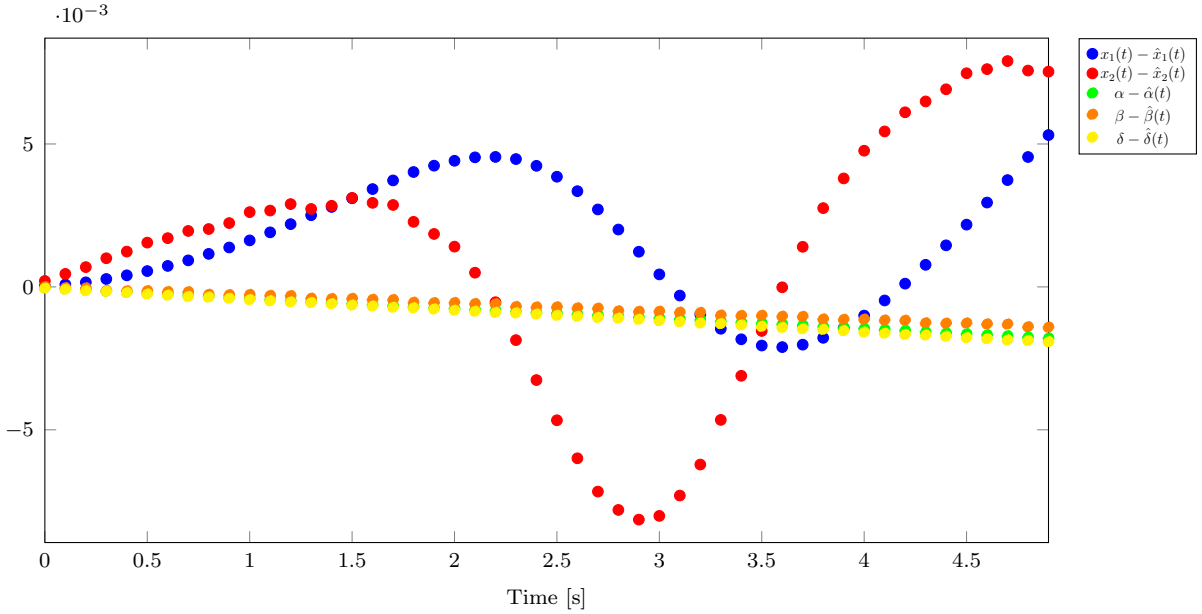

Figure 21: The difference between the trajectories of the true system and the identified system $x_1 - \hat{x}_1$ and $x_2 - \hat{x}_2$. And the difference between the true parameters α, β, δ and the change in the parameters as states form the identified system $\hat{\alpha}, \hat{\beta}, \hat{\delta}$.

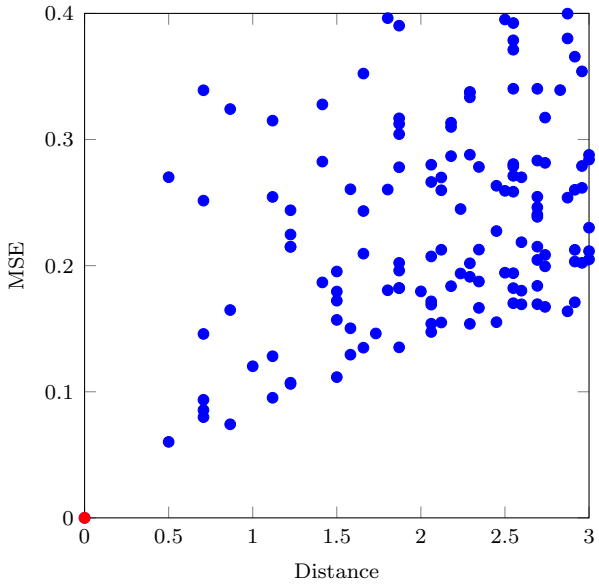

Figure 22: The MSE computed using different parameter values on the mesh that contains the true parameter values, plotted against the Euclidean distance of the mesh point from the exact values $\delta = 0$, $\alpha = 1$, and $\beta = -1$. The red dot shows the parameter values with the smallest MSE.

