March 2023

Data-Driven Learning Algorithm Via Densely-Defined Multiplication Operators and Occupation Kernels.

John Kyei
University of South Florida

Data-Driven Learning Algorithm Via Densely-Defined Multiplication Operators and Occupation Kernels.

by

John Kyei

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a concentration in Pure and Applied Mathematics
Department of Mathematics and Statistics
College of Arts and Sciences
University of South Florida

Co-Major Professor: Razvan Teodorescu, Ph.D.
Co-Major Professor: Joel A. Rosenfeld, Ph.D.
Catherine A. Beneteau, Ph.D.
Dmitry Khavinson, Ph.D.
Benjamin P. Russo, Ph.D.

Date of Approval:
March, 2023

Keywords: System Identification, Dynamical systems, integral transform, kernelized reconstruction

DEDICATION

I would like to dedicate this work to the memory of my mother.

ACKNOWLEDGMENTS

Many people have helped me immensely on this journey and I want to take a moment to acknowledge them. For what do I have that I did not receive? The answer of course is 'nothing'. First and foremost, glory and honor to the most high God for his divine enablement that got me this far. Indeed he makes everything beautiful in its time. Secondly, I would like to express my sincere gratitude to my advisors; Dr. Joel Rosenfeld and Dr. Razvan Teodorescu for their academic and intellectual investments in me. I could not have undertaken this journey without their guidance, support, and teachings. I am deeply indebted and forever grateful to them. A big thank you to Dr. Catherine Beneteau, Dr. Benjamin Russo, and Dr. Dmitry Khavinson for their invaluable patience, feedback, and sacrifices.

I would also like to acknowledge my mates and research teammates; Himanshu Singh, Efrain Gonzales, Nathan Hayford, and Chukwuebuka Amagwula. Our discussions and your reviews impacted and inspired me in numerous ways.

Lastly, none of this could have happened without my family and friends. The impact of their support has been instrumental in the success of this work, especially Juliana Ababio, Sarah Kyei, Ebenezer Eduafo, and Cecilia Afriyie. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Consider a nonautonomous nonlinear evolution $\dot{x} = f(x, t, \mu)$, where the vector $x(t) \in \mathbb{R}^n$ represents the state of the dynamical system at time t, μ contains system parameters, and $f(\cdot)$ represents a dynamic constraint. In most practical applications, the nonlinear dynamic constraint f is unknown analytically. The problem of approximating f directly from data measurements generated by the system is a main goal of this manuscript. In the postulates of the Nonlinear Autoregressive (NAR) framework, we show that the problem of approximating f can be studied through symbols of densely defined multiplication operators over a Reproducing Kernel Hilbert Spaces (RKHS). In this formulation, data is mapped into a RKHS by virtue of occupation kernels which are special functions that reside in a RKHS owing to an integration functional. The resulting scheme is a parameter identification algorithm where system parameters are approximated according to some induced structure on the symbols of the operator. The action of the adjoint multiplication on an occupation kernel induces a kernelized transform which is the subject of study in the second part of the dissertation. The work is concluded by a kernelized reconstruction of a solution to the classical Sturm-Liouville problem.

CHAPTER 1:
INTRODUCTION AND MAIN RESULTS

1.1 Introduction

Dynamical systems are among the most frequently encountered systems in engineering and in the physical sciences. These systems emerge as models for certain physical, chemical or biological states, also, as auxiliary tools for addressing other mathematical problems. While dynamical systems are widely studied, the majority lack explicit mathematical models that describe them. The main goal of this dissertation is to explore data-driven approaches to studying dynamical systems. Arguably, data is increasingly becoming abundant in modern systems of interest such as epidemiological systems, networks of neurons, financial markets, etc, but models are often elusive. This makes discovering governing equations from high-dimensional data a rapidly evolving field with great potential to address the challenge. System identification routines offer a data-centered approach that involves reproducing system input-output measurements. Mostly, the resulting model is applied to capture prevalent dynamical structure; an equally crucial part of the objective [44]. In this write-up, we present a non-linear system identification algorithm by employing *Occupation kernels* and *densely defined multiplication operators*. Occupation kernels are functions that reside in a reproducing kernel Hilbert space (RKHS) that presents an integration functional as an inner product and can be regarded as the generalization of *occupation measures* [50]. Linear models, as we know, are often deficient in their ability to represent complex nonlinear dynamics. Meanwhile, most every day systems are nonlinear to some degree, and this dictates the need for identification techniques specific to nonlinear systems [44]. A few such techniques are variants of the *Nonlinear Autoregressive Moving Average with Exogenous Inputs (NARMAX)*, including *Nonlinear Autoregressive Moving Average (NARMA)*, *Autoregressive Moving Average (ARMA)*, *Nonlinear Autoregressive NAR* etc . The theoretical background of this study is sourced from NARMAX introduced by Billings, S. A. and Leontaritis, I. J [4]. We introduce a kernelized learning algorithm that couples NAR

with symbols of densely defined multiplication to represent data as objects within a reproducing kernel Hilbert Space. The result is a parameter identification routine that approximates unknown system dynamics in a function space based on an induced norm. The norm which is the product of the interaction between the adjoint densely defined multiplication operators and the aforementioned occupation kernel imposes a needed structure on the symbols. Additionally, this interaction yields a kernelized Fourier transform which is the subject of study in the second half of the dissertation. The developed method shares close connections with Dynamic Mode Decomposition introduced by P. J. Schmid [58] and the Sparse Identification for Nonlinear Dynamics (SINDy) by S. L. Brunton, J. L. Proctor and J. N. Kutz [8]. In Dynamic Mode Decomposition, fundamental governing principles of continuous time systems are analyzed by discrete time proxies using high dimensional time series data. The algorithm identifies the best-fit linear model that advances high dimensional measurements forward in time. This is achieved by constructing an approximate locally linear dynamical system $\frac{d\mathbf{X}}{dt} = \mathcal{A}X$. Practically, a linear system is realized by describing an analogous discrete-time system sampled every Δt in time so that $\mathbf{X}_{k+1} = \mathbf{AX_k}$ where $\mathbf{A} = \exp(\mathcal{A}\Delta t)$ and $\mathcal{A}$ refers to the matrix in the continuous time dynamics. DMD produces a low-rank eigen decomposition of the matrix $\mathbf{A}$ that specially fits the measured trajectory by minimizing the deviation $\|\mathbf{X}_{k+1} - \mathbf{AX} + k\|_2$.

1.2 Main Results

The contributions of this work are in two folds. First, we develop an algorithm for estimating system parameters from data for approximating unknown dynamics. The method, which stems from the NAR architecture, incorporates past observations into predicting present values of the system. The algorithm is suitable for dynamics of the form $\gamma(t) = f(\gamma(t-\tau_1), \gamma(t-\tau_2), \cdots, \gamma(t-\tau_p))$, where $\tau_i, i = 1, 2, \cdots, p$ are time lags. The second part of the work introduces a kernelized approximation for the Fredholm integral operator. The framework is a formulation of a solution to the Sturm-Liouville differential equation subject to some boundary conditions. Here, solution to the Sturm-Liouville problem is sequentially reconstructed through through the kernel function associated to a Fredholm integral operator.

1.3 Organization of the Dissertation

The dissertation is organized as follows: Chapter 2 presents a background of relevant literature on system identification. Specifically, we look at data-driven modalities like the Dynamic Mode Decomposition (DMD) and Sparse Identification for Nonlinear Dynamics (SINDy) algorithms. Some topological and operator theoretic properties of the multiplication operator and its connection to the Nevanlinna-Pick interpolation problem are discussed. We conclude the chapter with a review on the Nonlinear Autoregressive Moving Average Model with Exogenous input (NARMAX). The main contribution of this work is presented in Chapters 3, 4 and 5. The third chapter introduces the notion of occupation kernels and formulates a system identification model. Numerical experiments are carried out to verify the methodology discussed. Chapter 4 highlights analytic properties of the interaction between the multiplication operator and the occupation kernel, and its role in the approximation process. Finally, Chapter 5 introduces a construction of a solution to the Sturm-Liouville problem through a kernelized approximation of the Fredholm integral operator.

CHAPTER 2:
PRELIMINARIES

2.1 Introduction

In this chapter, we present relevant background in the form of literature, fundamental theory, and definitions of key terms. We shall assume throughout this manuscript that all Hilbert spaces are separable. A Hilbert space $\mathcal{H}$ and its associated inner product and norm shall be denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ respectively. Hilbert function spaces shall be those comprising real-valued functions unless otherwise stated.

2.2 Reproducing Kernel Hilbert Space (RKHS)

Before we begin, let's specify some notations. We shall denote by $\mathbb{D}$, the unit disk in the complex plane; $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. The unit circle shall be denoted $\mathbb{T}$, representing the collection $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

A Hilbert Space is an inner product space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ that is complete with respect to the norm by the inner product. Suppose $\mathcal{X}$ is a topological space, a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ over the set $\mathcal{X}$ is a Hilbert space of real-valued functions such that for all $x \in \mathcal{X}$ the point evaluation functional $E_x g := g(x)$ is well-defined and bounded. The Riesz representation theorem states that for all $x \in \mathcal{X}$, there exists a function $k_x \in \mathcal{H}$ such that $\langle g, k_x \rangle_{\mathcal{H}} = g(x)$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on $\mathcal{H}$. The function k_x is said to have the reproducing property and is called the reproducing kernel function at x. Take for example the space of square integrable functions over the real numbers denoted $L^2(\mathbb{R})$ equipped with the inner product $\langle f, g \rangle = \int_{\mathbb{R}} f(s)g(s)ds$. $L^2(\mathbb{R})$ is a Hilbert space but not a RKHS as its evaluation functional is not bounded.

For any RKHS $\mathcal{H}$, the span of the reproducing kernel $k := span\{k_x : x \in \mathcal{X}\}$ is dense in $\mathcal{H}$. Equivalently, we say k $spans$ $\mathcal{H}$ with

$$\mathcal{H} = \overline{span\{k_x : x \in \mathcal{X}\}},$$

and the function $K(x, y) = \langle k_y, k_x \rangle_{\mathcal{H}}$ is a kernel function for the RKHS. An important property of reproducing kernels are that, the corresponding RKHS is unique for any symmetric kernel, a result known as the *Moore-Aronszajn theorem*. For any $\mathcal{X}$, the bijective correspondence between reproducing kernel Hilbert spaces and positive definite reproducing kernels $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ was first discovered by E.H. Moore [42]; also discussed by Aronszajn in [2]. Recall that a kernel k is said to be symmetric if for all pair of elements $x, y \in \mathcal{X}$, $k_y(x) = k(x, y) = k(y, x) = k_x(y)$.

Definition 2.2.1. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called positive semi-definite if for all $n \in \mathbb{N}, \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and all $x_1, \ldots, x_n \in \mathcal{X}$, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k\left(x_j, x_i\right) \geq 0. \tag{2.1}$$

Furthermore, k is said to be strictly positive definite if for mutually distinct points $x_1, \ldots, x_n \in \mathcal{X}$, equality in (2.1) only holds for $\alpha_1 = \cdots = \alpha_n = 0$. Lastly, k is called symmetric if $k\left(x, x'\right) = k\left(x', x\right)$ for all $x, x' \in \mathcal{X}$.

For fixed $x_1, x_2, \ldots, x_n \in \mathcal{X}$, the matrix

$$K = [k(x_i, x_j)]_{i,j=1}^{n} \tag{2.2}$$

is called the *Gram matrix* of k. Satisfying condition (2.1) is equivalent to saying the Gram matrix is positive semi-definite.

Symmetry and positive semi-definitness are in fact necessary and sufficient conditions for a function to be a kernel. Like most *kernel methods*, at the center of our current algorithm are kernel functions. We present a brief overview of kernels in the next few paragraphs. For a detailed treatment of the topic, refer to *Support Vector Machines by I. Steinwart, A. Christmann* [61].

Definition 2.2.2. Let $\mathcal{X}$ be a nonempty set. The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{K}(\mathbb{R} \text{ or } \mathbb{C})$ is called a kernel on $\mathcal{X}$ if there is a $\mathbb{K}$-Hilbert space $\mathcal{H}$ and a map $\Phi : \mathcal{X} \to \mathcal{H}$ such that for all $x \in \mathcal{X}$, we have

$$k(x, x') = \left\langle \Phi(x'), \Phi(x) \right\rangle . \tag{2.3}$$

Φ is referred to as a feature map and $\mathcal{H}$ a feature space of k.

A set of kernels does not necessarily form a linear space as the difference of two kernel is not a kernel, however, new kernels can be formed from various combinations of existing ones.

Lemma 2.2.1. *Let $\mathcal{X}, \mathcal{X}_1,$ and $\mathcal{X}_2$ be sets, and $\alpha \geq 0$, then*

1. *If k, k_1 and k_2 are kernels on $\mathcal{X}$ then αk and $k_1 + k_2$ are kernels on $\mathcal{X}$.*
2. *If k_1, k_2 are kernels on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, then $k_1(x_1, x_1') \cdot k_2(x_2, x_2')$ is a kernel on $\mathcal{X}_1 \times \mathcal{X}_2$, for $x_1, x_1' \in \mathcal{X}_1$ and $x_2, x_2' \in \mathcal{X}_2$.*

Below are two examples of kernel functions over $\mathbb{C}^d$ and $\mathbb{R}^d$ [61, Chapter 4] associated with a positive constant $\mu > 0$.

Polynomial kernels: For $d \geq 1$ and $c \geq 0$, $k(z, z') = (\langle z, z' \rangle + c)^m$ for $z, z' \in \mathbb{C}$.

Gaussian Radial Basis (RBF): For $d \in \mathbb{N}$, $z = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^d$ and $z' = (z_1', z_2', \ldots, z_n') \in \mathbb{C}^d$, define

$$k_{\mu,\mathbb{C}}(z, z') = \exp\left(-\frac{1}{\mu^2} \sum_{j=1}^{d} (z - z')^2 \right). \tag{2.4}$$

$k_{\mu\mathbb{C}}$ is a kernel on $\mathbb{C}^d$. Its restriction $k_\mu = k_{\mu,\mathbb{C}}|_{\mathbb{R}^d \times \mathbb{R}^d}$ is

$$k_\mu(x, x') = \exp\left(\frac{1}{\mu^2} \|x - x'\|_2^2 \right). \tag{2.5}$$

k_μ is called the *Gaussian Radial Basis kernel* and the parameter μ is called the *kernel width*.

Suppose $\mathcal{H}$ is a $\mathbb{K}-$ Hilbert function space over some set $\mathcal{X}$ that has a reproducing kernel k, then $\mathcal{H}$ is a RKHS on a feature space of k where the feature map $\Phi : \mathcal{X} \to \mathcal{H}$ is given by

$$\Phi(x) = k(\cdot, x). \tag{2.6}$$

As mentioned earlier, Theorem 2.2.2 states the one-to-one correspondence between reproducing kernels and RKHS.

Theorem 2.2.2. *[61] Let $\mathcal{H}$ be a RKHS over $\mathcal{X}$, then $k : \mathcal{X} \times \mathcal{X} \to \mathbb{K}$ defined by $k(x, x') = \langle \delta_x, \delta_{x'} \rangle_{\mathcal{H}}$, $x, x' \in \mathcal{X}$ is the reproducing kernel of $\mathcal{H}$. Furthermore, if $\{e_i\}_{i \in I}$ is an orthonormal basis of $\mathcal{H}$, then for all $x, x' \in \mathcal{X}$ we have*

$$k(x, x') = \sum_i e_i(x)\overline{e_i(x')}. \tag{2.7}$$

The *Hardy space* $\mathcal{H}^2$ (which is a subspace of the space of all Lebesgue measurable functions that are square-integrable over the unit circle $\mathbb{T}$, denoted $L^2(\mathbb{T})$) is an example of a RKHS. The general Hardy Space $(\mathcal{H}^p)$, for $p \in (0, \infty)$, is given in Definition 2.2.3 below.

Definition 2.2.3. (Hardy Space, [65]) For $0 < p < \infty$, the Hardy space $\mathcal{H}^p$ consists of analytic functions f over the unit disk $\mathbb{D}$ such that

$$\|f\|_p^p = \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} \left| f\left(re^{i\theta}\right) \right|^p d\theta < \infty.$$

When $p = \infty$, we use $\mathcal{H}^\infty$ to denote the space of bounded analytic functions in $\mathbb{D}$ with respect to the supremum norm. Thus

$$\|f\|_\infty = \sup\{|f(z)| : z \in \mathbb{D}\}.$$

The complex Hardy space over the disk can be thought of as the analytic continuation of square integrable functions on the unit circle $(\mathbb{T})$ whose Fourier coefficients $\hat{f}(n) := \int_{-\pi}^{\pi} f(e^{i\theta}) e^{-in\theta} d\theta$ vanishes for all negative values of n.

Consider the case $p = 2$, the Hardy space $\mathcal{H}^2(\mathbb{D})$ is a reproducing kernel Hilbert space with reproducing kernel function

$$k(z, \omega) = \sum_{n \geqslant 0} z^n \bar{\omega}^n = \frac{1}{1 - z\bar{\omega}},$$

commonly referred to as the *Szego kernel*.

Definition 2.2.4. (Mercer Kernel, [38]) Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is called a Mercer kernel if it is a continuous, positive semidefinite function, i.e. a continuous kernel function.

Theorem 2.2.3. *(Mercer's Theorem [38]). Let k be a Mercer kernel on $\mathcal{X}$, let μ be a finite Borel measure with support X, and let $T_k : L^2(\mathcal{X}, \mu) \to L^2(\mathcal{X}, \mu)$ be the associated integral operator. Then there is a countable collection of orthonormal continuous functions $\{e_n\}$ on $\mathcal{X}$ that are eigenvectors for T_K with corresponding eigenvalues $\{\lambda_n\}$ such that for every $g \in L^2(\mathcal{X}, \mu)$ we have*

$$T_k g = \sum_n \lambda_n \langle g, e_n \rangle e_n.$$

Furthermore, $k(x, y) = \sum_n \lambda_n e_n(x)\overline{e_n(y)}$.