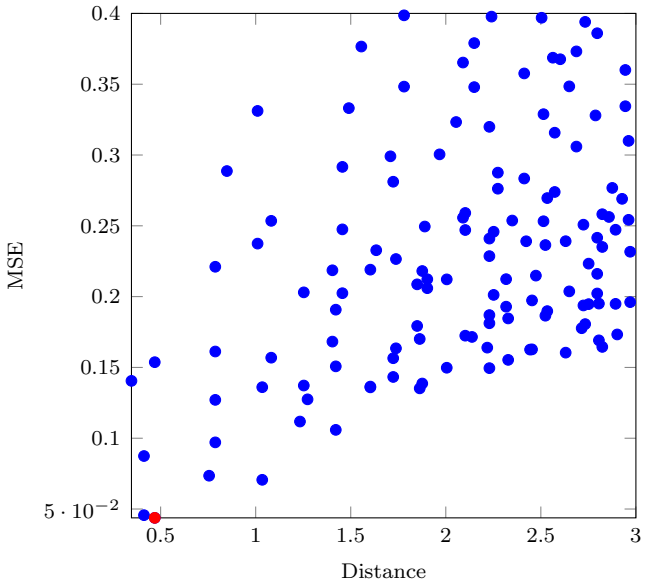

Figure 23: The MSE computed using different parameter values on the mesh that does not contain the true parameter values, plotted against the Euclidean distance of the mesh point from the exact values $\delta = 0$, $\alpha = 1$, and $\beta = -1$. The red dot shows the parameter values with the smallest MSE

7.7.1 Discussion

Figure 20 shows the trajectories of the true system and the identified system, along with the true parameters and the corresponding identified parameters as states, where the initial conditions of the parameters as states are determined to be $\hat{\alpha}(0) = 1$, $\hat{\beta}(0) = -1$, and $\hat{\delta}(0) = 0$. Figure 21 shows the trajectory differences between the true system and the identified system, along with the drift of the parameters as states from the true values, all of which are at most in the order of 10^{-3} over the span of 5 seconds. Note that naturally, obtaining the exact parameter using this method is not guaranteed. It was only possible in this example because the mesh that was used for parameter identification happened to contain the exact parameter vector. If the true parameter vector is not in the mesh, the algorithm returns the best-fitting parameter vector from the mesh.

Figure 22 shows values of the MSE computed using different parameter combinations on the mesh against the Euclidean distance of the vector of parameters $[\delta, \alpha, \beta]^\top$ form the vector $[0, 1, -1]^\top$ of true values. As seen in Figure 22, for the case where the true parameter is in the mesh, we see a clear separation between the error generated by the true parameter and those generated by other parameters on the mesh. Note that since the optimization problem in (7.2.4) is not expected to be convex in general, we expect such separation only locally, not globally as seen in Figure 22.

In order to demonstrate the effect of the choice of the mesh, another parameter identification experiment was conducted using the same training and query datasets but using a shifted mesh that does not contain the true parameter values $\delta = 0$, $\alpha = 1$, and $\beta = -1$. In this case we use a mesh composed of equally spaced nodes in $[-3.3, 2.7] \times [-3.3, 2.7] \times [-3.3, 2.7]] \subset \mathbb{R}^3$ with a spacing of 0.5. Figure 23 shows values of the MSE against the Euclidean distance from the true parameter in the parameter space, similar to 22. In this case, it can be seen that the parameter values that produce the smallest MSE ($\hat{\alpha} = 0.7$, $\hat{\beta} = -1.3$, and $\hat{\delta} = -0.3$.) is not the closest node in the mesh to the true values in Euclidean distance, but we still see a separation between a cluster of 'low-error' parameter vectors and other parameter vectors

on the mesh.

7.8 Conclusion

This chapter introduces a novel approach towards the construction of a finite-rank representation of the control Liouville operator. New results on the construction of singular values and functions of the finite rank operator using singular values and vectors of a matrix representation are also obtained. Once the singular values and functions are at hand, the drift dynamics and the control effectiveness can be approximated, which facilitates the prediction of the states of the dynamical system in response to an open-loop control signal. Once a predictive model is generated, its predictions can be compared with the trajectory of the true system using different guesses of the parameters as initial conditions of the augmented states. A weakness of the current method is that the parameters identified can only choose between the nodes of the mesh that is defined by the user. If the mesh does not include the exact parameters, it is impossible to identify the exact parameters.

A numerical experiment using the Duffing oscillator is presented to demonstrate the performance of the developed technique, while future work will focus on analyzing the effects of integration error and measurement noise on the DCLDMDPID method, incorporating partial system knowledge to enhance identification, investigating data-driven sensitivity analysis to enable gradient-based parameter identification and investigating other gradient-free global optimization techniques in place of the current grid-based approach.

CHAPTER VIII

CONCLUSION

8.1 Summary

This dissertation studied data-driven identification of nonlinear dynamical systems using lifting techniques. The central objective was to develop models that are useful both for prediction and for control-oriented analysis when the governing equations are unknown or only partially known. In contrast with classical linear system identification, where the model structure is usually fixed and the main task is to estimate a finite collection of parameters, nonlinear system identification requires the simultaneous selection of a representation and the estimation of the corresponding model. The lifting perspective addresses this difficulty by representing nonlinear dynamics through linear or approximately linear evolution in a higher-dimensional space.