Function-valued RKHS [2, 33, 54]

The theory of function-valued RKHS (equivalently referred to as vector-valued RKHS) parallels that of its scalar-valued counterparts. Let $\mathcal{X}$ be a nonempty set, $\mathcal{W}$ a real Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{W}}$. Let $\mathcal{B}(\mathcal{W})$ be the Banach space of bounded linear operators on $\mathcal{W}$ and denote by $\mathcal{W}^{\mathcal{X}}$, the vector space of function $f : \mathcal{X} \to \mathcal{W}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathcal{B}(\mathcal{W})$ is called an *operator-valued* positive definite kernel if for each pair $(x, z) \in \mathcal{X} \times \mathcal{X}$, $k(x, z) \in \mathcal{B}(\mathcal{W})$ is self-adjoint and

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \langle y_i, k(x_i, x_j) y_j \rangle_{\mathcal{W}} \geq 0 \tag{2.8}$$

for every finite set $\{x_i\}_{i=1}^N \subset \mathcal{X}$ and $\{y_i\}_{i=1}^N \subset \mathcal{W}$. Given $x \in \mathcal{X}$ and $w \in \mathcal{W}$, define the function $k_x w = k(\cdot, x)w \in \mathcal{W}^{\mathcal{X}}$ by

$$(k_x w)(z) = k(z, x)w \quad \text{for all} \quad z \in \mathcal{X}. \tag{2.9}$$

Then there is a unique $\mathcal{W}-$valued RKHS $\mathcal{H}_k$ which admits k as a reproducing kernel. Suppose we set $\mathcal{H}_0 = span\{k_x : x \in \mathcal{X}, w \in \mathcal{W}\}$, for $f = \sum_{i=1}^{N} k_{x_i} w_i$ and $g = \sum_{i=1}^{N} k_{z_i} y_i \in \mathcal{H}_0$, define the inner product

$$\langle f, g \rangle := \sum_{i=1}^{N} \sum_{j=1}^{N} \langle w_i, k(x_i, z_i) y_i \rangle_{\mathcal{W}}. \tag{2.10}$$

Completing $\mathcal{H}_0$ gives a Hilbert space $\mathcal{H}_k$ whose kernel has the reproducing property

$$\langle f(x), w \rangle_{\mathcal{W}} = \langle f, k_x w \rangle_{\mathcal{H}_k}. \tag{2.11}$$

A vector-valued RKHS $\mathcal{H}$ is a Hilbert space of vector-valued functions over a set $\mathcal{X}$ such that every point evaluation mapping $E_x : \mathcal{H} \to \mathbb{R}$, given by $E_x g := g(x)$, $\forall x \in \mathcal{X}$, is bounded. That is, for all $x \in \mathcal{X}$, there is a positive constant C_x such that

$$\|f(x)\| \leq C_x \|f\|_{\mathcal{H}}, \quad \text{for all} \quad f \in \mathcal{H}.$$

Therefore the reproducing kernel associated with $\mathcal{H}_k$ is defined to be the map $K : \mathcal{X} \times \mathcal{X} \to \mathcal{B}(\mathbb{V})$ such that $K(x, y) = E_x E_y^*$.

2.3 Multiplication Operator on a Reproducing Kernel Hilbert Space

In what follows, we look at some topological properties of the multiplication operator on a RKHS. Multiplication operators are well-studied objects in operator theory and functional analysis in general. Classiscal Analysis problems such as the *Nevanlinna-Pick Interpolation problem* use the computation of norms of the operator over a Hilbert spaces. Sarason demonstrates this in the Generalized Interpolation Theorem [56], where he formulates an operator theoretic solution to an interpolation problem. A brief overview of the problem is given below with emphasis on formulations that involve the multiplication operator.

Definition 2.3.1. For any measurable function φ on $\mathcal{X}$, the multiplication operator with symbol φ denoted by M_φ is defined on the RKHS $\mathcal{H}(\mathcal{X})$ by

$$M_\varphi [g] (x) = \varphi(x) g(x), \quad \text{for} \quad x \in \mathcal{X}. \tag{2.12}$$

M_φ is said to be bounded if there is a constant C such that the operator norm $M_\varphi f \leq C\|f\|$. If φ is a symbol of bounded multiplication operator, the $M_\varphi k_x = \overline{\varphi(x)} k_x$ for all $x \in \mathcal{X}$. The conjugation is dropped when the function $\varphi(x)$ is real-valued.

At times, practical applications dictates we pre-select a range space for the multiplication operator, in which case specific conditions may be imposed on the classes of allowable symbols and/or domain. For instance, we may define the multiplication operator between the two function spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$, $M_g : \mathcal{H} \to \tilde{\mathcal{H}}$, such that $M_g f \in \tilde{\mathcal{H}}$. The domain of $D(M_g)$ therefore comprises of all those functions $\{f \in \mathcal{H} : fg \in \tilde{\mathcal{H}}\}$.

Proposition 2.3.1. *Let L be a continuous linear operator on a RKHS $\mathcal{H}$ over the set $\mathcal{X}$. Suppose for the complex-valued function φ, $Lk_x = \overline{\varphi(x)}k_x$ for each $x \in \mathcal{X}$, then $L^* = M_\varphi$. That is, L^* is given by multiplication by φ.*

Proof. Let $f \in \mathcal{H}$, then

$$\begin{aligned}\langle L^* f, k_x\rangle &= \left\langle f, \overline{\varphi(x)}k_x\right\rangle \\ &= \varphi(x)\left\langle f, k_x\right\rangle \\ &= \varphi(x)f(x) \\ &= \left\langle \varphi f, k_x\right\rangle.\end{aligned}$$

Since the span of the set $\{k_x : x \in X\}$ is dense in $\mathcal{H}$, we have

$$\langle L^* f, g\rangle = \langle \varphi f, g\rangle$$

for each $g \in \mathcal{H}$, so that $L^* = M_\varphi$. □

Theorem 2.3.2. *[6] Let $\mathcal{X}$ be a Banach space of analytic functions on which point evaluation is bounded for each $z \in \mathbb{D}$. Suppose M_g is bounded on $\mathcal{X}$ for some $g \in \mathcal{X}$. Then*

$$|g(z)| \leq \|M_g\|. \tag{2.13}$$

Multipliers

For a given collection of functions, say a Hilbert function space $\mathcal{H}$ on the set $\mathcal{X}$, the multipliers of $\mathcal{H}$ is the collection of all those functions φ on $\mathcal{X}$ that multiply $\mathcal{H}$ into itself, that is, the set

$$\{\varphi : \varphi f \in \mathcal{H}, \forall f \in \mathcal{H}\}.$$

From Closed Graph Theorem, if φ is a multiplier of $\mathcal{H}$, then the multiplication operator with symbol φ, M_φ, is a bounded linear operator on $\mathcal{H}$.

We shall denote by $mult\,(\mathcal{H})$ the multipliers of $\mathcal{H}$. If there is a known structure on $\mathcal{H}$, such as the Hardy space, more can be said about M_φ and $mult(\mathcal{H})$, we may even be able characterize $mult(\mathcal{H})$. For instance, $\mathcal{H}^\infty$, the space of all bounded analytic functions on the disk is the multiplier algebra for the Hardy space $\mathcal{H}^2$. There is a thorough discussion of this by J. Angler and J. E. McCarthy in

[1]. It is worth noting that the Hardy space $\mathcal{H}^2$ is not the only space on the disk that has $\mathcal{H}^\infty$ as a multiplier algebra. The Bergman space A^2_a, comprising of those holomorphic functions on $\mathbb{D}$ that are square-summable with respect to the planar Lebesgue measure also has $\mathcal{H}^\infty$ as its multiplier algebra.

Definition 2.3.2. (Densely Defined and closed operators) Let $\mathcal{H}$ be a Hilbert space and $T : \mathcal{H} \to \mathcal{H}$ be a linear operator, and let the domain $D(T)$ be properly contained in $\mathcal{H}$. Then

1. T is called densely defined if $D(T)$ is a dense subspace of $\mathcal{H}$.

2. T is closed if for every sequence $\{g_n\} \subset D(T)$ such that $g_n \to g \in \mathcal{H}$ with $Tg_n \to h \in \mathcal{H}$, then $g \in D(T)$ and $Tg = h$.

Theorem 2.3.3. *[1] Suppose $\mathcal{H}$ is a RKHS on $\mathcal{X}$, and let $mult(\mathcal{H})$ be the corresponding multiplier algebra of $\mathcal{H}$. Then $mult(\mathcal{H})$ is closed in the operator norm topology.*

Proof. Let $\{M_{\varphi_n}\}_{n=1}^\infty$ be a Cauchy sequence in the space of bounded linear operators on $\mathcal{H}$, $(\mathcal{B}(\mathcal{H}))$. Then so is $\left\{M^*_{\varphi_n}\right\}_{n=1}^\infty$. Now, since $\mathcal{B}(\mathcal{H})$ is complete, there is a $T \in B(\mathcal{H})$ such that $M^*_{\varphi_n} \to T$ in operator norm . It follows from Theorem 2.3.7 that

$$\left\|M^*_{\varphi_m} - M^*_{\varphi_n}\right\| \geqslant |\varphi_m(x) - \varphi_n(x)|$$

for each $x \in \mathcal{X}$, so that the sequence $\{\varphi_n\}_{n=1}^\infty$ is Cauchy and converges for each x.

Define $\varphi(x) = \lim_{n\to\infty} \varphi_n(x)$, then

$$\begin{aligned} Tk_x &= \lim_{n\to\infty} M^*_{\varphi_n} k_x \\ &= \lim_{n\to\infty} \overline{\varphi_n(x)} K_x \\ &= \overline{\varphi(x)} k_x. \end{aligned}$$

That is, $T^* = \lim_{n\to\infty} M_{\varphi_n}$. See Proposition 2.3.1. □

Theorem 2.3.4. *[1] The algebra $\mathcal{H}^\infty$ is isometrically isomorphic and weak-* homeomorphic to $mult\left(\mathcal{H}^2\right)$.*

Proof. We will not discuss here the weak-* homeomorphism property since it is not as relevant to the topic at hand, but refer the reader to [1].

Consider $\mathcal{H}^2(\mathbb{D})$ as a subspace of $L^2(\mathbb{T}, d\mu)$ and $H^\infty(\mathbb{D}) \subset L^\infty(\mathbb{T}, d\mu)$.

Then for any $\varphi \in \mathcal{H}^\infty$, and for any $f \in \mathcal{H}^2$, we have that $\varphi f \in \mathcal{H}^2$. That is,

$$\begin{aligned}\|\varphi f\|_{\mathcal{H}^2}^2 &= \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |(\varphi f)(re^{i\theta})|^2 d\theta \\ &= \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |\varphi(re^{i\theta})|^2 |f(re^{i\theta})|^2 d\theta \\ &\leq \|\varphi\|_\infty^2 \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta = \|\varphi\|_\infty^2 \|f\|_{\mathcal{H}^2}^2.\end{aligned} \tag{2.14}$$

Thus $\|\varphi f\|_{\mathcal{H}^2} \leq \|\varphi\|_\infty \|f\|_{\mathcal{H}^2} < \infty$. Therefore $\varphi f \in \mathcal{H}^2$, thus $\varphi \in mult(\mathcal{H}^2)$ and $\mathcal{H}^\infty \subset mult(\mathcal{H}^2)$. Since $M_{\varphi f} = \varphi f$, this shows that

$$\|M_\varphi\| \leq \|\varphi\|_\infty . \tag{2.15}$$

Similarly, if $\varphi \in mult\left(\mathcal{H}^2\right)$ that is $\varphi f \in \mathcal{H}^2 \quad \forall f \in \mathcal{H}$, since the constant $1 \in \mathcal{H}^2$, and because $\varphi = \varphi \cdot 1 \in \mathcal{H}^2$, it means that $\varphi \in \mathcal{H}^2$ and is therefore analytic. Moreover, for any $\varphi \in \mathcal{H}^\infty$, M_φ is bounded, which can be shown using arguments from the Closed Graph Theorem and continuity of point evaluation. But $\|M_\varphi\| \geqslant \sup_{\lambda \in \mathcal{X}} |\varphi(\lambda)|$, so φ must be bounded. Therefore $\varphi \in \mathcal{H}^\infty(\mathbb{D})$ and hence $mult(\mathcal{H}^2) \subset \mathcal{H}^\infty(\mathbb{D})$.

However, from $M_\varphi^* k_\lambda = \overline{\varphi(\lambda)} k_\lambda$, we have the inequality:

$$\|M_\varphi\| = \left\|M_\varphi^*\right\| \geqslant \|\varphi\|_{\mathcal{H}^\infty}.$$

Therefore

$$\|M_\varphi\| \geqslant \|\varphi\|_{\mathcal{H}^\infty} \tag{2.16}$$

for every λ in $\mathbb{D}$. Combining (2.15) and (2.16), we get $\|M_\varphi\| = \|\varphi\|_{\mathcal{H}^\infty(\mathbb{D})}$. This concludes the proof that $\mathcal{H}^\infty(\mathbb{D})$ and the multiplier algebra of $\mathcal{H}^2$ are isomorphic. □

To generalize the conditions under which these norms coincide, we state the following proposition.

Proposition 2.3.5. *[63] Let $1 \leq p < \infty$. Then the following statements are equivalent.*

1. $\varphi \mathcal{H}^p \subset \mathcal{H}^p$.

2. *φ is analytic in $\mathbb{D}$ and $M_\varphi : \mathcal{H}^p \to \mathcal{H}^p$ is bounded.*

3. *$\varphi \in \mathcal{H}^\infty$.*

If any (and therefore each) of the above conditions is fulfilled, then the norm of the operator M_φ is

$$\|M_\varphi\| = \|\varphi\|_{\mathcal{H}^\infty}.$$

Proposition 2.3.6. *For all $\varphi \in \mathcal{H}^\infty$, the multiplication operator $M_\varphi : \mathcal{H}^2 \to \mathcal{H}^2$ is densely defined and closed with*

$$D(M_\varphi) = \mathcal{H}^2.$$

Proof. This is a direct consequence of Theorem 2.3.4. Recall that for $M_\varphi : \mathcal{H}^2 \to \mathcal{H}^2$

$$D(M_\varphi) = \{f \in \mathcal{H}^2 : M_\varphi f \in \mathcal{H}^2\}. \tag{2.17}$$

That is, if $\varphi \in \mathcal{H}^\infty$ then $\varphi \in \operatorname{mult}\left(\mathcal{H}^2\right)$ that is $\varphi f \in H^2$ for all $f \in \mathcal{H}^2$. Hence $\mathcal{H}^\infty(\mathbb{D}) \subset D(M_\varphi)$. But $M_\varphi : D(M_\varphi) \subset \mathcal{H}^2 \to \mathcal{H}^2$, implying equality. Consequently, the domain $D(M_\varphi)$ is dense in $\mathcal{H}^\infty$. While closedness was shown above. □

Recall that if an operator is closed and densely defined, then it admits a well-defined adjoint which is also closed and densely defined. Therefore, for $\varphi \in \mathcal{H}^\infty$, M_φ^* is closed and densely defined.

Theorem 2.3.7. *Let $\mathcal{H}$ be a RKHS on $\mathcal{X}$ and let φ be a multiplier on $\mathcal{H}$. Then the multiplication operator M_φ acting on $\mathcal{H}$ is continuous (bounded).*

Proof. Here, it is enough to verify the hypothesis of the *Closed Graph Theorem*. Suppose that $\{f_n\}_{n=1} \subset \mathcal{H}$. Such that f_n converges uniformly in norm to $f \in \mathcal{H}$. Suppose also that $M_\varphi f_n$ converges to g with $g \in \mathcal{H}$. We wish to demonstrate that $M_\varphi f = g$. From the *Cauchy-Schwarz inequality*, we have that norm convergence implies pointwise convergence, hence $f_n \to f$ and $(M_\varphi f_n)(z) = \varphi(z) f_n(z) \to g(z)$ for each $z \in \mathbb{C}$; thus $\varphi(z) f(z) = g(z)$, which implies $M_\varphi f = g$.
As an immediate consequence of this, notice that if M_φ is continuous, then it has a well-defined adjoint M_φ^*. By definition, the domain of the the adjoint operator comprises all functions $\psi \in \mathcal{H}$ for which the functional $L(f) = \langle M_\varphi f, \psi \rangle$ is continuous. Suppose $k_x \in D\left(M_\varphi^*\right)$. Then

$\left\langle f, M_\varphi^* k_x \right\rangle = \langle M_\varphi f, k_x \rangle = \langle (\varphi f)(x), k_x \rangle = \varphi(x) \langle f, k_x \rangle = \left\langle f, \overline{\varphi(x)} k_x \right\rangle$, therefore $M_\varphi^* k_x = \overline{\varphi(x)} k_x$
This holds for all $f \in D(M_\varphi)$. □

Theorem 2.3.8. *If T is a densely defined operator on $\mathcal{H}$, then T^* is a closed operator.*

Proof. The double adjoint relationship $(T^*)^* = T$ is enough justification of this fact. □

Self-adjoint and compact operators are well-studied with extensive literature. A self-adjoint operator T which includes requirements that $D(T^*) = D(T)$, has real-valued eigenvalues and are hermitian and symmetric. A form of the spectral theory of compact, self-adjoint operators involving the multiplication operator is summarized in Proposition 2.3.9. See [17]

Proposition 2.3.9. *[17] Suppose $T \in \mathcal{B}(\mathcal{H})$ is self-adjoint. Then there exists a σ-finite measure space (X, μ), a bounded, measurable, real-valued function h on $\mathcal{X}$, and a unitary map $U : \mathcal{H} \to L^2(X, \mu)$ such that*

$$\left[UTU^{-1}(\psi)\right](\lambda) = h(\lambda)\psi(\lambda)$$

for all $\psi \in L^2(X, \mu)$.

2.4 Operator Theoretic Approaches to the Nevanlinna-Pick Interpolation Problem

The *Nevanlinna-Pick interpolation problem* was first studied by Pick around 1916 [40] and separately by Nevanlinna in 1919 [36]. Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ be the unit disk in the complex plane $\mathbb{C}$. The original Nevanlinna-Pick problem is stated as follows. Given n distinct points $\lambda_1, \lambda_2, \ldots, \lambda_n$ in the open unit disk $\mathbb{D}$, and n complex numbers $w_1, w_e, \ldots, w_n$ in $\mathbb{D}$, is there an analytic function $\phi : \mathbb{D} \to \mathbb{D}$ that interpolates the given data? i.e.

$$\phi(\lambda_i) = w_i, \text{ for } i = 1, 2, \cdots n \tag{2.18}$$

Pick's existence requirement relied on the *Schwarz lemma* and was that such a function ϕ exists if and only if the matrix

$$P = \left[\frac{1 - w_i \overline{w_j}}{1 - \lambda_i \overline{\lambda_j}}\right]_{i,j=1}^{n} \tag{2.19}$$

now known as the *Pick Matrix*, is positive semi-definite. Pick further established that the interpolating function is unique if and only if $rank(P) < n$ (P is singular) in which case ϕ is a *finite Blaschke product* of degree $m = rank(P)$.

Recall that, for a finite sequence $z_1, z_2, \ldots, z_n$ in $\mathbb{D}$ and $\gamma \in \mathbb{T}$, the function

$$B(z) = \gamma \prod_{k=1}^{n} \frac{z - z_k}{1 - \overline{z_k} z} \tag{2.20}$$

is a finite Blaschke product [14, chapter 3].