The dissertation focused on two complementary classes of lifting methods. The first class is based on Koopman-type and Liouville-type operator representations, reproducing kernel Hilbert spaces, and occupation kernels. These methods lift trajectories or observable functions into function spaces where the dynamics can be represented through linear operators. The second class is based on Carleman lifting, where nonlinear state dynamics are lifted into a monomial basis so that an infinite-dimensional linear system represents the original nonlinear system. In both cases, the infinite-dimensional representation must be approximated from finite data. The main problem addressed throughout the dissertation is therefore the construction of finite-dimensional lifted models that are computationally useful, data-driven, and accompanied by meaningful analytical or numerical guarantees.

A recurring theme in the dissertation is the replacement of numerical differentiation by integration along trajectories. Numerical derivatives are sensitive to measurement noise, especially in nonlinear system identification, where derivative estimates often enter directly into regression problems. Occupation kernels provide a way to encode trajectory information through integral functionals. When the unknown dynamics belong to an appropriate RKHS, endpoint differences of trajectories provide linear measurements of the dynamics without explicitly differentiating the measured signal. This observation motivates the occupation-kernel regression methods and the Liouville-operator DMD methods developed in the dissertation.

Another recurring theme is the role of operator compactness and convergence. Koopman and DMD methods are attractive because they connect nonlinear dynamics to linear spectral analysis. However, finite-rank approximations of Koopman-type operators do not automatically provide reliable spectral approximations. The continuous-time Liouville framework used in this dissertation provides a path toward compact operators and norm-convergent finite-rank approximations under appropriate assumptions. This strengthens the connection between the data-driven finite-dimensional models and the underlying infinite-dimensional dynamics.

The dissertation also addressed the use of lifted models for controlled systems and parameter identification. In controlled systems, the model must account not only for state evolution, but also for the effect of input signals. Control occupation kernels and control Liouville operators provide a natural way to incorporate inputs into the RKHS/operator-theoretic framework. For parameter identification, the dissertation treated unknown parameters as augmented state variables and used a lifted operator-theoretic model to compare predicted and observed trajectories under candidate parameter values. This extends the lifting-based identification framework beyond prediction of state trajectories to recovery of unknown physical quantities.

8.2 Summary of Contributions

The first main contribution was the development of control occupation kernel regression for nonlinear higher-order control-affine systems. The method applies to systems of the form

$$\frac{d^s x}{dt^s} = f(x) + g(x)u, \tag{8.2.1}$$

where the drift dynamics f and the control effectiveness matrix g are unknown. Higher-order control occupation kernels were introduced to avoid state augmentation for certain higher-order systems and to represent the controlled dynamics through integral measurements in a vector-valued RKHS. The resulting regression problem is regularized in the RKHS norm and, by the representer theorem, its solution lies in the span of the control occupation kernels generated by the observed trajectories. This gives a finite-dimensional matrix formulation while preserving a basis that is naturally induced by the dynamics. Numerical experiments on the Duffing oscillator and a two-link robot manipulator demonstrated that the method can identify control-affine dynamics from noisy data and can produce models useful for control-oriented simulation.

The second contribution was a data-driven Carleman lifting method for nonlinear system identification with guaranteed error bounds. Classical Carleman linearization provides an infinite-dimensional linear representation of nonlinear dynamics, but practical computation requires truncation. The dissertation developed an identification method in which a finite-dimensional lifted linear system is estimated from trajectory data, while the truncation order is selected using a prescribed-error analysis. The resulting model produces trajectories that remain within a computable error bound of the nonlinear system trajectory under the assumptions of the chapter. The numerical study showed the expected trade-off between increasing the Carleman truncation order and the numerical conditioning of the data-driven identification problem. Although higher truncation order improves the model-based Carleman approximation, the identified system may become less accurate beyond a critical order because of conditioning and finite-data effects. This contribution therefore clarifies that, in

data-driven Carleman identification, the truncation order should be selected not only from the approximation power of the lifted representation, but also from the numerical stability of the identification problem.

The third contribution was the development of a continuous-time dynamic mode decomposition method for nonlinear control-affine systems using finite-rank approximations of control Liouville operators. The method constructs a finite-rank representation of the control Liouville operator from controlled trajectory data and then uses the singular values and singular functions of that operator to approximate the drift dynamics and the control effectiveness. Under sufficiently rich data, the finite-rank representation converges in norm to the true control Liouville operator. This is important because norm convergence gives a stronger approximation framework than strong-operator-topology convergence and is more compatible with spectral computations. The Duffing oscillator example showed that the identified model can accurately reproduce the response of the true system to a new open-loop input, and that both the drift and control effectiveness can be approximated over the region covered by the data.