Nevanlinna approached the problem differently with a idea based on Schur's algorithm. Nevalinna provided a unified parameterization for all such solutions of the problem in the case where the Pick matrix is invertible. Other formulations have since been recorded, including operator theoretic techniques pioneered by D. Sarason, whose methodology is centered around the multiplication operator [55, 56]. To apply operator theoretic techniques, a reformulation of the problem involved the use of the Maximum Modulus Principle to show that such interpolating functions come in the form of elements $f \in \mathcal{H}^\infty$ such that

$$\|f\|_\infty \leq 1 \quad \text{and} \quad f(\lambda_i) = w_i \quad \text{for } i = 1, 2, \cdots, n \tag{2.21}$$

Todiscuss the approach, let $\mathcal{H}$ be a Hilbert space over some set $\mathcal{X}$ with kernel function k. If f is an element of the multiplier algebra of $\mathcal{H}$, then by application of the Closed Graph Theorem, we have that $M_f : \mathcal{H} \to \mathcal{H}$ is a bounded operator on $\mathcal{H}$. Also, given $x \in \mathcal{X}$, we have shown that $M_f^* k_x = \overline{f(x)} k_x$, where k_x is the kernel function at x. This indicates that k_x is an eigenvector of M_f^* and f is bounded with $\|f\|_\infty \leq \|M_f\|$

Suppose such an interpolating function ϕ exists, then for each i,

$$M_\phi^* k_{\lambda_i} = \overline{\phi(\lambda_i)} k_{\lambda_i}, \tag{2.22}$$

implying that for any such ϕ, the action of M_ϕ^* on $m = span\{k_{\lambda_i}\}_i$ is completely determined. Pick's theorem may then be interpreted as saying if the norm of M_ϕ^* restricted to m is less or equal to 1, then there is some interpolating function ϕ such that $\|M_\phi\| \leq 1$ on all of $\mathcal{H}$, and Pick's problem has an affirmative answer [1].

Thinking of $\mathcal{H}^\infty$ as a multiplier algebra of a holomorphic space, the Pick problem can be stated in general as: *Let $\mathcal{H}$ be a Hilbert function space on a set $\mathcal{X}$, let $\lambda_1, \ldots, \lambda_N$ be points of $\mathcal{X}$, and let $w_1, \ldots, w_N$ be complex numbers. When does there exist a multiplier ϕ of $\mathcal{H}$ of norm at most one that interpolates each λ_i to w_i?* A statement of the necessary conditions are given in the following theorem

Theorem 2.4.1. *[1] Let $\mathcal{H}$ be a Hilbert function space on a set X, let $\lambda_1, \ldots, \lambda_N$ be points of X, and let $w_1, \ldots, w_n$ be complex numbers. A necessary condition to be able to solve the interpolation problem*

$$\phi : \lambda_i \mapsto w_i \tag{2.23}$$

with a function ϕ in $mult(\mathcal{H})$ of norm at most one, is that the Pick matrix

$$\left[\left(1 - w_i \bar{w}_j\right) k\left(\lambda_i, \lambda_j\right)\right]_{i,j} \tag{2.24}$$

be positive semi-definite, where k is the kernel function of $\mathcal{H}$.

Suppose ϕ exists and satisfies (2.23). Then M_ϕ is a contraction. Recalling that an operator T is a contraction if and only if $I - TT^* \geq 0$, we get that

$$I - M_\phi M_\phi^* \geq 0. \tag{2.25}$$

Equivalently, $\left\langle \left(I - M_\phi M_\phi^*\right) v, v \right\rangle \geq 0$ for any v, in particular, this must hold for any $v \in m = span\{k_{\lambda_i}\}$. Choose $v = \sum_{i=1}^n a_i k_{\lambda_i}$ then

$$\left\langle \left(I - M_\phi M_\phi^*\right) \sum_{j=1}^n a_j k_{\lambda_j}, \sum_{i=1}^n a_i k_{\lambda_i} \right\rangle \geq 0. \tag{2.26}$$

Noting that $\phi(\lambda_i) = w_i$ and $M_\phi^* k_{\lambda_i} = \overline{\phi(\lambda_i)} k_{\lambda_i}$, therefore we get

$$\sum_{i,j=1}^n a_j \bar{a}_i \left(1 - w_i \bar{w}_j\right) \left\langle k_{\lambda_j}, k_{\lambda_i} \right\rangle \tag{2.27a}$$

$$= \sum_{i,j=1}^n a_j \bar{a}_i \left(1 - w_i \bar{w}_j\right) k\left(\lambda_i, k_{\lambda_j}\right) \geq 0. \tag{2.27b}$$

Considering that $\{a_i\}_i$ are arbitrary, the condition in (2.27b) says precisely that the Pick matrix is positive semi-definite.

Notice that the Hilbert space $\mathcal{H}$ may be decomposed as $\mathcal{H} = m \oplus m^{\perp}$, where $m^{\perp}$ is comprised of functions that vanish at $\lambda_1, \lambda_2, \ldots, \lambda_n$. If we let P be an orthogonal projection onto m, then $PM_{\phi}|_m$ is an operator on the finite dimensional subspace m that depend only on $\phi(\lambda_i), \ldots, \phi(\lambda_n)$. Raghupathi in his dissertation [43] formulated the following problem. Suppose we denote by I the ideal of functions that vanish at the n points $\lambda_i, \lambda_2, \ldots, \lambda_n$, and $\mathcal{A} \subset \mathcal{H}^{\infty}$ algebra. Then the solution to the interpolation problem is a function $f \in \mathcal{A}$ such that $\|f\|_{\infty} \leq 1$. Therefore the solution is an interpolating function that resides in a unit closed ball of $\mathcal{A}$.

If we define the map $\pi_m : \mathcal{A} \subset \mathcal{H}^{\infty} \to \mathcal{B}(m)$ given by $\pi_m(f) = PM_fP$, the map π_m is a homomorphism and the kernel of this homomorphism is the set $\{f \in \mathcal{A} : M_f^* k_{\lambda_i} = \overline{f(\lambda_i)} k_{\lambda_i} = 0 \quad \text{for} \quad i = 1, 2, \ldots, n\}$. If $\{k_{\lambda_i}\}$ are nonzero, then $f(\lambda_i) = 0$. The kernel of the homomorphism is therefore I. This gives a contractive, unital representation of $\mathcal{A}/I$ on $\mathcal{B}(m)$. It can be observed that $\|PM_fP\| \leq C$ if and only if $C^2 I - (PM_fP)(PM_fP)^* \geq 0$. Equivalently, $C^2 P - PM_f M_f^* P \geq 0$. An element $k \in m$ has the form $k = \sum_{j=1}^{n} \alpha_j k_{\lambda_j}$, where $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$. Using the fact that $M_f^* k_{\lambda_j} = \overline{f(\lambda_j)} k_{\lambda_j}$, it follows that

$$\begin{aligned} C^2\|k\|^2 - \left\langle M_f M_f^* k, k \right\rangle &= \sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} \left(C^2 - f(x_i)\,\overline{f(x_j)} \right) K(x_i, x_j) \\ &= \left\langle \left[\left(C^2 - f(x_i)\,\overline{f(x_j)} \right) K(x_i, x_j) \right] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \right\rangle. \end{aligned} \tag{2.28}$$

indicating that $\|\pi_m(f)\| \leq 1$ if and only if $\left[\left(1 - f(\lambda_i)\overline{f(\lambda_j)} \right) k(\lambda_i, \lambda_j) \right]_{i,j}^{n}$.

Again, we see a reformulation of Pick's condition as an intrinsic property of the multiplication operator with symbols in $\mathcal{H}^{\infty}$.

2.5 Dynamic Mode Decomposition Approach to System Identification

Dynamic Mode Decomposition (DMD) highlights the connection between Koopman operators, dimensionality reduction algorithms and linear models. DMD is a matrix decomposition technique that allows for simulation of governing equations of a dynamical system directly from data measurements. The method was developed originally by Schmid and Sesterhenn [26, 58] as a way to decompose complex fluid flows into simple representations in time and space variables. Mezić et al [30, 31, 32] later showed that there is a connection between DMD and the underlying nonlinear dynamics through *Koopman operator*, and is consistent with dynamical system techniques. The DMD algorithm has been applied to data originating from dynamical systems in various disciplines; in epidemiology, DMD modes has been used to gain insight into infectious disease data such as inteprating large-scale dynamic patterns of infectious disease spread [26, 41]. DMD computes *Dynamic modes* from dimensionally-reduced neural recordings to extract meaningful spatial patterns. In [7], the authors couple DMD with unsupervised clustering to uncover distinct sleep spindle networks during sleep from recording of dynamic brain activity. In what follows, we give a brief overview of the mathematical structure of the Koopman-based DMD. Our discussion here is mainly sourced from the book *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* [26] by Kutz, J Nathan, Brunton, Steven L, Brunton, Bingni W, and Proctor, Joshua L. For an extensive treatment of the subject, we refer the reader to this material.

2.6 Koopman DMD

Given the dynamical system

$$\frac{d\mathbf{X}}{dt} = f(\mathbf{X}, t, \mu), \tag{2.29}$$

suppose time series data from system (2.29) are collected at discrete instances as $\mathbf{X}_k$, $k = 1, 2, 3 \ldots, m-1$. The DMD algorithm analyzes the relationship between data pairs consisting of past measurements $\mathbf{X}_k$ and future measurement $\mathbf{X}_{k+1}$, where $\mathbf{X} \in \mathbb{R}^n$. For all such pairs of data, it is assumed that there is a linear operator $\mathbf{A} \in \mathbb{R}^{n \times n}$ that provides a relationship in the form

$$\mathbf{X}_{k+1} = A\mathbf{X}_k. \tag{2.30}$$

Here, the relationship is a locally linear approximation of a possible nonlinear dynamics. While the relationship does not need to hold, there are theoretical justifications for the use of this approximation on data generated by nonlinear systems [41]. Operator $\mathbf{A}$ is constructed by searching for the best-fit solution for all pairs. This is done by choosing $\mathbf{A}$ that minimizes the error $\|\mathbf{X}_{k+1} - \mathbf{A}\mathbf{X}_k\|_2$. Intuitively, the approach uses numerical solutions to evolve observation at present time t_k to future state at time t_{k+1}.

Suppose snapshots of data measurements are collected on the $n-$dimensional state variable at m different times, the data may be arranged into finite dimensional state and time-shifted data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively as

$$\mathbf{X}_1 = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix} \tag{2.31a}$$

$$\mathbf{X}_2 = \begin{bmatrix} | & | & | & & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \cdots & \mathbf{x}_m \\ | & | & | & & | \end{bmatrix} \tag{2.31b}$$

The local linear approximation in equation (2.30) can then be described in terms of the data matrices as

$$\mathbf{X}_2 \approx \mathbf{A}\mathbf{X}_1. \tag{2.32}$$

The best-fit matrix $\mathbf{A}$ is closely related to the *Koopman operator* as we'll see shortly. $\mathbf{A}$ is expressed in terms of the measurement as

$$\mathbf{A} = \mathbf{X_2}\mathbf{X}_1^{\dagger}, \tag{2.33}$$

and the solution minimizes

$$\|\mathbf{X}_2 - \mathbf{A}\mathbf{X}_1\|_F. \tag{2.34}$$

where $\dagger$ is the *Moore-Penrose pseudoinverse*, and $\|\cdot\|_F$ is the *Frobenius norm*. The *Singular Value Decomposition (SVD)* algorithm is a commonly used method in computing the pseudoinverse, a method which involves a low-rank truncation of the possibly noisy data. In the implementation of the algorithm,

the method seeks a discrete-time system sampled at Δt analogous to

$$\frac{d\mathbf{X}}{dt} = \mathcal{A}\mathbf{X}. \tag{2.35}$$

where $\mathcal{A}$ represents the matrix in continuous-time dynamics with $\mathbf{A} = \exp(\mathcal{A}\Delta t)$. It is shown that solution to (2.30) can be expressed as

$$\mathbf{X}_k = \sum_{j=1}^{r} \boldsymbol{\phi}_j \lambda_j^k b_j = \boldsymbol{\Phi}\boldsymbol{\Lambda}^k \mathbf{b} \tag{2.36}$$

where λ_k, ϕ_k are eigenvalues and eigenvectors of the discrete-time map $\mathbf{A}$. Equation (2.36) therefore represents a low-rank eigendecomposition of $\mathbf{A}$. Heuristically, this construction produces a finite-rank representation of the possibly infinite dimensional linear operator that describes the evolution of the state of the dynamics a step in time [26].

DMD modes, frequently called dynamic modes are the eigenvectors of the matrix $\mathbf{A}$. Suppose $\mathbf{X}$ is an $n \times m$ *rank-r* data matrix, then the corresponding SVD decomposition of the data $\mathbf{X}$ is given as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*, \tag{2.37}$$

where * denotes the conjugate transpose, $\mathbf{U} \in \mathbb{C}^{n\times r}$ and $\mathbf{V} \in \mathbb{C}^{m\times r}$ are unitary matrices and $\Sigma \in \mathbb{C}^{r\times r}$ is a diagonal matrix. The matrix $\mathbf{U}$ contains *Proper Orthogonal Decomposition (POD)* modes and the diagonal values of the matrix Σ represent contributions of each mode.

To demonstrate, a schematic overview of DMD algorithm applied on a sample fluid flow data [26] is shown in Figure 1.

In practice, the matrix $\mathbf{A}$ can be very large in size, hence computing the eigendecomposition can become computationally expensive. As such, we solve the eigenvalue problem for an approximation matrix $\tilde{\mathbf{A}}$ derived from the SVD of $\mathbf{A}$. Subsequently, the low-rank approximation $\tilde{\mathbf{A}}$ is computed as a projection

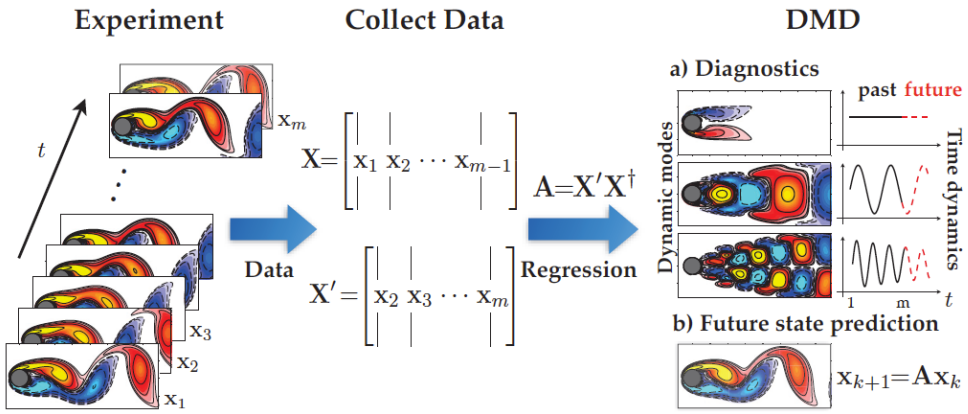


Figure 1. An overview of the DMD scheme on a sample fluid flow data. We note that the regression step $\mathbf{A} = \mathbf{X}'\mathbf{X}^\dagger$ does not typically construct the best-fit matrix $\mathbf{A}$, but instead an approximation $\tilde{\mathbf{A}}$ which evolves the dynamics in time. Eigendecomposition of the high-dimensional matrix $\mathbf{A}$ is then approximated by the eigendecompostion of $\tilde{\mathbf{A}}$. The eigenvectors of $\mathbf{A}$ are called DMD modes. Image source [26]

of $\mathbf{A}$ onto *POD* modes

$$\mathbf{A} = \mathbf{X}_2\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^* \tag{2.38a}$$

$$\tilde{\mathbf{A}} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{U}^*\mathbf{X}_2\mathbf{V}\mathbf{\Sigma}^{-1}. \tag{2.38b}$$

A low-dimensional linear model of the original system is defined on *POD* coordinates using $\tilde{\mathbf{A}}$, and the corresponding reconstruction of the high-dimensional state is given respectively in equation (2.39) below

$$\tilde{\mathbf{X}}_{k+1} = \tilde{\mathbf{A}}\tilde{\mathbf{X}}_k \tag{2.39a}$$

$$\mathbf{X}_k = \mathbf{U}\tilde{\mathbf{X}}_k. \tag{2.39b}$$

Let $\mathbf{W}$ be a matrix whose columns comprise the eigenvectors of $\tilde{\mathbf{A}}$ and Λ a diagonal matrix containing corresponding eigenvalues λ_k. Then we can compute the eigendecomposition of $\tilde{\mathbf{A}}$ as

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}. \tag{2.40}$$

To reconstruct the eigendecomposition of $\mathbf{A}$, eigenvalues of $\mathbf{A}$ are given by $\mathbf{\Lambda}$ whiles the corresponding eigenvectors of $\mathbf{A}$ (*DMD modes*) are given by

$$\Phi = \mathbf{X}_2 \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{W}. \tag{2.41}$$

Koopman Operator:

Let $(t, \mathcal{M}, \mathbf{F})$ denote the discrete-time dynamical system, where $t \in \mathbb{Z}$ is time, $\mathcal{M} \in \mathbb{R}^N$ is the state space and $\mathbf{F}$ is a *flow map*, so that $\mathbf{X} \mapsto \mathbf{F}(\mathbf{X})$ evolves the dynamics forward in time, that is $\mathbf{X}_{k+1} = \mathbf{F}(\mathbf{X}_k)$ in the discrete sense. Given a Hilbert space $\mathcal{H}$ of scalar-valued measurable functions $g : \mathbb{C}^n \to \mathbb{C}$, the Koopman operator is an infinite dimensional operator that acts on $g \in \mathcal{H}$ by composition

$$\mathscr{K} g = g \circ \mathbf{F} \tag{2.42a}$$

$$\mathscr{K} g\left(\mathbf{x}_k\right) = g\left(\mathbf{F}\left(\mathbf{x}_k\right)\right) = g\left(\mathbf{x}_{k+1}\right). \tag{2.42b}$$

The Koopman operator, also know as the composition operator, advances the function g on the measurements by mapping $g(\mathbf{x}_k)$ to $g(\mathbf{x}_{k+1})$.