The fourth contribution was the construction of a norm-convergent Liouville DMD formulation and the demonstration of its equivalence with occupation kernel regression. This result connects two viewpoints that initially appear different. The first viewpoint constructs a finite-rank approximation of a Liouville operator and obtains a DMD-type model from its singular functions. The second viewpoint directly solves a regression problem using occupation kernels as basis functions. The dissertation showed that, in the unregularized setting, these approaches produce equivalent models. This equivalence gives an operator-theoretic interpretation of occupation kernel regression and a regression-theoretic interpretation of Liouville DMD. The numerical experiments further showed that regularization can improve performance in the presence of measurement noise and numerical integration error, even though regularization introduces bias in the noise-free idealized setting.

The fifth contribution was a dynamic mode decomposition approach to parameter iden-

tification using discrete control Liouville operators. The method augments the system state with unknown parameters and constructs an RKHS-based lifted model of the augmented dynamics. Once the predictive model is constructed, candidate parameter values are used as initial conditions for the parameter states, and the predicted state trajectory is compared with an observed query trajectory. The parameter vector that minimizes the resulting prediction error is selected as the estimate. The Duffing oscillator example demonstrated that, when the true parameter vector lies on the search mesh, the method can recover the exact parameter values in the tested case. When the true parameter vector is not on the mesh, the method returns the best-fitting mesh point and still separates a low-error cluster of candidate parameters from higher-error candidates. This contribution shows that operator-theoretic lifted models can be used not only for trajectory prediction, but also for data-driven recovery of unknown parameters when the explicit dependence of the dynamics on those parameters is unavailable.

Taken together, these contributions advance a unified lifting-based approach to nonlinear system identification. Occupation kernels reduce sensitivity to numerical differentiation. Liouville operators provide continuous-time operator models and a route to norm-convergent DMD. Control occupation kernels and control Liouville operators extend the framework to input-driven systems. Carleman lifting provides a complementary finite-dimensional linear approximation with prescribed-error analysis. Augmented-state DCLDMD extends the framework to parameter identification. The common thread is the construction of useful finite-dimensional models from trajectory data while preserving as much analytical structure as possible from the underlying infinite-dimensional lifted representation.

8.3 Limitations

The methods developed in this dissertation also have limitations that clarify the scope of the results and motivate future work. For occupation-kernel and RKHS-based methods, the quality of the identified model depends on the selected kernel, the regularization parameter,

and the coverage of the state-input domain by the training trajectories. If the observed trajectories do not sufficiently excite a region of the state space or do not contain sufficiently rich control inputs, the identified drift and control effectiveness may generalize poorly in that region. This behavior was visible in the numerical examples, where errors tended to increase near the boundary of the domain covered by the data.

Computational cost is another important limitation. Occupation-kernel models are expressed as combinations of kernel functions associated with the training trajectories. As the number of trajectories grows, evaluating the identified dynamics at a new state may become expensive, especially for real-time feedback applications. The two-link robot manipulator example demonstrated that the identified model can be useful for control computation, but also highlighted the need for faster evaluation methods. Without additional compression, sparsification, or low-rank approximation, direct implementation of the identified model may be too slow for high-rate feedback control.

The Carleman-based method is limited by the growth of the lifted dimension and by numerical conditioning. Increasing the truncation order increases the expressive power of the lifted model, but it also increases the dimension of the linear system to be identified. This can amplify numerical errors and make matrix inversions ill-conditioned. The error bound developed in the Carleman chapter gives a conservative way to select a truncation order, but the numerical results indicate that the best observed trajectory approximation and the best guaranteed bound may occur at different truncation orders. This is partly because the current analysis bounds the full lifted state, while the states of interest may only be the original physical coordinates or a low-dimensional projection of the lifted state.

For the DMD and Liouville-operator methods, the theoretical convergence results rely on assumptions about the RKHS, the compactness of the relevant operators, and the richness of the data. In practice, data are finite, noisy, and generated by numerical integration or physical sensors. Therefore, integration error, measurement noise, and finite-sample effects can influence the Gram matrices and the resulting operator approximations. The numerical

experiments suggest that regularization may mitigate these effects, but a complete statistical and numerical perturbation analysis remains open.

For the parameter-identification method, the current implementation depends on a user-defined mesh in the parameter space. This creates an inherent resolution limitation: if the true parameter vector is not contained in the mesh, the method cannot return the exact value. In addition, the prediction-error landscape need not be convex, so a coarse grid may miss good parameter estimates or may select a parameter vector that fits the available query trajectory well but is not globally identifiable. This limitation suggests that the current grid-based implementation should be viewed as a proof of concept rather than a final optimization strategy.

8.4 Future Work

Several directions follow naturally from the results of the dissertation. First, future work should develop computationally efficient evaluation methods for occupation-kernel models. Possible approaches include sparse kernel expansions, greedy trajectory selection, low-rank kernel approximations, randomized numerical linear algebra, and offline compression of the occupation-kernel basis. These tools would make the identified models more practical for real-time prediction and feedback control.

Second, the effects of measurement noise and numerical integration error should be analyzed more systematically. The regression and DMD formulations in the dissertation already suggest that integration-based methods are less sensitive to high-frequency noise than derivative-based methods. However, the finite-dimensional matrices used by the algorithms are still affected by noisy trajectories, quadrature error, and finite sampling. Perturbation bounds for the Gram matrices, singular functions, identified vector fields, and predicted trajectories would make it possible to choose regularization parameters and sampling rates in a more principled way.

Third, kernel and hyperparameter selection remain important open problems. The RKHS

framework provides a flexible model class, but the selected kernel encodes assumptions about smoothness, length scale, and function structure. Automated or data-adaptive kernel selection methods would improve the robustness of the proposed algorithms. In the context of control-affine systems, future work should also investigate kernels that exploit partial structural knowledge, such as known state dependencies, sparsity, mechanical-system structure, or separability between drift and control terms.

Fourth, the Carleman lifting results can be strengthened by deriving less conservative error bounds for the states of interest. In many applications, the full lifted state is not directly meaningful; the objective is to approximate the original state variables over a prescribed time horizon. Bounds that directly estimate the error in the truncated physical coordinates could reduce conservatism and produce truncation-order selection rules that better match the observed numerical performance. Further investigation is also needed to quantify machine-precision effects and conditioning at high truncation orders.

Fifth, future work should extend the control Liouville and occupation-kernel methods toward closed-loop control design. The dissertation demonstrates that the identified models can predict responses to new open-loop inputs and can be used in control-oriented computations. The next step is to integrate the identified models into feedback design, model predictive control, adaptive control, or reinforcement-learning schemes while preserving stability and robustness guarantees. This will require understanding how model error propagates through the controller and how uncertainty in the identified dynamics can be accounted for during synthesis.

Sixth, the parameter-identification method should be extended beyond grid search. Data-driven sensitivity analysis could provide gradients of the prediction error with respect to candidate parameters, enabling gradient-based optimization. Alternatively, derivative-free global optimization methods could replace the mesh-based search while preserving the ability to handle nonconvex prediction-error landscapes. Incorporating partial physical knowledge, such as known signs, bounds, or conservation laws, may also improve identifiability and

reduce the dimension of the search space.

Finally, experimental validation is an important direction. The dissertation uses numerical examples to demonstrate the proposed methods, including the Duffing oscillator, the Van der Pol oscillator, and a two-link robot manipulator model. Applying the algorithms to laboratory systems would test their performance under sensor noise, actuator limits, unmodeled dynamics, sampling irregularity, and imperfect state measurements. Such experiments would also clarify which computational bottlenecks are most important in real-time settings.

8.5 Closing Remarks

The results of this dissertation show that lifting techniques provide a powerful framework for data-driven identification of nonlinear systems. By moving from the original nonlinear state dynamics to lifted function-space or monomial representations, it becomes possible to use linear operator ideas, regression, spectral approximation, and error analysis in settings where direct nonlinear modeling is difficult. The occupation-kernel and Liouville-operator methods developed here emphasize integration along trajectories, compact operator approximation, and control-aware RKHS constructions. The Carleman method emphasizes finite-dimensional lifted linear identification with prescribed-error guarantees. The parameter-identification method shows that lifted operator models can also be used to infer unknown physical quantities when their coupling to the dynamics is not explicitly known.

The broader conclusion is that lifting does not remove the difficulty of nonlinear system identification; rather, it reorganizes the difficulty into questions of representation, approximation, data richness, regularization, and computation. This reorganization is valuable because it exposes mathematical structure that can be analyzed and exploited. The dissertation contributes to this direction by developing new algorithms and theoretical connections that make lifting-based models more useful for prediction, control-oriented modeling, and parameter identification of nonlinear dynamical systems.

REFERENCES

[1] L. Ljung, *System identification — theory for the user*, Prentice Hall, 1999.

[2] Jan C. Willems, Paolo Rapisarda, Ivan Markovsky, and Bart L. M. De Moor, *A note on persistency of excitation*, Systems Control Lett. **54** (2005), no. 4, 325–329.

[3] Henk J van Waarde, Claudio De Persis, M Kanat Camlibel, and Pietro Tesi, *Willems' fundamental lemma for state-space systems and its extension to multiple datasets*, IEEE Control Systems Letters **4** (2020), no. 3, 602–607.

[4] Karel J Keesman and Karel J Keesman, *System identification: an introduction*, vol. 2, Springer, 2011.

[5] Oliver Nelles, *Nonlinear system identification*, 2002.

[6] Torsten Carleman, *Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires*, Acta Mathematica **59** (1932), 63–87.

[7] Bernard O Koopman, *Hamiltonian systems and transformation in hilbert space*, Proceedings of the National Academy of Sciences **17** (1931), no. 5, 315–318.

[8] P. Ioannou and J. Sun, *Robust adaptive control*, Prentice Hall, 1996.