In the case of continuous dynamics, consider a continuous dynamical system $\frac{d\mathrm{X}}{dt} = \mathbf{f}(\mathbf{X})$, where $\mathrm{X} \in \mathcal{M}$ is a state on a smooth n-dimensional manifold $\mathcal{M}$. The Koopman operator $\mathscr{K}$ is an infinite dimensional linear operator that acts on all observable functions $g : \mathscr{M} \to \mathbb{C}$ so that

$$\mathscr{K} g(\mathrm{X}) = g(\mathbf{f}(\mathrm{X})). \tag{2.43}$$

Essentially, the finite dimensional nonlinear dynamical system defined by $\frac{d\mathrm{X}}{dt} = \mathbf{f}(\mathbf{X})$ and the infinite-dimensional linear dynamics defined by $\mathscr{K}$ in (2.43) are equivalent representations. The transformation from the state space representation of the dynamical system to the Koopman representation trades nonlinear finite-dimensional dynamics for linear infinite-dimensional dynamics. The continuous system $\frac{d\mathrm{X}}{dt} = \mathbf{f}(\mathbf{X})$ induces a discrete time dynamical system by the flow map $\mathbf{F}_t : \mathscr{M} \to \mathscr{M}$ that maps $\mathrm{X}(t_0)$ to $\mathrm{X}(t_0 + t)$ via:

$$\mathbf{F}_t(X(t_0)) = X(t_0 + t) = X(t_0) + \int_{t_0}^{t_0+t} \mathbf{f}(X(\tau))d\tau. \tag{2.44}$$

This induces the discrete-time dynamical system

$$\mathbf{X}_{k+1} = \mathbf{F}_t(\mathbf{X}_k). \tag{2.45}$$

The analogous discrete-time Koopman operator is what was given as

$$\mathscr{K}_t g(\mathbf{X}_k) = g(\mathbf{F}_t(\mathbf{X}_k)) = g(\mathbf{X}_{k+1}), \tag{2.46}$$

with a corresponding eigenvalue problem of the Koopman operator formulated as

$$\mathscr{K}\varphi_k = \lambda_k \varphi_k \tag{2.47}$$

where $\varphi_k(\mathbf{X})$ is the Koopman eigenfunction with corresponding eigenvalue λ_k.

Connection between DMD and the Koopman operator: The DMD algorithm determines the Koopman eigenvalues and modes directly from data. This is done under certain conditions that rely heavily on the choice of observables. Figure 2 shows the contrast between the DMD and Koopman mode decomposition

Connection between DMD and the Koopman operator is such that DMD eigenvalues are Koopman eigenvalues provided so that: (1) The set of observables is sufficiently large; (2) The data is *rich*, as described in the theorem below.

Theorem 2.6.1. *(Koopman mode decomposition and DMD:[26]) Let φ_k be an eigenfunction of $\mathscr{K}$ with eigenvalue λ_k and suppose $\varphi_k \in$ span$\{g_j\}_j$ so that*

$$\varphi_k(\mathbf{X}) = w_1 g_1(\mathbf{X}) + w_2 g_2(\mathbf{X}) + \cdots + w_p g_p(\mathbf{X}) = \mathbf{w} \cdot g \tag{2.48}$$

for some $\mathbf{w} = [w_1, w_2, \cdots, w_p]^T \in \mathbb{C}^p$. If $\mathbf{w} \in range(\mathbf{Y})$, then $\mathbf{W}$ is a left eigenvector of $A_{\mathbf{Y}}$ with eigenvalue λ_k so that $\tilde{\mathbf{W}}^* A_{\mathbf{Y}} = \lambda_k \tilde{\mathbf{W}}^*$.

Similarly to the standard DMD algorithm, the following overview summarizes the Koopman process.

1. From the data matrices $\mathbf{X}$ and $\mathbf{X}'$, create data matrices of observables $\mathbf{Y}$ and $\mathbf{Y}'$

$$\mathbf{Y} = \begin{bmatrix} | & | & & | \\ \mathrm{g}\left(\mathbf{x}_1\right) & \mathrm{g}\left(\mathbf{x}_2\right) & \cdots & \mathrm{g}\left(\mathrm{x}_{m-1}\right) \\ | & | & & | \end{bmatrix}, \tag{2.49a}$$

$$\mathbf{Y}' = \begin{bmatrix} | & | & & | \\ \mathrm{g}\left(\mathbf{x}_1'\right) & \mathrm{g}\left(\mathrm{x}_2'\right) & \cdots & \mathrm{g}\left(\mathrm{x}_{m-1}'\right) \\ | & | & & | \end{bmatrix} \tag{2.49b}$$

   where each column is given by $y_k = g(\mathrm{x_k})$ or $\mathrm{y}_k' = g(\mathrm{x}_k')$

2. Perform the DMD algorithm to compute

$$\mathbf{A}_Y = \mathbf{Y}'\mathbf{Y}^\dagger \tag{2.50}$$

   where † is the pseudoinverse along with the low-rank approximation $\tilde{\mathbf{A}}_Y$. The eigenvalues and eigenvectors of $\mathbf{A}_Y$ may approximate Koopman eigenvalues and modes depending on the set of chosen observables

3. DMD may be used to compute the modes Φ_Y, which may approximate the Koopman modes

$$\mathbf{\Phi}_Y = \mathbf{Y}'\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}, \tag{2.51}$$

   where $\mathbf{W}$ comes from $\tilde{\mathbf{A_Y}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$ and $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$.

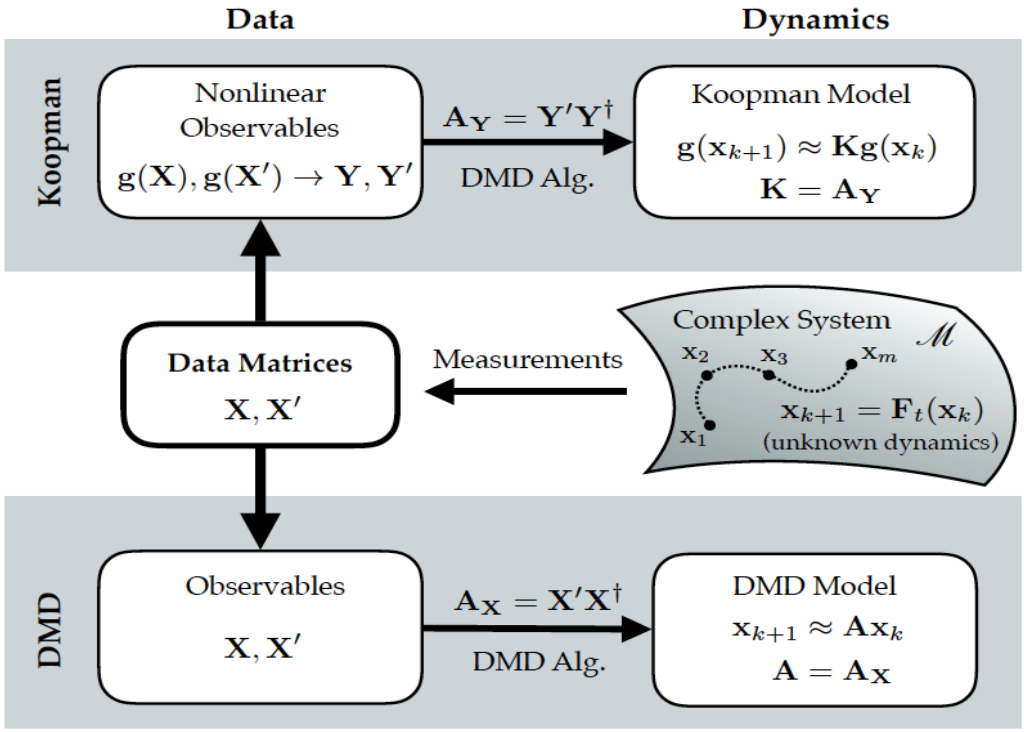
4. In the state of observables, the future state is given by

$$\mathrm{y}(t) = \mathbf{\Phi}_Y \mathrm{diag}\left(\exp(\omega t)\right)\mathbf{b}, \tag{2.52}$$

   where $\mathbf{b} = \mathbf{\Phi}_Y^\dagger \mathrm{y}_1$ is determined by projecting back to the initial data observables and ω are the set of eigenvalues λ_k generated from the matrix Λ where $\omega_k = \ln\left(\lambda_k\right)/\Delta t$.

5. To transform observables back to state space:

$$\mathrm{y}_k = \mathbf{g}(\mathrm{X}_k) \rightarrow \mathrm{X}_k = \mathbf{g}^{-1}(\mathrm{y}_k). \tag{2.53}$$



.

Figure 2. *Schematics of DMD/Koopman framework. In standard DMD, state measurements are taken to construct a model that advances the state in time from* $\mathbf{X}$ *to* $\mathbf{X}'$*. Koopman spectral analysis uses measurements with nonlinear observables* $y = g(\mathbf{X})$ *to provide a better map from* $\mathbf{Y}$ *to* $\mathbf{Y}'$*. Both methods does not rely on knowing the flow map* F*.* Image source: *Dynamic mode decomposition: data-driven modeling of complex systems*, Kutz et al [26]

Most machine learning algorithms are primarily concerned with clustering and classifications, the DMD algorithm can however be modified in various ways to potentially incorporate the nonlinear manifold on which the dynamics occur [26]. Using techniques from kernel methods in machine learning, one

technique for choosing system observables g_j is to think of

$$\mathbf{g}(\mathrm{X}) = \begin{bmatrix} g_1(\mathrm{X}) \\ g_2(\mathrm{X}) \\ \vdots \\ g_n(\mathrm{X}) \end{bmatrix}$$

as a map from the physical space into a feature space, in which case the dynamical system is constructed as

$$\mathbf{g}(\mathbf{X}) = \sum_{k=1}^{\infty} \mathbf{v}_k \varphi_k(\mathbf{X}) \tag{2.54a}$$

$$g(f(\mathbf{X})) = \sum_{k=1}^{\infty} \mathbf{v}_k \lambda_k \varphi_k(\mathbf{X}), \tag{2.54b}$$

where λ_k, φ_k are eigenvalues and eigenfunctions of the Koopman operator and $\mathbf{v}_k$ are Koopman modes associated with the eigenfunction. Popular ways of constructing the feature space $\mathbf{g}(\mathbf{X})$ includes using the set of polynomials

$$g_j(x) = \{x, x^2, x^3, \ldots, x^n\}. \tag{2.55}$$

Alternatively, there are other kernel functions commonly used such as radial basis function and *Hermite polynomials*.

Extended DMD: DMD can incorporate machine learning ideas such as kernels and extended observables aimed at reducing computational cost and to robustify the algorithm. Given data matrices of observables

$$\mathbf{Y} = \begin{bmatrix} | & | & & | \\ \mathrm{g}\left(\mathbf{x}_1\right) & \mathrm{g}\left(\mathbf{x}_2\right) & \cdots & \mathrm{g}\left(\mathrm{x}_{m-1}\right) \\ | & | & & | \end{bmatrix}, \tag{2.56a}$$

$$\mathbf{Y}' = \begin{bmatrix} | & | & & | \\ \mathrm{g}\left(\mathbf{x}_1'\right) & \mathrm{g}\left(\mathrm{x}_2'\right) & \cdots & \mathrm{g}\left(\mathrm{x}_{m-1}'\right) \\ | & | & & | \end{bmatrix} \tag{2.56b}$$

the DMD algorithm produces the matrix decomposition

$$\mathbf{A}_Y = \mathbf{Y}'\mathbf{Y}^\dagger \tag{2.57}$$

along with approximate $\tilde{\mathbf{A}}_Y$ in the instance where the observable matrices is extremely large ($n \gg m$ or $m \gg n$), where $\mathbf{A}_Y$ becomes computationally intractable. Extended DMD and kernel techniques produce numerically efficient ways to approximate the $\tilde{\mathbf{A}}_Y$, a finite dimensional approximation of the Koopman operator. In this case, the Koopman operator generated by using a kernel trick reduces computational cost. Extended DMD is a method developed to reduce the cost of evaluating the Koopman operator when the number of snapshots is extremely large ($m \gg n$). Suppose $m \gg n$, one can consider the efficient computation for $\mathscr{K}$ by

$$\begin{aligned}
\mathbf{A_Y} &= \mathbf{Y}'\mathbf{Y}^\dagger \\
&= \mathbf{Y}'\mathbf{I}^\dagger \\
&= \mathbf{Y}'\left(\mathbf{Y}^T\mathbf{Y}^{T\dagger}\right)\mathbf{Y}^\dagger \\
&= \left(\mathbf{Y}'\mathbf{Y}^T\right)\left(\mathbf{Y}\mathbf{Y}^T\right)^\dagger \\
&= \mathbf{A}_1\mathbf{A}_2^\dagger,
\end{aligned} \tag{2.58}$$

where I is the identity matrix of appropriate size and

$$\mathbf{A}_1 = \mathbf{Y}'\mathbf{Y}^T \tag{2.59a}$$

$$\mathbf{A}_2 = \mathbf{Y}\mathbf{Y}^T. \tag{2.59b}$$

It can be observed that $\mathbf{A}_1$, $\mathbf{A}_2 \in \mathbb{C}^{n\times n}$ are much smaller matrices than the original $n \times m$ formulation. Inversion of $\mathbf{Y}$ is traded for that of a much smaller matrix $\mathbf{A}_2$. This architecture is illustrated in Figure 3.

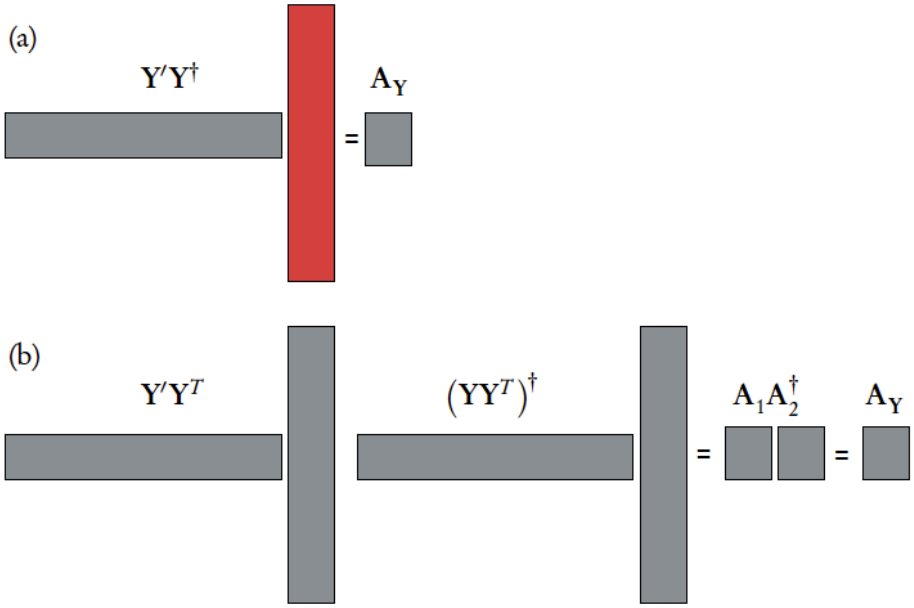


Figure 3. Extended DMD architecture with more snapshots m much greater than the number of observables n. Image source: [26, 41]

Kernel DMD: Kernel DMD is more suited for cases when $n \gg m$ resulting in long and *skinny* observable matrices. We consider an efficient computation of the Koopman operator by projecting to the principal component space obtained through the SVD of the data matrix $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{*}.$$

For the Koopman eigenvalue problem $\mathscr{K}\,\mathbf{y}_k = \lambda_k \mathbf{y}_k$, we suppose the eigenvector $\mathbf{y}_k$ can be constructed by expansion,

$$\mathbf{y}_k = \mathbf{U}\mathbf{y}_k. \tag{2.60}$$

Substituting this into the eigenvalue problem

$$
\begin{aligned}
\lambda_k \mathbf{y}_k &= \mathscr{K}\, \mathbf{y}_k, \\
\lambda_k \mathrm{U}\hat{\mathbf{y}}_k &= \mathscr{K}\, \mathbf{U}\hat{\mathbf{y}}_k \\
&= \mathrm{Y}'\mathrm{Y}^{\dagger}\mathrm{U}\hat{\mathbf{y}}_k \\
&= \mathrm{Y}'\mathrm{Y}^{\dagger}\left(\mathrm{Y}\mathrm{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \mathrm{Y}'\left(\mathrm{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \mathrm{I}\mathrm{Y}'\left(\mathrm{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \left(\mathrm{Y}^{T}\right)^{\dagger}\mathrm{Y}^{T}\mathrm{Y}'\left(\mathrm{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \left(\mathrm{Y}^{T}\right)^{\dagger}\left(\mathrm{Y}^{T}\mathrm{Y}'\right)\left(\mathrm{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \mathrm{U}\left(\Sigma\mathbf{V}^{*}\right)\left(\mathrm{Y}^{T}\mathrm{Y}'\right)\left(\mathbf{V}\Sigma^{\dagger}\right)\hat{\mathbf{y}}_k \\
&= \mathbf{U}\mathscr{K}\,\mathbf{y}_k,
\end{aligned}
\tag{2.61}
$$

where I is the identity matrix. The Koopman operator is evaluated by the expression

$$
\mathbf{A}_{\mathbf{Y}} = \left(\Sigma\mathbf{V}^{*}\right)\left(\mathbf{Y}^{T}\mathbf{Y}'\right)\left(\mathbf{V}\Sigma^{\dagger}\right). \tag{2.62}
$$

We notice that $\left(\Sigma\mathbf{V}^{*}\right) \in \mathbb{C}^{m\times m}$, $\left(\mathbf{Y}^{T}\mathbf{Y}'\right) \in \mathbb{C}^{m\times m}$, and $\left(\mathbf{V}\Sigma^{\dagger}\right) \in \mathbb{C}^{m\times m}$, suggesting that $\mathbf{A}_{\mathbf{Y}}$ is determined by the number of snapshots taken rather than the number of features. Figure (3) illustrates this computational infrastructure. The kernel method reduces the computation of the Koopman operator to the product of the three matrices that are projected to the feature space embedding $\mathbf{U}$.

Application: Infectious Disease Data

Among the many disciplines that makes use of modern DMD is epidemiology. The method has been applied to many different kinds of disease data to answer a myriad of questions. Such data may be from experiments, numerical simulations or historical data with state variables that may vary from study to study. For example, the number of infections in a given region, average duration of incubation for the disease or other demographic information [41]. The well-curated data is arranged into a DMD data

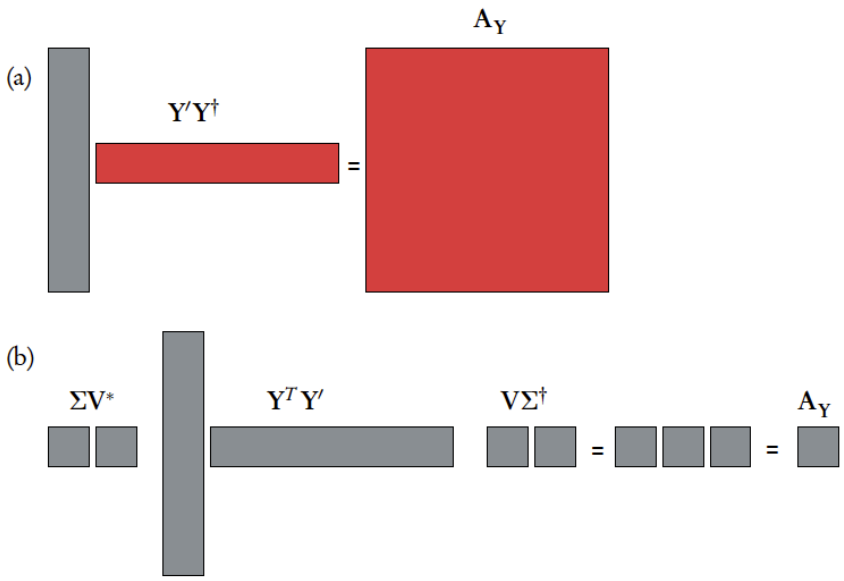


Figure 4. Illustration of the DMD algorithm for infectious disease data. Image source: [26]

matrix as

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix} \tag{2.63a}$$

$$\mathbf{X}' = \begin{bmatrix} | & | & | & & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \cdots & \mathbf{x}_m \\ | & | & | & & | \end{bmatrix}, \tag{2.63b}$$

where columns represent different snapshots in time and rows describe specific state of the system. Figure 5 gives an illustration of the data collection and the DMD architecture in the studies. The set of eigenvalues give dynamic characteristics such as growth, decay, and oscillatory behavior of the dynamic mode. Figure 6 shows an eigenvalue spectrum with two pairs of complex conjugate eigenvalues; the red pair represent purely oscillatory modes since they lie on the unit circle whereas the blue pair have decaying dynamic characteristics since they lie within the circle and thus are stable. It is convenient to convert the oscillatory frequency of the discrete eigenvalue of the approximation low rank approximation

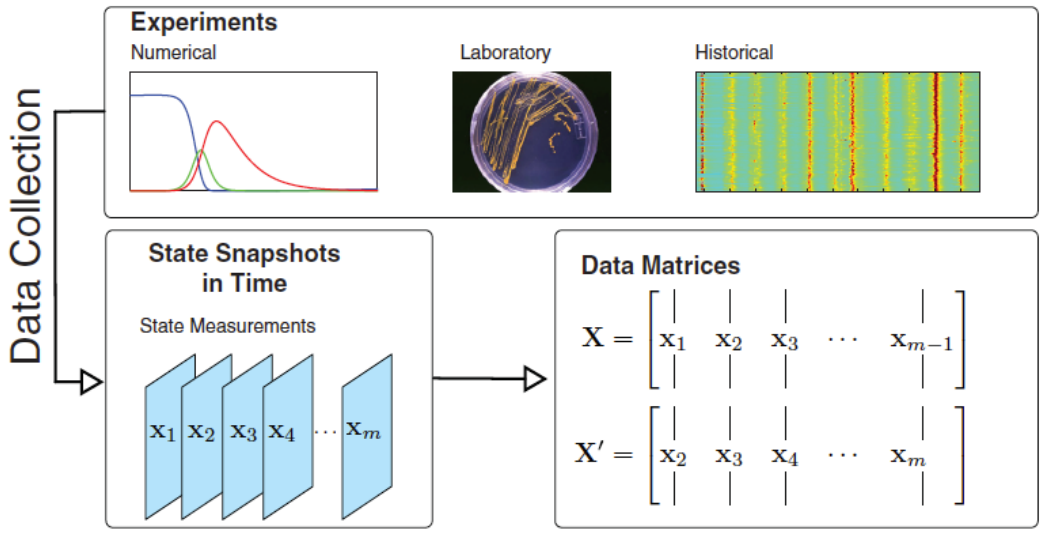


Figure 5. Data collection and arrangement of snapshot in the data matrix for DMD analysis. Image source: [26, 41]

map $\tilde{\mathbf{A}}$ into a continuous time oscillatory frequency. The relation is given by

$$\text{frequency}_j = \frac{Im(\ln \lambda_j)}{2\pi\Delta t}. \tag{2.64}$$

This allows for each eigenvalue to be examined based on intuitive and interpretable continuous frequency with units of per year [41]. Each element in a dynamic mode vector has two essential pieces of information. Firstly, the absolute value of the element provides a measure of the spatial location's participation in the mode. For complex valued elements, the angle between the real and imaginary components of the element provides a measure of a location's phase of oscillation relative to others for that mode's frequency.

Example: Google Flu Trends [26, 41] In this example, DMD is applied to infectious disease data from Google's flu trend tool. Google has studied how specific search terms can be indicators for predicting flu spread within the United States. Google used curated historical flu data and search data to construct a method for estimating current flu activity in the country. Though recent scientific investigations have

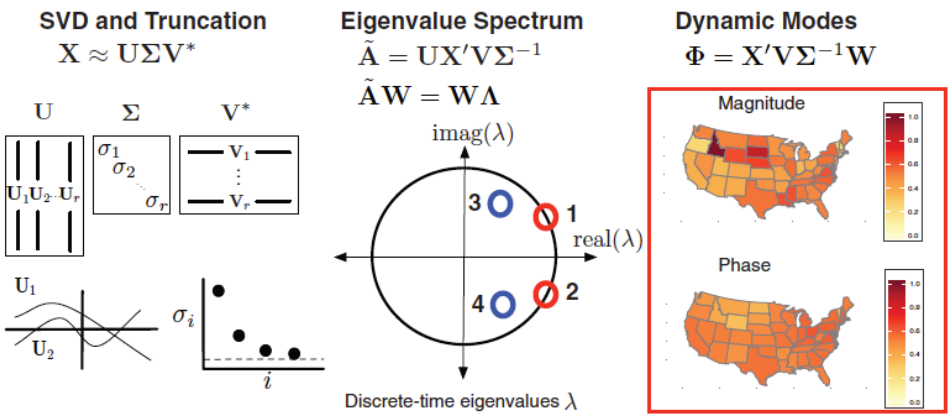


Figure 6. Illustration of the DMD algorithm for infectious disease data. Image source: [41]

cast doubt on the validity of the analysis, the data set used in this example was solely to demonstrate the implementation of the DMD framework. Figure 7 indicates 4 traces of the unprocessed data from these 4 states Alaska (black), California (red), Texas (green), and New York (blue). The data is provided for every 7 days (serving as Δt) and represented as a two-dimensional array with state, city and the health-human-services regional breakdown. A normalization is performed to condition the data for application of DMD. The methods focus on state information in order to visualize every element of the dynamic mode on the map of the US. The output of DMD analysis is shown to the right of the data visualization in Figure 7. As shown, the eigenvalue spectrum indicates a number of modes that are within the unit circle, which shows fast decaying eigenvalues and modes that do not have a significant contribution to the broader structure of the dynamics. The mode selection plot illustrates a yearly frequency. The mode and frequency aligns with seasonal varying profiles in Figure 7d. The phase of the dynamic mode associated with this yearly frequency is plotted in Figure 7e. (phase is scaled between 0 and 1 indicating time of the year.)

2.7 Occupation Kernel DMD

Despite the robustness of Koopman-based DMD, there are a few drawbacks that potentially limit its application on certain classes of dynamics. Convergence of the DMD operator to the Koopman operator is typically in the *strong operator topology (SOT)* sense which is a pointwise convergence. This is not enough to justify the use of DMD to interpolate or extrapolate the system behavior from a collection of diverse samples. In contrast, norm convergence is uniform convergence for operators and yields a

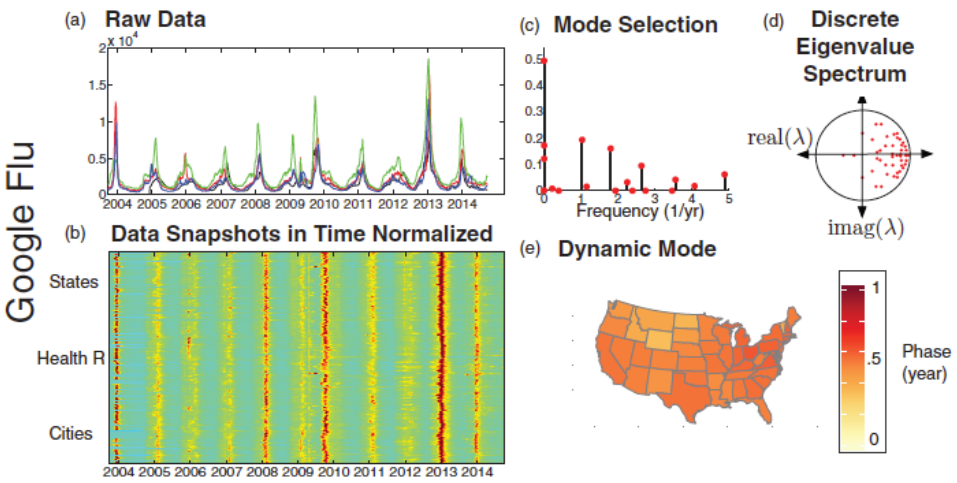


Figure 7. Illustration of the DMD algorithm for infectious disease data. Image source: [41]

bound on the error over the kernels corresponding to the data set, however this result is only possible for compact Koopman operators which hardly exist in many applications of interest.

A fundamental assumption of Koopman-based DMD is the idea that a finite dimensional nonlinear dynamical system can be expressed as a linear operator over an infinite dimensional space. Additionally, the method assumes *forward completeness* of the system, for example, by assuming the dynamics are globally Lipschitz. These limitations restrict the class of systems suitable for the Koopman DMD [47]. As an example, consider the continuous time dynamics $\dot{x} = 1 + x^2$. Discretizing this system with time step 1 produces the following discrete dynamics: $x_{k+1} = \mathbf{F}(x_k) = \tan\left(1 + \tan^{-1}(x_k)\right)$. It can be observed that $\mathbf{F}$ has a finite escape time, thus is not defined over $\mathbb{R}$, and since the symbol for the Koopman operator must be defined over the whole domain, there is no well-defined Koopman operator arising from this discretization.

In an effort to address these limitations, the authors in [49] couple densely-defined *Liouville operators* over a RKHS with *occupation kernels* to develop a Liouville-based DMD algorithm for continuous time dynamical systems. Occupation kernels are special functions that reside in a RKHS by virtue of an integration functional [46, 47, 49, 50]. These "kernels" remove the responsibility of approximation from that of the operators and places it on the estimation of occupation kernels from time-series data which require less theoretical conditioning. The authors show that the incorporation of Liouville operators in the routine allows for the study of dynamics that are locally rather than globally Lipschitz.

Definition 2.7.1. Suppose $\mathcal{H}$ is a RKHS over a compact set $\mathcal{X} \subset \mathbb{R}^n$, let $\dot{x} = f(x)$ be a dynamical system with locally Lipschitz and continuous dynamics $f : \mathbb{R}^n \to \mathbb{R}^n$. The Liouville operator with symbol f denoted by A_f is the map $A_f : \mathcal{D}(A_f) \to \mathcal{H}$ defined by

$$\begin{aligned} A_f g &:= \nabla_x g \cdot f \quad where \\ \mathcal{D}(A_f) &:= \{g \in \mathcal{H} : \nabla_x g \cdot f \in \mathcal{H}\}. \end{aligned} \tag{2.65}$$

Considering classes of dynamics for which the associated Liouville operator is bounded and densely defined, it can be shown that the action of the adjoint operator on an occupation kernel admits a closed form that depends directly on the data.

Proposition 2.7.1. *Let $\mathcal{H}$ be a RKHS of continuously differentiable functions over a compact set $\mathcal{X}$, and suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous. If $\gamma : [0, T] \to \mathcal{X}$ is a trajectory as in Definition 3.4.3 that satisfies $\dot{\gamma} = f(\gamma)$, then $\Gamma_\gamma \in \mathcal{D}\left(A_f^*\right)$, and $A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$ where $K(\cdot, \cdot)$ is the reproducing kernel function of $\mathcal{H}$.*

The action A_f^* gives the difference of the kernel function centered at end points of the trajectory, which completes the integration of the nonlinear dynamics into the RKHS. This connection sets the state for finite dimensional nonlinear dynamics to be expressed as linear system in infinite dimensions. Suppose $\alpha = \{\Gamma_{\gamma_i}\}_{i=1}^M \subset \mathcal{H}$, given as $\Gamma_{\gamma_i}(x) = \int_0^{T_i} K(x, \gamma_i(t))dt$ for a collection of trajectories $\gamma_i : [0, T_i] \to \mathcal{X}$ satisfying $\dot{\gamma}_i = f(\gamma_i)$. If α is selected as the basis for a vector space, the action A_f^* is known on the $span(\alpha)$. Per this setup, the DMD procedure is to express a matrix representation of A_f^* on the finite dimensional space spanned by α followed by a projection onto $span(\alpha)$ [49]. Take $w_1, w_2, \cdots, w_M$ to be the coefficients for the projection of $g \in \mathcal{H}$ onto $span(\alpha) \subset \mathcal{H}$, written as $P_\alpha g = \sum_{i=1}^M w_i \Gamma_{\gamma_i}$. The coefficients $w_1, w_2, \cdots, w_M$ may be obtain through the solution of a linear system given as

$$\begin{pmatrix} \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_M \end{pmatrix} = \begin{pmatrix} \langle g, \Gamma_{\gamma_1} \rangle_H \\ \vdots \\ \langle g, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \tag{2.66}$$

Where the inner products may be expressed as a single or double integral

$$\left\langle \Gamma_{\gamma_j}, \Gamma_{\gamma_i} \right\rangle_H = \int_0^{T_i} \int_0^{T_j} K\left(\gamma_i(\tau), \gamma_j(t)\right) dt d\tau$$
$$\left\langle g, \Gamma_{\gamma_i} \right\rangle_H = \int_0^{T_i} g\left(\gamma_i(t)\right) dt. \tag{2.67}$$

2.8 Sparse Identification For Nonlinear Dynamics (SINDy)

The *SINDy* algorithm, like its close cousin DMD, aims at extracting governing equations from data measurements from a dynamical system. Introduced by Brunton et al in 2016 [8], the method combines sparsity-promoting techniques and machine learning to obtain numerical and analytic insight about noisy data. The authors assert that the dynamics of most physical systems are fully expressible in a few relevant terms thereby making governing equations sparse in a high-dimensional space.

Consider the dynamical system

$$\frac{d\mathbf{X}(t)}{dt} = f(\mathbf{x}(t)), \tag{2.68}$$

where $\mathbf{X}$ is the state and f is a dynamic constraint. To determine f from data, a time history of state information is collated and stored in data matrices $\mathbf{X}$ and $\dot{\mathbf{X}}$ as shown in equations (2.69a) and (2.69b) respectively. Next, a library of candidate nonlinear functions is constructed with each column representing a candidate function for f, indicated as the array of symbolic function Θ in equation (2.69c). Under the assumption that a few terms are active, a sparse regression problem is setup to

determine those nonlinearities so that system (2.68) is approximated as shown in equation (2.69d).

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \cdots & x_n(t_m) \end{bmatrix} \tag{2.69a}$$

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^T(t_1) \\ \dot{\mathbf{x}}^T(t_2) \\ \vdots \\ \dot{\mathbf{x}}^T(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \cdots & \dot{x}_n(t_m) \end{bmatrix} \tag{2.69b}$$

$$\Theta(\mathbf{X}) = \begin{bmatrix} 1 & 1 & \overset{|}{1} & \prime & & | & | & \\ 1 & \mathbf{X} & \mathbf{X}^{P_2} & \mathbf{X}^{P_3} & \cdots & \sin(\mathbf{X}) & \cos(\mathbf{X}) & \cdots \\ | & | & | & | & & | & | & \end{bmatrix} \tag{2.69c}$$

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi \tag{2.69d}$$

where Ξ is a sparse vector of coefficients.

A model of each row of $f(\mathbf{x}(t))$ is determined as

$$\dot{\mathbf{x}}_k = f_k(\mathbf{x}) = \Theta(\mathbf{x})\xi_k \tag{2.70a}$$

$$\dot{\mathbf{x}} = f(\mathbf{x}) = \Xi^\top \left(\Theta(\mathbf{x}^\top)\right)^\top \tag{2.70b}$$

where ξ_k is the kth column of Ξ.

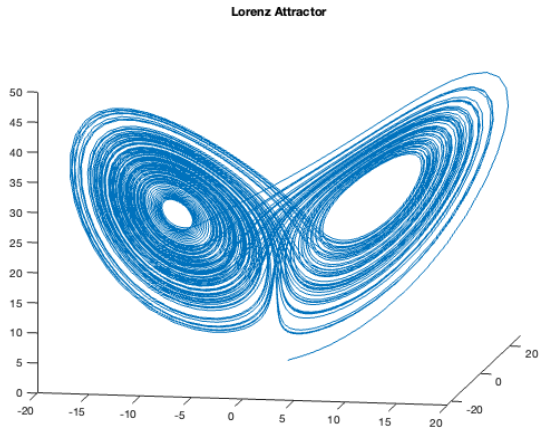
Realistically, $\dot{\mathbf{X}}$ is approximated from $\mathbf{X}$, in addition to the possibility of noise contamination, a regularized version of equation (2.69d) is used

$$\mathbf{X} = \Theta(\mathbf{X})\Xi + \eta\mathbf{z}. \tag{2.71}$$

where $\mathbf{z}$ is modeled as a matrix of Gaussian noise with zero mean and magnitude η. Equation (2.71) is commonly an overdetermined system with noise, to which sparse promoting techniques like the *least absolute shrinkage and selection operator (LASSO)* or *sequential threshold least-square* algorithms may

be used. As it is rightly acknowledged by the authors, there is no single method that will solve all problems in nonlinear system identification. One of the challenges of the *SINDy* algorithm is that the sparse identification procedure relies on the fortunate choice of coordinate and basis functions that facilitate sparse representation of the dynamics [8]. The identification may fail if the dynamics is not sparse in a chosen basis. To mitigate the cost of making the incorrect choices in basis selection, it is helpful to test many different function bases.

Among the experiments the authors performed to validate the algorithm is the chaotic dynamic; the Lorenz system as shown in (2.72)



$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z. \end{aligned} \tag{2.72}$$

Figure 8. Lorenz equation (left), Solution to the Lorenz system with $[\delta, \beta, \rho] = [10, 8/3, 28]$ (right)

with $[\delta, \beta, \rho] = [10, 8/3, 28]$ and initial conditions $[x, y, z]^T = [-8, 7, 27]$. Data is collected from $t = 0$ to $t = 100$ with $\Delta t = 0.0001$ and stacked in matrices $\mathbf{X}$ and $\dot{\mathbf{X}}$. A symbolic library of basis functions is chosen to be ploynomials in (x, y, z) up to fifth order.

$$\Theta(\mathbf{X}) = \begin{bmatrix} | & | & | & | & | & & | \\ \mathbf{x}(t) & \mathbf{y}(t) & \mathbf{z}(t) & \mathbf{x}(t)^2 & \mathbf{x}(t)\mathbf{y}(t) & \cdots & \mathbf{z}(t)^5 \\ | & | & | & | & | & & | \end{bmatrix} \tag{2.73}$$

A schematic framework of the algorithm demonstrated on the Lorenz equation is given in Figure 9 .