[9] Jean-Jacques E. Slotine and Weiping Li, *On the adaptive control of robot manipulators*, Int. J. Robot. Res. **6** (1987), no. 3, 49–59.

[10] Muhammad Nasiruddin Mahyuddin, Jing Na, Guido Herrmann, Xuemei Ren, and Phil Barber, *Adaptive observer-based parameter estimation with application to road gradient and vehicle mass estimation*, IEEE Trans. Ind. Electron. **61** (2013), no. 6, 2851–2863.

[11] Sheng Chen, Stephen A Billings, and PM Grant, *Non-linear system identification using neural networks*, International journal of control **51** (1990), no. 6, 1191–1214.

[12] Sarah McCabe, Patricia Davies, and Detlev Seidel, *On the use of nonlinear autoregressive moving average models for simulation and system identification*, 1991 American Control Conference, IEEE, 1991, pp. 1758–1763.

[13] Mohammad Khalil, Sondipon Adhikari, and Abhijit Sarkar, *Linear system identification using proper orthogonal decomposition*, Mechanical Systems and Signal Processing **21** (2007), no. 8, 3123–3145.

[14] Vincent Laurain, Roland Tóth, Dario Piga, and Wei Xing Zheng, *An instrumental least squares support vector machine for nonlinear system identification*, Automatica **54** (2015), 340–347.

[15] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Nat. Acad. Sci. U.S.A. **113** (2016), no. 15, 3932–3937.

[16] ———, *Sparse identification of nonlinear dynamics with control (SINDYc)*, IFAC-PapersOnLine **49** (2016), no. 18, 710–715.

[17] J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor, *Dynanic mode decomposition - data-driven modeling of complex systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[18] Marko Budišić, Ryan Mohr, and Igor Mezić, *Applied Koopmanism*, Chaos: An Interdisciplinary Journal of Nonlinear Science **22** (2012), no. 4, 047510.

[19] Igor Mezić, *Spectral properties of dynamical systems, model reduction and decompositions*, Nonlinear Dyn. **41** (2005), no. 1, 309–325.

[20] Milan Korda and Igor Mezić, *On convergence of extended dynamic mode decomposition to the Koopman operator*, J. Nonlinear Sci. **28** (2018), no. 2, 687–710.

[21] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley, *A data-driven approximation of the Koopman operator: extending dynamic mode decomposition*, J. Nonlinear Sci. **25** (2015), no. 6, 1307–1346.

[22] Gert K. Pedersen, *Analysis now*, vol. 118, Springer Science & Business Media, 2012.

[23] Igor Mezić, *Analysis of fluid flows via spectral properties of the Koopman operator*, Annu. Rev. Fluid Mech. **45** (2013), 357–378.

[24] Efrain Gonzalez, Moad Abudia, Michael Jury, Rushikesh Kamalapurkar, and Joel A. Rosenfeld, *Anti-Koopmanism*, arXiv:2106.00106, Submitted to the Conference on Neural Information Processing Systems.

[25] Qianxiao Li, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis, *Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator*, Chaos: An Interdisciplinary Journal of Nonlinear Science **27** (2017), no. 10, 103111.

[26] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton, *Data-driven discovery of koopman eigenfunctions for control*, Machine Learning: Science and Technology **2** (2021), no. 3, 035023.

[27] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz, *Dynamic mode decomposition with control*, SIAM J. Appl. Dyn. Syst. **15** (2016), no. 1, 142–161.

[28] Milan Korda and Igor Mezić, *Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control*, Automatica **93** (2018), 149–160.

[29] Amit Surana, *Koopman operator based observer synthesis for control-affine nonlinear systems*, Proc. IEEE Conf. Decis. Control, IEEE, 2016, pp. 6492–6499.

[30] Debdipta Goswami and Derek A. Paley, *Bilinearization, reachability, and optimal control of control-affine nonlinear systems: a Koopman spectral approach*, IEEE Trans. Autom. Control **67** (2022), no. 6, 2715–2728.

[31] Robin Strässer, Manuel Schaller, Karl Worthmann, Julian Berberich, and Frank Allgöwer, *Koopman-based feedback design with stability guarantees*, IEEE Transactions on Automatic Control (2024).

[32] Suddhasattwa Das, Dimitrios Giannakis, and Joanna Slawinska, *Reproducing kernel Hilbert space compactification of unitary evolution groups*, Appl. Comput. Harmon. Anal. **54** (2021), 75–136.

[33] Joel A. Rosenfeld, Rushikesh Kamalapurkar, L. Forest Gruss, and Taylor T. Johnson, *Dynamic mode decomposition for continuous time systems with the Liouville operator*, J. Nonlinear Sci. **32** (2022), no. 1, 1–30.

[34] Joel A. Rosenfeld and Rushikesh Kamalapurkar, *Singular dynamic mode decomposition*, SIAM J. Appl. Dyn. Syst. **22** (2023), no. 3, 2357–2381.