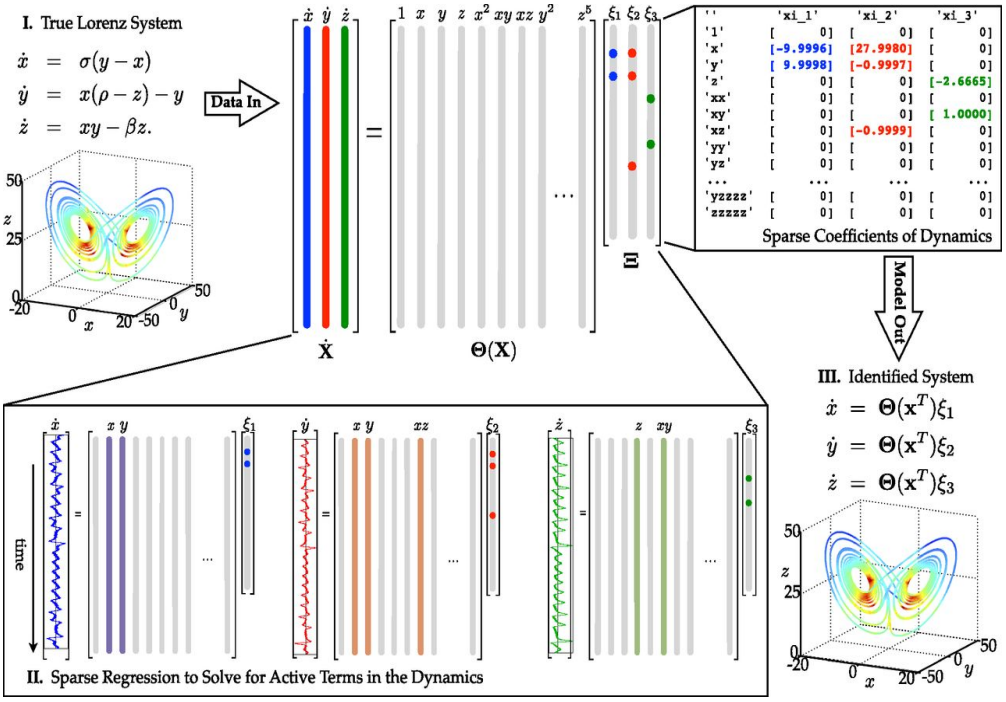


Figure 9. Schematic of the SINDy algorithm demonstrated on the Lorenz system. Image source: Reprinted with permission from the authors and copyright owner [8]

To conclude the chapter, we address yet another data-driven method which is the platform on which the model presented in Chapter 3 is built.

2.9 Nonlinear Autoregressive Moving Average Model With Exogenous Input (NARMAX)

Most dynamical systems in sciences such as physics, engineering and biology either do not have completely established scientific laws or closed form models for their dynamics. *System Identification* routines such as *Nonlinear AutoRegressive Moving Average model with eXogenous inputs (NARMAX)* offer an alternative method for establishing a mathematical description using observed inputs and outputs of the system [5]. In instances where the algorithm relies solely on data samples to reproduce unknown dynamics with no *a priori* information, it is chiefly called *black box* system identification. Nonlinearities exist in most systems to some degree, and complexities in their behavior limit the ability for linear models to capture fully their rich dynamics. Hence there is a rising need for nonlinear techniques that address such shortfalls while maintaining applicability and interpretability. The non-parametric nonlinear *NARMAX* model was introduced in a seminal work by S. A. Billings and I. J. Leontaritis in the early 80s. It was later presented in the publication [4] by the duo. The model has since seen a number of

revisions and extensions as well as practical applications [**e.g see**, 3, 9, 10, 25, 64]. In this section, we survey polynomial *NARMAX* models. For reference see the book [4] by Stephen A. Billings. A brief overview of multiresolution analysis which involves a hierarchical function approximation will also be gleaned. Subsequently in chapter 3, the *NAR* a derivative of model is interfaced with occupation kernel to perform parameter identification.

Early works in nonlinear system identification methods were influenced by the *Volterra Series* stated in discrete time as

$$\begin{aligned} y(k) =& h_0 + \sum_{m_1=1}^{M} h_1(m_1)\, u(k-m_1) + \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} h_2(m_1,m_2)\, u(k-m_1)\, u(k-m_2) \\ &+ \sum_{m_1=1}^{M}\sum_{m_2=1}^{M}\sum_{m_3=1}^{M} h_3(m_1,m_2,m_3)\, u(k-m_1)\, u(k-m_2)\, u(k-m_3) + \cdots \end{aligned} \tag{2.74}$$

where $u(k), y(k), k = 0, 1, \ldots$ are measured inputs and outputs respectively. $h(m_1, m_2, \ldots, m_l)$ is the ℓth order Volterra kernel. Initial implementations of this model involved at most the first two terms and a Gaussian white noise, until it was later extended to include more terms [5]. Common shortfalls of volterra series include; the number of terms needed to adequately represent a function is not known; more so, the number of data points needed to identified for good estimation can easily become excessively large. Example, suppose the $1st$ order kernel $h_1(m_1)$ is described by 30 samples, then 30×30 points will be required for $h_2(m_1, m_2)$, and $30 \times 30 \times 30$ for $h_3(m_1, m_2, m_3)$ and so on [5].

Consider for example the Duffings Equation, $\ddot{x} = x - x^3$, $x \in \mathbb{R}$ which may be augmented as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 - x_1^3 \end{bmatrix} \tag{2.75}$$

The linear dynamics are second order but the model contains a cubic term and so the Volterra series will contain at least 3 kernels. This may involve estimation of a large number of values to characterize the system and consequently a huge data size. *NARMAX* on the other hand is constructed as an expansion of a nonlinear function in terms of past inputs, outputs and noise terms. The *Single-Input Single-output*

(SISO) NARMAX model is defined as

$$y(k) = f\left[y(k-1), \ldots, y(k-n_y), u(t-d-1), \ldots, u(k-d-n_u), e(k-1), \ldots, e(k-n_e)\right] + e(k) \tag{2.76}$$

where the map $f : \mathcal{Y}^{n_y} \times U^{n_u} \times \Xi^{n_e} \to \mathcal{Y}$ is an unknown nonlinear dynamics of $n_y + n_u + n_e$ variables. $u \in U$ is an m-dimensional input vector and $y \in \mathcal{Y}$ an $l-$dimensional output vector. n_y, n_u are the maximum time lags for output and input respectively. d is time delay which typically taken to be 1. The noise variable $e(t)$ is assumed to be bounded $|e(k)| < \delta$ and uncorrelated with maximum lag n_e. Typically, $e(t)$ is taken as the error $e(k) = y(k) - \hat{y}(k-1)$.

The philosophy of the *NARMAX* system identification process consists of structure detection, parameter estimation, model validation, prediction, and analysis. Billings highlights in his book that, an advantage of the model in (2.76) is the inclusion of past outputs $(y(k-1), y(k-2), \cdots, y(k-n)y))$, as the approach yields a more concise model and requires fewer data points to implement as compared the Volterra series. He illustrates this through the Finite Impulse Response (FIR) model whose expansion involves only past inputs compared to its counterpart, the Infinite Impulse Response (IIR) filter.

$$\begin{aligned} \text{FIR}: \quad & y(k) = b_1 u(k-1) + b_2 u(k-2) + \cdots + b_{nb} u(k-n_{nb}) \\ \text{IIR}: \quad & y(k) + a_1 y(k-1) + a_2 y(k-2) + \cdots + a_{na} y(k-na) \quad = b_1 u(k-1) + \\ & b_2 u(k-2) + \cdots + b_{nb} u(k-n_{nb}) \end{aligned} \tag{2.77}$$

where na and nb are model orders.

Normally, a simple linear system of a FIR filter may need 50 weights $nb = 50$, whereas the IIR filter would need about 4 , due to the incorporation of the output lagged terms [3, 5].

NARMAX is able to represent a wide class of nonlinear systems as well as perform a structure selection to include only highly weighted regressors [3, 4]. Both deterministic and stochastic cases of *NARMAX* exist for some choice of $f[\cdot]$ provided the system is finitely realisable, and a linearized model would exist if the system were operated close to an equilibrium point, supporting the versatility of *NARMAX* models. Linear and nonlinear model types like Autoregressive models, Autoregressive Moving Average with Exogenous (*ARMAX*), and Volterra series can be viewed as special cases of NARMAX. *Power form polynomial, rational models, Neural networks, wavelet expansions and Radial Basis Function (RBF)*

networks are sample model structures used to approximate a nonlinear function $f[\cdot]$ that satisfies equation (2.76).

Polynomial models: The polynomial NARMAX model estimates $f[\cdot]$ by multivariable polynomials with finite degree of the form

$$\begin{aligned} y(k) = \theta_0 + \sum_{i_1=1}^{n} f_{i_1}(x_{i_1}(k)) + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} f_{i_1 i_2}(x_{i_1}(k), x_{i_2}(k)) + \dots \\ + \sum_{i_1=1}^{n} \dots \sum_{i_l=i_{l-1}}^{n} f_{i_1 i_2 \dots i_l}(x_{i_1}(k), x_{i_2}(k), \dots, x_{i_l}(k)) + e(k) \end{aligned} \tag{2.78}$$

Where l is the degree of polynomial nonlinearity, $n = n_y + n_u + n_e$

$$f_{i_1 i_2 \dots i_l}(x_{i_1}(k), \dots, x_{i_l}(k)) = \theta_{i_1 i_2 \dots i_M} \prod_{k=1}^{M} x_{i_k}(k) \quad 1 \le m \le l$$

with $\theta_{i_1 i_2 \dots i_M}$ as model parameters and

$$x_m(k) = \begin{cases} y(k-m), & 1 \le m \le n_y \\ u(k-(m-n_y)), & n_y + 1 \le m \le n_y + n_u \\ e(k-(m-n_y-n_u)), & n_y + n_u + 1 \le m \le n \end{cases}$$

Specifically, (2.78) is explicitly written as

$$\begin{aligned} y(k) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) x_{i_2}(k) + \dots \\ + \sum_{i_1=1}^{n} \dots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1 i_2 \dots i_l} x_{i_1}(k) x_{i_2}(k) \dots x_{i_l}(k) + e(k). \end{aligned} \tag{2.79}$$

CHAPTER 3:
OPERATOR THEORETIC PARAMETER IDENTIFICATION SCHEME VIA NAR ALGORITHM

3.1 Introduction and Motivation

Given a dynamical system $\dot{x} = f(x,t)$, the problem of the current chapter is to identify parameters that uniquely approximate the unknown function f in the basis of a reproducing Kernel Hilbert space (RKHS) using data measurement from the system. We explore an alternative formulation for the *Nonlinear Autoregressive (NAR)*-type model that leverages densely defined multiplication operators. Over the past two decades, operator theoretic methods for modeling of dynamical systems has been gaining substantial popularity. These methods most often utilize semi-groups of Koopman operators and their generators for studying continuous time dynamical systems arising from differential equations. While [51] provided a direct connection between these generators and a data driven inner product for a parameter identification routine for such systems, this methodology does not directly apply to NAR type dynamical systems, refer to [8, 22, 29]. Dynamical systems appear widely across numerous disciplines, including neuro-science [34], biology [13, 59], engineering [11, 20, 23], and physics [62]. System identification routines involve reproducing system input-output measurements, and often times, the application of the resulting model to capture prevalent dynamical structure [44]. Linear models as we know, can be deficient in their ability to represent complex nonlinear dynamics [44], while most everyday systems are nonlinear to some degree and this dictates the need for identification techniques specific to nonlinear systems.

Several examples in the field of data-driven modeling have been illustrated where either the parameters of the concerned dynamics are partially or entirely unknown and are often referred to as *gray-box* and *black-box* respectively [35].

There exist various modeling techniques for addressing the non-linear system identification problem such as neural networks [35], Lyapunov methods [37], and Volterra series [15]. One of the main technical challenges that arises from the aforementioned methods is the estimation of the state derivative, which unfortunately is prone to error, and introduces an artificial noise component that requires additional filtering [8].

The contribution of this dissertation is to address a non-linear system identification routine by employing *Occupation kernel* techniques with *densely defined multiplication operators*. We develope a new parameter idetification algorithm to determine dynamics of the form $\dot{x} = f(x(t-\tau_1), x(t-\tau_2), \cdots, x(t-\tau_p))$, for $p \in \mathbb{Z}^+$ and $\tau_i \in \mathbb{R}$. An advantage of this approach is that, properties of the underlying RKHS are included in the identification process, which grants us access to orthonormal basis of the space for the approximation. Occupation kernels [50] can be regarded as the generalization of *occupation measures*. Multiplication operators are well studied objects over a variety of RKHSs, such as the Hardy and Fock spaces [1, 66]. One aspect of this study is the characterization of the impact of conditions on the operators on the symbols themselves. For example, the requirement of boundedness of the operator restricts the symbol of the multiplication operator to be a bounded function, which leads to the collection of bounded multiplication operators over the Hardy space [39], whereas the only bounded analytic symbols for the Fock space are constant functions. Restricting to densely defined multiplication operators allows for a much broader collection of symbols for multiplication operators, which includes multivariate polynomials and many other functions for spaces like the exponential dot product and Gaussian radial basis function's native spaces. This is important in the present context, since the dynamics arise as symbols of multiplication operators, it allows for a broader collection of dynamics to be identified with this routine.

The focus of this chapter is to develop the method exhibiting a novel approach to nonlinear system identification with respect to *NAR* [5] using densely defined Multiplication operators and a new "kernel" function that represents an integration functional over a reproducing kernel Hilbert space dubbed an occupation kernel [50, 52].

3.2 Problem Description

A key assumption in most system identification efforts is that, only a few relevant terms define any unknown dynamics, given a dynamical system $\dot{x} = f(x)$, where x is a state variable and f, a dynamic constraint. We suppose that the unknown dynamics f lies in the linear span of a set of functions $\{f_j\}$. The objective is to determine corresponding weights $\{\theta_i\}$ of these terms that combine to produce the governing equation for the system. $f(\cdot) = \sum_j^N \theta_j f_j(\cdot)$, where θ_j are parameters. The unknown parameters in the model may be identified through a series of approaches commonly referred to as *parameter identification*. We develop a parameter identification technique using a version of the NAR model. Time series data with recurrent feedforward network free of exogenous inputs is explored. Here, we consider an autoregressive part of order τ_n and total moving average contribution of zero. This results in a Nonlinear AutoRegressive (NAR) model defined by

$$y(s) = f[y(s-\tau_1), y(s-\tau_2), \ldots, y(s-\tau_n)], \tag{3.1}$$

where $f[\cdot]$ is a non-linear function, $y(s) \in \mathbb{R}^m$ is the system output and $\tau_1, \tau_2, \ldots, \tau_n$ are time lags for the output. A large family of systems admit models that are in the form of equation (3.1) [24, 25].
In contrast to popular formulations of the identification problem, the function $f[\cdot]$ assumes the form of a symbol of a densely defined multiplication operator mapping between two pre-specified RKHSs. Subsequently, f is demonstrated to admit a decomposition in terms of an orthogonal set over a Hilbert space .

3.3 Preliminaries

Definition 3.3.1. A reproducing kernel Hilbert space (RKHS), $\mathcal{H}$, over a set X is a Hilbert space of real valued functions over the set X such that for all $x \in \mathcal{X}$ the evaluation functional $E_x g := g(x)$ is bounded. As such, the Riesz representation theorem guarantees, for all $x \in X$, the existence of a function $k_x \in \mathcal{H}$ such that $\langle g, k_x \rangle_{\mathcal{H}} = g(x)$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product for $\mathcal{H}$.

The function k_x is called the reproducing kernel function at x, and the function $k(x, y) = \langle k_y, k_x \rangle_{\mathcal{H}}$ is called the kernel function corresponding to $\mathcal{H}$.

Definition 3.3.2. Let $\mathcal{X} \subset \mathbb{R}^n$ be compact, and $\mathcal{H}$, a RKHS of continuous functions over $\mathcal{X}$. Let $\gamma : [0,T] \to \mathcal{X}$ be a continuous trajectory. The functional $g \longmapsto \int_0^T g(\gamma(t))dt$ is bounded over $\mathcal{H}$, and by Riesz representation theorem, may be represented as $\int_0^T g(\gamma(t))dt = \langle g, \Gamma_\gamma \rangle_{\mathcal{H}}$ for Some $\Gamma_\gamma \in \mathcal{H}$. The function Γ_γ is called the occupation kernel corresponding to γ in $\mathcal{H}$.

For a given trajectory $\gamma : [0,T] \to \mathcal{X}$, the occupation kernel with respect to the trajectory γ assumes the integral form shown in proposition (3.3.1) below.

Proposition 3.3.1. *Let $\mathcal{H}$ be a RKHS over a compact set $\mathcal{X}$ consisting of continuous functions and let $\gamma : [0,T] \to \mathcal{X}$ be a continuous trajectory as in Definition 3.4.3. The occupation kernel corresponding to γ ; Γ_γ in $\mathcal{H}$, may be expressed as*

$$\Gamma_\gamma(x) = \int_0^T k(x, \gamma(t))dt \tag{3.2}$$

Proof:

Notice that $\Gamma_\gamma(x) = \langle \Gamma_\gamma, K(\cdot, x) \rangle_H$, by the reproducing property of k. Consequently,

$$\begin{aligned} \Gamma_\gamma(x) &= \langle \Gamma_\gamma, k(\cdot, x) \rangle_H = \langle k(\cdot, x), \Gamma_\gamma \rangle_H \\ &= \int_0^T k(\gamma(t), x)dt = \int_0^T k(x, \gamma(t))dt \end{aligned}$$

which establishes the result.

Proposition (3.3.2) defines a road map to examining interactions between occupation kernel corresponding to a trajectory and the multiplication operator

Proposition 3.3.2. *Let $\mathcal{H}$ be a RKHS of continuous functions over a compact set $\mathcal{X} \subset \mathbb{R}^n$, and suppose $f : \mathbb{C} \to \mathbb{C}$ is Lipschitz continuous. If $\gamma : [0,T] \to \mathcal{X}$ is a trajectory as in Definition 3.3.2 above, then $\Gamma_\gamma \in D(M_\varphi^*)$ and*

$$M_\varphi^* [\Gamma_\gamma] (\cdot) = \int_0^T \overline{\varphi(\gamma(t))} k(\cdot, \gamma(t))dt \tag{3.3}$$

Proof. Given the reproducing kernel function centered at z, k_z, per the reproducing kernel property, we get that

$$
\begin{aligned}
M_{\varphi}^{*}\left[\Gamma_{\gamma}\right](z) &= \left\langle M_{\varphi}^{*}\Gamma_{\gamma}, k_{z}\right\rangle_{\mathcal{H}} = \left\langle \Gamma_{\gamma}, M_{\varphi}k_{z}\right\rangle_{\mathcal{H}} \\
&= \overline{\left\langle M_{\varphi}k_{z}, \Gamma_{\gamma}\right\rangle_{\mathcal{H}}} = \overline{\int_{0}^{T}\left(M_{\varphi}k_{z}\right)\left(\gamma(t)\right)dt} \\
&= \int_{0}^{T}\overline{\varphi(\gamma(t))k_{z}(\gamma(t))}dt = \int_{0}^{T}\overline{\varphi(\gamma(t))}k(z,\gamma(t))dt.
\end{aligned}
\tag{3.4}
$$

where $k_z(\gamma(t)) = k(\gamma(t), z) = \overline{k(z, \gamma(t))}$.
Thus $M_{\varphi}^{*}\Gamma_{\gamma}(z) = \int_{0}^{T}\overline{\varphi(\gamma(t))}k(z,\gamma(t))dt$ as desired. Over the real numbers, the inner product is symmetric and thus the conjugation in (3.4) is dropped □

We shall denote by S, the collection of all symbols corresponding to densely defined multiplication operator on a given RKHS. S admits a well-defined structure beginning with an inner product. Let $\psi, \varphi \in S$, define the bilinear form on the $S \times S$ by

$$
\langle \varphi, \psi \rangle_{S} := \left\langle M_{\psi}^{*}\Gamma_{\gamma}, M_{\varphi}^{*}\Gamma_{\gamma}\right\rangle_{\mathcal{H}}. \tag{3.5}
$$

defines an inner product on S conferred by the inner product on the RKHS $\mathcal{H}$

Proposition 3.3.3. *Equation* (3.5) *defines a inner product on* S.