[35] Moad Abudia, Joel A. Rosenfeld, and Rushikesh Kamalapurkar, *On convergent dynamic mode decomposition and its equivalence with occupation kernel regression*, IFAC-PapersOnLine, 2024, to appear.

[36] Gianluigi Pillonetto, Francesco Dinuzzo, Tianshi Chen, Giuseppe De Nicolao, and Lennart Ljung, *Kernel methods in system identification, machine learning and function estimation: A survey*, Automatica **50** (2014), no. 3, 657–682.

[37] Gianluigi Pillonetto and Giuseppe De Nicolao, *A new kernel-based approach for linear system identification*, Automatica **46** (2010), no. 1, 81–93.

[38] Gianluigi Pillonetto, Minh Ha Quang, and Alessandro Chiuso, *A new kernel-based approach for nonlinearsystem identification*, IEEE Transactions on Automatic Control **56** (2011), no. 12, 2825–2840.

[39] Gianluigi Pillonetto and Alessandro Chiuso, *Tuning complexity in regularized kernel-based regression and linear system identification: The robustness of the marginal likelihood estimator*, Automatica **58** (2015), 106–117.

[40] Joel A. Rosenfeld, Benjamin Russo, Rushikesh Kamalapurkar, and Taylor Johnson, *The occupation kernel method for nonlinear system identification*, SIAM J. Control Optim. (to appear), see arXiv:1909.11792.

[41] Jean B. Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat, *Nonlinear optimal control via occupation measures and LMI-relaxations*, SIAM J. Control Optim. **47** (2008), no. 4, 1643–1666.

[42] Anirudha Majumdar, Ram Vasudevan, Mark M. Tobenkin, and Russ Tedrake, *Convex optimization of nonlinear feedback controllers via occupation measures*, Int. J. Robot. Res. **33** (2014), no. 9, 1209–1230.

[43] Milan Korda, Didier Henrion, and Colin N. Jones, *Controller design and value function approximation for nonlinear dynamical systems*, Automatica **67** (2016), 54 – 66.

[44] Dongsik Chang, Wencen Wu, Catherine R Edwards, and Fumin Zhang, *Motion tomography: Mapping flow fields using autonomous underwater vehicles*, The International Journal of Robotics Research **36** (2017), no. 3, 320–336.

[45] Benjamin P. Russo, Rushikesh Kamalapurkar, Dongsik Chang, and Joel A. Rosenfeld, *Motion tomography via occupation kernels*, J. Comput. Dyn. **9** (2022), no. 1, 27–45.

[46] Xiuying Li and Joel A. Rosenfeld, *Fractional order system identification with occupation kernel regression*, IEEE Control Syst. Lett. (2020), 19 – 24.

[47] Joel A. Rosenfeld, Rushikesh Kamalapurkar, L. Forest Gruss, and Taylor T. Johnson, *Dynamic mode decomposition for continuous time systems with the Liouville operator*, J. Nonlinear Sci. **32** (2021), no. 1.

[48] Joel A. Rosenfeld and Rushikesh Kamalapurkar, *Dynamic mode decomposition with control Liouville operators*, Int. Symp. Math. Theory Netw. Syst., 2021, to appear, seearXiv:2101.02620.

[49] Marcelo Forets and Amaury Pouly, *Explicit error bounds for carleman linearization*, arXiv preprint arXiv:1711.02552 (2017).

[50] Marcelo Forets and Christian Schilling, *Reachability of weakly nonlinear systems using carleman linearization*, International Conference on Reachability Problems, Springer, 2021, pp. 85–99.

[51] Arash Amini, Qiyu Sun, and Nader Motee, *Carleman state feedback control design of a class of nonlinear control systems*, IFAC-PapersOnLine **52** (2019), no. 20, 229–234.

[52] ———, *Error bounds for carleman linearization of general nonlinear systems*, 2021 Proceedings of the Conference on Control and its Applications, SIAM, 2021, pp. 1–8.

[53] K. J. Åström and P. Eykhoff, *System identification—a survey*, Automatica **7** (1971), no. 2, 123–162.

[54] Tochukwu Elijah Ogri, Zachary I. Bell, and Rushikesh Kamalapurkar, *State and parameter estimation for affine nonlinear systems*, Proc. IEEE Conf. Decis. Control (Singapore), December 2023, pp. 1517–1522.

[55] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys. **378** (2019), 686–707.

[56] Yihong Zhou, Yanjiao Wang, Fengying Ma, Feng Ding, and Tasawar Hayat, *Parameter estimation for a class of radial basis function-based nonlinear time-series models with moving average noises*, J. Franklin Inst. **358** (2021), no. 4, 2576–2595.

[57] Philippe Bisaillon, Brandon Robinson, Mohammad Khalil, Chris L. Pettit, Dominique Poirel, and Abhijit Sarkar, *Robust Bayesian state and parameter estimation framework for stochastic dynamical systems with combined time-varying and time-invariant parameters*, J. Sound Vib. **575** (2024), 118106.

[58] Tapas Tripura and Souvik Chakraborty, *A sparse Bayesian framework for discovering interpretable nonlinear stochastic dynamical systems with Gaussian white noise*, Mech. Syst. Signal Process. **187** (2023), 109939.

[59] Daniel Gedon, Niklas Wahlström, Thomas B. Schön, and Lennart Ljung, *Deep state space models for nonlinear system identification*, IFAC-PapersOnLine **54** (2021), no. 7, 481–486.

[60] Lawrence J. Alder and Stephen M. Rock, *Adaptive control of a flexible-link robotic manipulator with unknown payload dynamics*, Proc. Amer. Control Conf. (San Francisco, CA, USA), June 1993, p. 467.

[61] Shicong Dai, Taeyoung Lee, and Dennis S. Bernstein, *Adaptive control of a quadrotor UAV transporting a cable-suspended load with unknown mass*, Proc. IEEE Conf. Decis. Control (Los Angeles, CA, USA), December 2014, pp. 6149–6154.

[62] Peter Klanatsky, François Veynandt, and Christian Heschl, *Grey-box model for model predictive control of buildings*, Energy Build. **300** (2023), no. 1, 113624.

[63] Thore Wietzke, Jan Gall, and Knut Graichen, *Occupancy prediction for building energy systems with latent force models*, Energy Build. **307** (2024), 113968.

[64] Z. Morrison, M. Abudia, J. A. Rosenfeld, and R. Kamalapurkar, *Dynamic mode decomposition of control-affine nonlinear systems using discrete control liouville operators*, IEEE Control Systems Letters **8** (2024), 79–84.

[65] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Information Science and Statistics, Springer, New York, 2008. MR 2450103 (2010f:62002)

[66] N. Aronszajn, *Theory of reproducing kernels*, Trans. of the Am. Math. Soc. **68** (1950), no. 3, 337–404.

[67] Joel A. Rosenfeld, Rushikesh Kamalapurkar, Benjamin Russo, and Taylor T. Johnson, *Occupation kernels and densely defined Liouville operators for system identification*, Proc. IEEE Conf. Decis. Control, December 2019, pp. 6455–6460.

[68] Joel A. Rosenfeld and Rushikesh Kamalapurkar, *Dynamic mode decomposition with control liouville operators*, 2024.

[69] George S. Kimeldorf and Grace Wahba, *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Statist. **41** (1970), 495–502. MR 0254999 (40 #8206)

[70] Fedor Zhdanov and Yuri Kalnishkan, *An identity for kernel ridge regression*, Theoretical Computer Science **473** (2013), 157–178.

[71] Fedor Zhdanov and Vladimir Vovk, *Competing with gaussian linear experts*, arXiv:0910.4683, 2009.

[72] Kevin Kochersberger, Kenneth Kroeger, Bryan Krawiec, Eric Brewer, and Thomas Weber, *Post-disaster remote sensing and sampling via an autonomous helicopter*, Journal of Field Robotics **31** (2014), no. 4, 510–521.

[73] Frank Natterer, *The mathematics of computerized tomography*, SIAM, 2001.

[74] Vladimir Vovk, *Kernel ridge regression*, Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, Springer, 2013, pp. 105–116.

[75] H. V. Henderson and S. R. Searle, *On deriving the inverse of a sum of matrices*, SIAM Review **23** (1981), no. 1, 53–60.

[76] Holger Wendland, *Scattered data approximation*, Cambridge University Press, 2004.

[77] Steven Roman, S Axler, and FW Gehring, *Advanced linear algebra*, vol. 3, Springer, 2005.

[78] Joel A. Rosenfeld, Rushikesh Kamalapurkar, and Warren E. Dixon, *The state following approximation method*, IEEE Trans. Neural Netw. Learn. Syst. **30** (2019), no. 6, 1716–1730.

[79] Fedor Zhdanov and Yuri Kalnishkan, *An identity for kernel ridge regression*, Theor. Comput. Sci. **473** (2013), 157–178.

VITA

Moad K. Abudia

Candidate for the Degree of

Doctor of Philosophy

Dissertation: LIFTING TECHNIQUES FOR DATA-DRIVEN IDENTIFICATION OF NONLINEAR SYSTEMS

Major Field: Mechanical Engineering

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Mechanical Engineering at Oklahoma State University, Stillwater, Oklahoma in July, 2026.

Completed the requirements for the Master of Science in Mechanical Engineering at Oklahoma State University, Stillwater, Oklahoma in December, 2020.

Completed the requirements for the Bachelor of Science in Mechanical Engineering at Oklahoma State University, Stillwater, Oklahoma in May, 2016.