Proof. Given that $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a well defined inner product on $\mathcal{H}$, the result follows directly □

For scalar-valued functions φ and ψ, the inner product can be explicitly expressed as

$$
\begin{aligned}
&\left\langle M_{\varphi}^{*}\Gamma_{\gamma}, M_{\psi}^{*}\Gamma_{\gamma}\right\rangle_{\mathcal{H}} = \left\langle M_{\psi}M_{\varphi}^{*}\Gamma_{\gamma}, \Gamma_{\gamma}\right\rangle = \int_{0}^{T}\left(M_{\psi}M_{\varphi}^{*}\Gamma_{\gamma}\right)\left(\gamma(t)\right)dt. \\
&= \int_{0}^{T}\left(M_{\psi}\int_{0}^{T}\overline{\varphi(\gamma(t))}k(\cdot,\gamma(t))dt\right)\left(\gamma(\tau)\right)d\tau \\
&= \int_{0}^{T}\int_{0}^{T}\left(M_{\psi}\overline{\varphi(\gamma(t))}k(\cdot,\gamma(t))dt\right)\gamma(\tau)d\tau \\
&= \int_{0}^{T}\int_{0}^{T}\psi(\gamma(\tau))\overline{\varphi(\gamma(t)}k(\gamma(\tau),\gamma(t))dtd\tau
\end{aligned}
$$

Suppose $\{e_m\}_{m=1}^{\infty}$ is an orthonormal basis set for $\mathcal{H}$, So that

$$K(x,y) = \sum_{m=1}^{\infty} e_m(x)\overline{e_m(y)} \tag{3.6}$$

The integral can be decomposed as

$$\langle \psi\varphi \rangle_S = \int_0^1 \int_0^1 \psi(\gamma(\tau))\overline{\varphi(\gamma(t))} \sum_{m=1}^{\infty} e_m(\gamma(\tau))\overline{e_m(\gamma(t))}dtd\tau$$

But $e_m \in \mathcal{H} \quad \forall m$ therefore

$$= \sum_{m=1}^{\infty} \int_0^T \varphi(\gamma(\tau))e_m(\gamma(\tau))d\tau \cdot \int_0^T \overline{\varphi(\gamma(t))} \quad \overline{e_m(\gamma(t))}dt \tag{3.7a}$$

$$= \sum_{m=1}^{\infty} \int_0^T \varphi(\gamma(\tau))e_m(\gamma(\tau))d\tau \cdot \int_0^T \overline{\varphi(\gamma(t))e_m(\gamma(t))}dt \tag{3.7b}$$

The integrals $\int_0^T \varphi(\gamma(\tau))e_m(\gamma(\tau))d\tau$ and $\int_0^T \overline{\varphi(\gamma(t))e_m(\gamma(t))}dt$ are both convergent for all m.

Furthermore, let's define a feature Ψ map on S by

$$\begin{aligned} \Psi :& S \to \mathbb{C} \\ & \varphi \mapsto \left[\int_0^T \varphi(\gamma(t))e_m(\gamma(t))dt\right]_{m=1}^{\infty} \end{aligned} \tag{3.8}$$

As such, the inner product expression in equation (3.7b) may be written

$$\langle \psi, \varphi \rangle_{S,\gamma} = \sum_{m=1}^{\infty} \Psi(\varphi)\overline{\Psi(\psi)} \tag{3.9}$$

If we take the perspective of the mapping $\varphi \mapsto M_{\varphi}^{*}\Gamma_{\gamma}$ as transformation and assume the kernel k is universal, it immediately follows that if $M_{\varphi}^{*}\Gamma_{\gamma} = M_{\psi}^{*}\Gamma_{\gamma}$ then ψ agrees with φ in the L^2 sense over $\gamma([0,T])$. Hence the mapping $\varphi \mapsto M_{\varphi}^{*}\Gamma_{\gamma}$ provides a generalized Fourier transform. We defer the exploration of this kernelized transform to Chapter 4.

Definition 3.3.3. Given a Hilbert space $\mathcal{Y}$ and a set $\mathcal{X}$, a vector valued reproducing kernel Hilbert space (vv-RKHS), $\mathcal{H}$, is a Hilbert space of functions mapping $\mathcal{X}$ to $\mathcal{Y}$, where for each $x \in \mathcal{X}$ the evaluation mapping $E_x : \mathcal{H} \to \mathcal{Y}$ given by $E_x(f) = f(x)$ is bounded. The operator valued kernel for a vector valued reproducing kernel Hilbert space is given by $k : \mathcal{X} \times \mathcal{X} \to \mathcal{B}(\mathcal{Y})$, $k(x, y) = E_x E_y^*$, here $\mathcal{B}(\mathcal{Y})$ denotes the bounded operators on $\mathcal{Y}$.

Boundedness of the mapping $E_x : \mathcal{H} \to \mathcal{Y}$ is equivalent to the boundedness of the functional $\mathcal{H} \ni g \mapsto \langle g(x), y\rangle_{\mathcal{Y}}$ for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The Riesz representation theorem guarantees for each $x \in \mathcal{X}$ and $v \in \mathcal{Y}$ the existence of a function $k_{x,v} \in \mathcal{H}$ such that $\langle g, k_{x,v}\rangle_{\mathcal{H}} = \langle g(x), v\rangle_{\mathcal{Y}}$ for all $g \in \mathcal{H}$. In general, for a vector valued reproducing kernel Hilbert space $\mathcal{H}$ with kernel k, we can define $k_x v(\cdot) := k(\cdot, x)v \in \mathcal{H}$. We note that $k_{x,v} = k_x v = E_x^*(v)$ since

$$(E_x^*(v))(y) = E_y E_x^*(v) = k(y, x)v = (k_x v)(y)$$

for all y and the reproducing property is given by

$$\langle f(x), v\rangle_{\mathcal{Y}} = \langle E_x(y), v\rangle_{\mathcal{Y}} = \langle f, E_x^*(v)\rangle_{\mathcal{H}} = \langle f, k_{x,v}\rangle_{\mathcal{H}}.$$

As an example of kernel for a vv-RKHS, we let P be a positive definite matrix and $k(x, y)$ a scalar valued kernel. The function $K(x, y) = k(x, y)P$ defines a kernel for a vv-RKHS. We recall that a RKHS is uniquely determined by its kernel function.

3.4 Multiplication Operators and Occupation Kernels

As opposed to the definition we saw in chapter 2, mutatis mutandis, a slightly different definition of the multiplication operator is presented here. It is important to note that the definition of a multiplication operator in this setting is ambiguous without specifying how the products act on the arguments. In order to ensure the adjoint relationships hold as written, we will make a particular choice in how we define the multiplication operator.

Definition 3.4.1. Let $\mathcal{H}$ be a scalar-valued RKHS over a set $\mathcal{X} \subset \mathbb{R}^n$ and $\tilde{\mathcal{H}}$ be a $\mathbb{R}^k$-valued RKHS over $Y \subset (\mathbb{R}^n)^{p+1}$ for $k \in \mathbb{N}$. For a measurable function of the form $\varphi : (\mathbb{R}^n)^p \to \mathbb{R}^k$, define the

multiplication operator with symbol φ,

$$M_\varphi : D\left(M_\varphi\right) \subset \mathcal{H} \to \tilde{\mathcal{H}} \quad \text{by} \quad M_\varphi(g) = h$$

where

$$h\left(x_1, x_2, \ldots, x_p, x_{p+1}\right) := g\left(x_1\right) \varphi\left(x_2, \ldots, x_p, x_{p+1}\right)$$

and

$$D(M_\varphi) = \{g \in \mathcal{H} : M_\varphi g = \varphi g \in \tilde{\mathcal{H}}\}.$$

Definition 3.4.2. Let $\mathcal{H}$ and $\tilde{\mathcal{H}}$ be two RKHSs and T be a linear operator whose domain $D(T) \subset \mathcal{H}$ and $T : D(T) \to \tilde{\mathcal{H}}$. T is *densely defined* in $\mathcal{H}$ if $D(T)$ is a dense subspace of $\mathcal{H}$ (cf. [39, Chapter 5]).

Given T and $D(T)$ as above, the *adjoint* of a densely defined (possibly unbounded) operator T has the domain $D(T^*) := \{g \in \tilde{\mathcal{H}} \mid h \mapsto \langle Th, g, \rangle_{\mathcal{H}}$ is bounded over $D(T)\}$; refer [39].
The operator T is *closed* if for every sequence $\{g_m\}_{m=0}^{\infty}(T)$ such that $g_m \to g \in \mathcal{H}$ and $Tg_m \to h \in \mathcal{H}$ then $g \in D(T)$ and $Tg = h$.

Lemma 3.4.1. *The adjoint of a closed densely defined operator over a Hilbert space is closed and densely defined.*

Proof. Refer [39, Chapter 5]. □

Occupation kernels, introduced by *Rosenfeld et al* in [48], have been used for a variety of data driven methods in dynamical systems [50, 51, 52]. A variant formulation of the occupation kernel presented in the original manuscript is adopted for the present context. This modification ensues that it is suitable for vector-valued symbols as used here. In addition, we present a vector-valued function formulation of Propositions 3.3.2 and 3.3.3

Definition 3.4.3. Let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Y} := \mathcal{X}^{p+1} \subset (\mathbb{R}^n)^{p+1}$, and $\tilde{\mathcal{H}}$ be RKHS of continuous functions over $\mathcal{Y}$. Let $\tau_1 < \tau_2 < \ldots < \tau_p$ be real numbers. Let $\gamma : [-\tau_p, T] \to \mathcal{X}$ be a continuous trajectory. The functional $g \mapsto \int_0^T g\left(\gamma(t), \gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)\right) dt$ is bounded over $\tilde{\mathcal{H}}$, and by the Riesz representation theorem, can be represented as

$$\int_0^T g\left(\gamma(t), \gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)\right) dt = \langle g, \Gamma_\gamma \rangle_{\tilde{\mathcal{H}}}$$

for some $\Gamma_\gamma \in \tilde{\mathcal{H}}$. The function Γ_γ is called the occupation kernel corresponding to γ in $\tilde{\mathcal{H}}$.

Proposition 3.4.2. *Let $\mathcal{H}$ and $\tilde{\mathcal{H}}$ be RKHSs of continuous functions over the sets $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} := (\mathcal{X})^{p+1} \subset (\mathbb{R}^n)^{p+1}$. Suppose $f : (\mathcal{X})^p \to \mathbb{R}$ is Lipschitz continuous and a symbol of a densely defined multiplication operator, $M_f : D(M_f) \to \tilde{\mathcal{H}}$, given as $M_f g = gf$ for $g \in D(M_f)$. Let $\gamma : [-\tau_p, T] \to \mathcal{X}$ be a continuous trajectory, then $\Gamma_\gamma \in D(M_f^*)$ and for $x \in \mathcal{X}$*

$$M_f^* \Gamma_\gamma(x) = \int_0^T f(\gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)) k(x, \gamma(t)) dt \tag{3.10}$$

Proof. : Note,

$$\begin{aligned} M_f^* \Gamma_\gamma(x) &= \langle M_f^* \Gamma_\gamma, k_x \rangle_{\mathcal{H}} \\ &= \langle \Gamma_\gamma, M_f k_x \rangle_{\tilde{\mathcal{H}}} \\ &= \int_0^T f(\gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)) k(x, \gamma(t))\, dt \end{aligned}$$

□

Proposition 3.4.3. *Let ψ and φ be symbols for densely defined multiplication operators, and $\gamma_1, \gamma_2 : [-\tau_p, T] \to \mathcal{X}$ be trajectories as in Definition 3.4.3. Suppose $c_1, c_2 \in \mathbb{R}$, then*

$$M^*_{(c_1\psi + c_2\varphi)} \Gamma_\gamma = M^*_{c_1\psi} \Gamma_\gamma + M^*_{c_2\varphi} \Gamma_\gamma, \text{ and}$$

$$M^*_{c\varphi} \Gamma_\gamma = c M^*_\varphi \Gamma_\gamma$$

$$M^*_\varphi \left(c_1 \Gamma_{\gamma_1} + c_2 \Gamma_{\gamma_2}\right) = c_1 M^*_\varphi \Gamma_{\gamma_1} + c_2 M^*_\varphi \Gamma_{\gamma_2}$$

Proof. The proof of this proposition follows from the definition, i.e. let $\{c_i\}_{i=1}^N \subset \mathbb{R}$ be a set of real constants, and $\{\varphi_i\}_{i=1}^N$ be a collection of symbols of densely defined multiplication operators such that

equation (3.10) holds for all $t \in [-\tau_p, T]$. It follows that

$$\begin{aligned} M^*_{\left(\sum_{i=1}^N c_i\varphi_i\right)}\Gamma_\gamma &= \int_0^T \left(\sum_{i=1}^N c_i\varphi_i\right)(\gamma(t-\tau_1),\gamma(t-\tau_2),\ldots,\gamma(t-\tau_p))K(\cdot,\gamma(t))dt \\ &= \sum_{i=1}^N c_i \int_0^T \varphi_i(\gamma(t-\tau_1),\gamma(t-\tau_2),\ldots,\gamma(t-\tau_p))K(\cdot,\gamma(t))dt \\ &= \sum_{i=1}^N c_i M^*_{\varphi_i}\Gamma_\gamma \end{aligned}$$

Suppose γ_1, γ_2 are continuous trajectories satisfying Proposition 3.4.2, then we have that

$$M^*_\varphi\left(\Gamma_{\gamma_1} + \Gamma_{\gamma_2}\right) = M^*_\varphi\Gamma_{\gamma_1} + M^*_\varphi\Gamma_{\gamma_1}$$

since $M^*_\varphi\Gamma_{\gamma_1}$ is linear □

We shall denote by S the collection of symbols corresponding to densely defined multiplication operators on the RKHS $\mathcal{H}$ of continuous functions as described above.

For $\psi, \varphi \in S$, with $\psi, \varphi : (\mathbb{R}^n)^p \to \mathbb{R}^n$, we define a bi-linear form over S by

$$\begin{aligned} \langle\cdot,\cdot\rangle_S &: S \times S \to \mathbb{R} \\ \langle\psi,\varphi\rangle_S &:= \left\langle M^*_\varphi\Gamma_\gamma, M^*_\psi\Gamma_\gamma\right\rangle_{\mathcal{H}}. \end{aligned} \tag{3.11}$$

Henceforth, we shall adopt the following notation for convenience,.

$$\gamma^{[t]} := (\gamma(t-\tau_1),\gamma(t-\tau_2),\ldots,\gamma(t-\tau_p)).$$

Proposition 3.4.4. *Equation* (3.11) *defines an inner product on* S/Λ, *where* $\Lambda := \{\varphi \in S : \langle\varphi,\varphi\rangle_S = 0\}$. *Moreover, if* $\{e_m\}_{m=1}^\infty \subset \mathcal{H}$ *is an orthonormal basis for* $\mathcal{H}$, *then*

$$\begin{aligned} \langle\psi,\varphi\rangle_S &= \int_0^T\int_0^T \psi(\gamma^{[\tau]})\varphi(\gamma^{[t]})K(\gamma(\tau),\gamma(t))dtd\tau \\ &= \sum_{m=1}^\infty \int_0^T \psi(\gamma^{[\tau]})e_m(\gamma(\tau))d\tau \cdot \int_0^T \varphi(\gamma^{[t]})e_m(\gamma(t))dt, \end{aligned}$$

Proof. Linearity of the transform with respect to the symbols φ and ψ expressed in Proposition 3.4.3 together with the linearity of the scalar product on $\mathcal{H}$ establishes the necessary linearity properties of $\langle \cdot, \cdot \rangle_S$ as a pre-inner product on S. Non-negativity of our pre-inner product $\langle \varphi, \varphi \rangle$ for all symbols φ is established by noting $\langle \varphi, \varphi \rangle_S = \langle M_\varphi^* \Gamma_\gamma, M_\varphi^* \Gamma_\gamma \rangle_{\mathcal{H}}$. Modding out by Λ and taking $\langle \varphi + \Lambda, \psi + \Lambda \rangle_{S/\Lambda} = \langle \hat{\varphi}, \hat{\psi} \rangle_S$ with $\hat{\varphi} \in \varphi + \Lambda$ and $\hat{\psi} \in \psi + \Lambda$, the pre-inner product can be expressed as an inner product over S/Λ by standard arguments (cf. [39]).

By definition, $g \mapsto \langle g, M_\psi^* \Gamma_\gamma \rangle_{\mathcal{H}}$ is a bounded functional on $\mathcal{H}$, extended off of the dense set $D(M_\psi)$ to all of $\mathcal{H}$, and given as

$$g \mapsto \int_0^T g(\gamma(t)) \psi(\gamma^{[t]}) dt, \text{ for all } g \in \mathcal{H}, \tag{3.12}$$

where $\gamma^{[t]} = (\gamma(t-\tau_1), \gamma(t-\tau_2), \ldots, \gamma(t-\tau_p))$. Hence, $\left\langle M_\varphi^* \Gamma_\gamma, M_\psi^* \Gamma_\gamma \right\rangle_{\mathcal{H}}$ is expressible as the functional (3.12) applied to $M_\varphi^* \Gamma_\gamma$. Thus,

$$\begin{aligned} \left\langle M_\varphi^* \Gamma_\gamma, M_\psi^* \Gamma_\gamma \right\rangle_{\mathcal{H}} &= \int_0^T \psi(\gamma^{[\tau]}) \int_0^T \varphi(\gamma^{[t]}) K(\gamma(\tau), \gamma(t)) dt d\tau \\ &= \int_0^T \int_0^T \psi(\gamma^{[\tau]}) \varphi(\gamma^{[t]}) K(\gamma(\tau), \gamma(t)) dt d\tau. \end{aligned}$$

□

This formulation relies on the fortunate coincidence that the dynamics are scalar-valued. Though the experiments in section 3.6 use scalar-valued symbols, their vector valued counterparts are addressed below for completeness.

3.4.1 Vector-Valued Symbols

There are natural questions that lead to an equivalent formulation involving other forms of the symbol of the multiplication operator. In many practical applications, governing equations are best represented as vector valued functions, such that

$$\begin{aligned} x(t) &= f\left(x(t-\tau_1), x(t-\tau_2), \ldots, x(t-\tau_p)\right) \\ f &: \mathbb{R}^{n \times p} \to \mathbb{R}^k. \end{aligned} \tag{3.13}$$

In such scenarios, it is useful that the RKHS in which the operator embeds the symbol is chosen to be vector-valued.

Let $\mathcal{H}$ be a scalar valued RKHS over a set $\mathcal{X} \subset \mathbb{R}^n$ and $\tilde{\mathcal{H}}$ a $\mathbb{R}^k$-valued RKHS over the set $(\mathcal{X})^{p+1}$ as in Definitions 3.3.1 and 3.3.3 respectively. In this section we consider the subset of all multiplication operators equipped with a vector-valued symbols f that maps a scalar function $g \in \mathcal{H}$ into $\tilde{\mathcal{H}}$ as given in Definition 3.4.1. Given a monotonic sequence $\tau_1 < \tau_2 < \tau_3 < \cdots < \tau_p$, and a continuous trajectory γ, if $h \in \tilde{\mathcal{H}}$, the mapping

$$h \mapsto \left\langle \int_0^T h(\gamma(t), \gamma(t-\tau_1), \dots \gamma(t-\tau_p))dt, v \right\rangle_{\mathbb{R}^k}$$

is bounded for all $v \in \mathbb{R}^k$. As such, there exists a function $\Gamma_{\gamma,v}$ such that

$$\begin{aligned}\langle h, \Gamma_{\gamma,v}\rangle_{\tilde{\mathcal{H}}} &= \left\langle \int_0^T h(\gamma(t), \gamma(t-\tau_1), \dots \gamma(t-\tau_p))dt, v \right\rangle_{\mathbb{R}^k} \\ &= \int_0^T h(\gamma(t), \gamma(t-\tau_1), \gamma(t-\tau_2), \dots, \gamma(t-\tau_p))^\intercal v \, dt\end{aligned}$$

where $\intercal$ is the transpose operator. If $g \in \mathcal{D}(M_f)$, then

$$\langle M_f g, \Gamma_{\gamma,v}\rangle_{\tilde{\mathcal{H}}} = \int_0^T g(\gamma(t)) f(\gamma(t-\tau_1), \gamma(t-\tau_2), \cdots, \gamma(t-\tau_p))^\intercal v \, dt. \tag{3.14}$$

Additionally for a kernel function $K_x \in \mathcal{H}$,

$$\left\langle M_f^* \Gamma_{\gamma,v}, K_x \right\rangle_{\mathcal{H}} = \langle \Gamma_{\gamma,v}, M_f K_x\rangle_{\tilde{\mathcal{H}}} = \int_0^T K(\gamma(t), x) f(\gamma^{[t]})^\intercal v \, dt.$$

For two symbols $F, G : (\mathbb{R}^n)^p \to \mathbb{R}^k$ and vectors $v, w \in \mathbb{R}^k$ we have,

$$\begin{aligned}\langle M_G^* \Gamma_{\gamma,w}, M_F^* \Gamma_{\gamma,v}\rangle_{\mathcal{H}} &= \int_0^T \left(M_G^* \Gamma_{\gamma,w}\right) \gamma(t) F(\gamma^{[t]})^\intercal v dt \\ &= \int_0^T \left(\int_0^T K(\gamma(\tau), \gamma(t)) G(\gamma^{[\tau]})^\intercal w \, d\tau \right) F(\gamma^{[t]})^\intercal v \, dt \\ &= \int_0^T \int_0^T K(\gamma(\tau), \gamma(t)) \left(G(\gamma^{[\tau]})\right)^\intercal w \left(F(\gamma^{[t]})\right)^\intercal v \, d\tau dt \\ &= \int_0^T \int_0^T K(\gamma(\tau), \gamma(t)) \left(G(\gamma^{[\tau]})\right)^\intercal w v^\intercal F(\gamma^{[t]}) \, d\tau dt.\end{aligned} \tag{3.15}$$

For a finite basis $\{v_r\}_{r=1}^k$ of $\mathbb{R}^k$, we define the scalar product

$$\langle G, F\rangle_S = \sum_{r=1}^{k} \langle M_G^* \Gamma_{\gamma,v_r}, M_F^* \Gamma_{\gamma,v_r}\rangle_{\tilde{\mathcal{H}}}\,. \tag{3.16}$$

The scalar product under this definition depends on the choice of trajectory γ and basis $\{v_r\}_{r=1}^k$. Moreover, this formula enforces non-negativity of the inner product when F is equal to G.

3.5 System Identification

It is assumed that the unknown system dynamics are a symbol for a densely defined multiplication operator from $\mathcal{H}$ to $\tilde{\mathcal{H}}$, where $\mathcal{H}$ can be infinite dimensional. The algorithm seeks to project the dynamics onto a subspace $\mathcal{K} \subset S$ spanned by a finite collection of basis elements, thereby finding the best approximation of f with respect to functions in $\mathcal{K}$ according to the metric induced by the inner product on S/Λ. This is achieved by leveraging properties of the action of the adjoint multiplication operator on an occupation kernel, $M_f^*\Gamma_\gamma$. Within this setup, $M_f^*\Gamma_\gamma$ is then decomposed into a linear combination of actions involving basis elements as $M_{Y_i}^*\Gamma_\gamma$, where $\gamma : [-\tau_p, T] \to \mathcal{X}$ is an observed trajectory of the system serving as data units. Parameters are estimated through a constrained optimization problem by virtue of the induced norm on the symbols.

Let $\{Y_i\}_{i=1}^M$ be a collection of basis functions such that $f = \sum_{i=1}^M \theta_i Y_i$. Thus, satisfies the dynamics

$$\begin{aligned}\gamma(t) =& f(\gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)) \\ =& \sum_{i=1}^{M} \theta_i Y_i(\gamma(t-\tau_1), \ldots, \gamma(t-\tau_p)),\end{aligned} \tag{3.17}$$

where $\gamma : [-\tau_p, T] \to \mathcal{X} \subset \mathbb{R}^n$, and $Y_i : (\mathbb{R}^n)^p \to \mathbb{R}^k$. In the context of *gray box* system identification, given the parameterizing basis above, we seek to estimate the coefficients $\theta_i, i = 1, 2, \ldots M$ so that the dynamics in (3.17) may be used to reproduce data points defined in terms of system trajectories

$$\gamma(t) = f(\gamma(t-\tau_1), \gamma(t-\tau_2), \cdots, \gamma(t-\tau_p)).$$

Suppose f and Y_i corresponds to symbols of densely defined multiplication operators over $\mathcal{H}$ for all i. Given $T \in \mathbb{R}$, let $\gamma : [-\tau_p, T] \to \mathbb{R}^n$ be a specified system measurement as in Proposition $3.4.2$ and

$g \in D(M_f)$. The *unknown* parameters θ_i are sought out by minimizing deviation of the approximations from the actual dynamics. That is, find $\{\theta_i\}$ that minimize

$$\min_{\theta} \|f - \sum_{i=1}^{M} \theta_i Y_i\|_S^2 \tag{3.18}$$

where $Y_i = Y_i(\gamma(t-\tau_1), \gamma(t-\tau_2), \cdots, \gamma(t-\tau_n))$. Note that

$$\left\| f - \sum_{i=1} \theta_i Y_i \right\|_S^2 = \|f\|_S^2 - 2\sum_{i=1} \theta_i \langle f, Y_i \rangle + \sum_{i=1}\sum_{j=1} \theta_i \theta_j \langle Y_i, Y_j \rangle.$$

The right side of above can be explicitly written as follows:

$$\|f\|_S^2 - 2\Theta^\top \begin{bmatrix} \langle f, Y_1 \rangle \\ \langle f, Y_2 \rangle \\ \vdots \\ \langle f, Y_M \rangle \end{bmatrix} + \Theta^\top \begin{bmatrix} \langle Y_1, Y_1 \rangle & \langle Y_1, Y_2 \rangle & \dots & \langle Y_1, Y_M \rangle \\ \langle Y_2, Y_1 \rangle & \langle Y_2, Y_2 \rangle & \dots & \langle Y_2, Y_M \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle Y_M, Y_1 \rangle & \langle Y_M, Y_2 \rangle & \dots & \langle Y_M, Y_M \rangle \end{bmatrix} \Theta \tag{3.19}$$

where $\Theta = [\theta_1, \theta_2, \theta_3, \dots, \theta_M]^\intercal$. Setting the gradient of (3.19) to 0 and solving for Θ, we get

$$\Theta = \begin{bmatrix} \langle Y_1, Y_1 \rangle & \langle Y_1, Y_2 \rangle & \dots & \langle Y_1, Y_M \rangle \\ \langle Y_2, Y_1 \rangle & \langle Y_2, Y_2 \rangle & \dots & \langle Y_2, Y_M \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle Y_M, Y_1 \rangle & \langle Y_M, Y_2 \rangle & \dots & \langle Y_M, Y_M \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle f, Y_1 \rangle \\ \langle f, Y_2 \rangle \\ \vdots \\ \langle f, Y_M \rangle \end{bmatrix}.$$

The inner product between basis pair $\langle Y_i, Y_j \rangle_{\tilde{\mathcal{H}}}$ is computed as

$$\langle Y_i, Y_j \rangle_S = \sum_{r=1}^{k} \int_0^T \int_0^T K(\gamma(\tau), \gamma(t)) \left(Y_i(\gamma^{[\tau]}) \right)^\intercal v_r v_r^\intercal \left(Y_j(\gamma^{[t]}) \right) d\tau dt.$$

Likewise, under the assumption that $\gamma(t) = f(\gamma^{[t]})$,

$$\begin{aligned} \langle f, Y_i \rangle_S &= \sum_{r=1}^{k} \int_0^T \int_0^T K(\gamma(\tau), \gamma(t)) \left(f(\gamma^{[\tau]}) \right)^\intercal v_r v_r^\intercal \left(Y_j(\gamma^{[t]}) \right) d\tau dt \\ &= \sum_{r=1}^{k} \int_0^T \int_0^T K(\gamma(\tau), \gamma(t)) \left(\gamma(\tau) \right)^\intercal v_r v_r^\intercal \left(Y_j(\gamma^{[t]}) \right) d\tau dt. \end{aligned}$$

Note that if we exclude the basis $\{v_r\}_{r=1}^{k}$ the scalar case follows.

In the event that the target function f lies in the linear span of the chosen basis functions Y_i and the Gram matrix $[\langle Y_i, Y_j \rangle_S]_{i,j=1,1}^{M,M}$ is positive definite, then f will be identified as it is the unique solution to the minimization problem. Outside of this scenario, the result will be the best approximation in the norm $\|\cdot\|_S$.

3.6 Numerical Experiments

The identification approach introduced in this chapter is centered around a new implementation within the NAR framework. Verification of the developed model is carried out through a series of numerical experiments, two of which are presented below. In particular, this work is not concerned with drawing a comparison with existing methods. However, the effectiveness of the model in identifying system parameters is evident in these experiments. In both experiments, RBF and exponential dot product kernels were used and performance of the model with respect to each kernel function is accessed. A kernel width of $\mu = 0.01$ is used, step sizes of 0.01 and 0.05 is employed for numerical integration in experiments 1 and 2 respectively. Three integer lags $\tau_1 = 1, \tau_2 = 3, \tau_3 = 3$ were utilized. With the identified parameters, candidate input values are sampled over a pre-selected interval to carry out prediction. While there is no unique way of selecting the predictor values, we rely on central tendencies of the training data to sample these inputs.

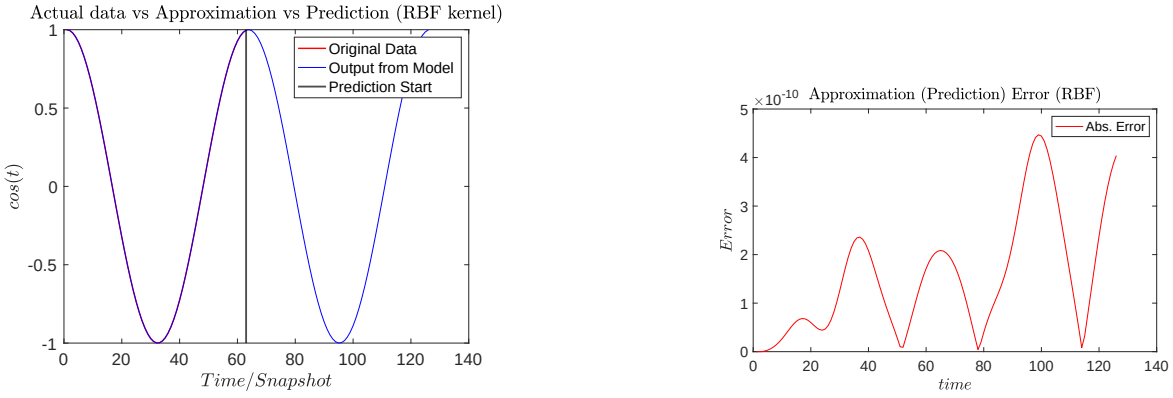
Experiment 1:

In experiment 1, simulated data from the cosine function is used to train the model and generate system parameters. The function $\cos(t)$ is discretized according to a step size h over interval $[0, 2\pi]$, the data is split into two parts, the first half is used to train the model and the second half for prediction. Basis functions are constructed as tensor products of monomials of up to degree three (3), that is $Y_i = x_1^{i_1} x_2^{i_2} x_3^{i_3}$, for nonnegative integers i_1, i_2, i_3 so that $0 \leq i_1 + i_2 + i_3 \leq 3$. With $n = 1, p = 3$, the approximation is such that

$$\cos(t) \approx \sum_{i=1} \theta_i Y_i(\gamma(t-\tau_1), \gamma(t-\tau_2), \gamma(t-\tau_3))$$

where

$$Y_i(\gamma(t-\tau_1), \gamma(t-\tau_2), \gamma(t-\tau_3)) = (\gamma(t-\tau_1))^{i_1} (\gamma(t-\tau_2))^{i_2} (\gamma(t-\tau_3))^{i_3}.$$



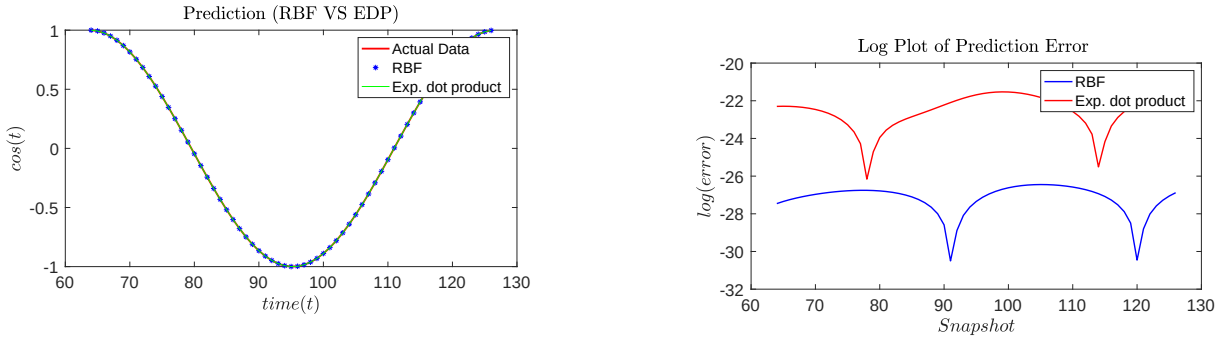
(a) **(b)**

Figure 10. Experiment 1: (a) Actual data (red) over $[0, 2\pi]$ versus approximation by model (blue) over training interval (left of gray line) followed by prediction by model (right of gray line). (b) Absolute pointwise error for Approximation and prediction associated with the trained model using RBF.

Figure 10a shows the approximation of the function values over $[0, 2\pi]$ and the prediction values by the model over $[2\pi, 4\pi]$. Figure 10b shows the associated approximation error. Prediction over $[2\pi, 4\pi]$ is carried out by interchanging between the two kernel functions. Figure 11a shows both predictions involving RBF and exponential dot product kernels whiles Figure 11b is a plot of the associated absolute pointwise error.

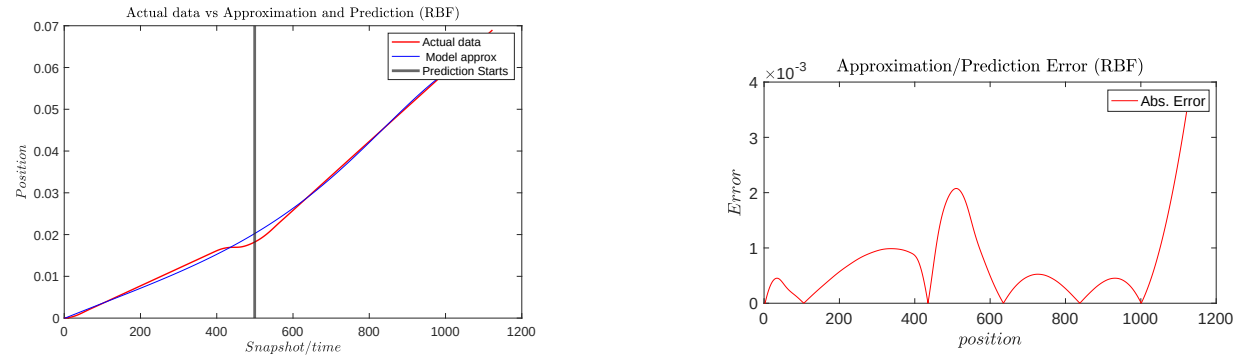
Experiment 2:

In Experiment 2, time series data measurements from an Electro-Mechanical Positioning System (EMPS) [19]; a standard configuration of a drive system for prismatic joint of robots or machine tools. The data comprises measurements of joint positions q_m and control signal at a sampling frequency of $1kHz$ with a duration of approximately 25 seconds. The primary goal of the experiment is to use the trained model to reproduce the recorded one-dimensional position data as well as predict positions over an extended interval. Using a subset of q_m comprising 1000 data points (snapshots), the data is divided into 2 parts for training and prediction respectively. The same basis collection as in Experiment 1 was adopted. Figure 12a compares the true data with approximations as well as prediction over adjacent interval. Next, Figure 13a shows prediction curves of the model using RBF and the exponential dot product while Figure 13b indicates their corresponding absolute pointwise error associated with the prediction.



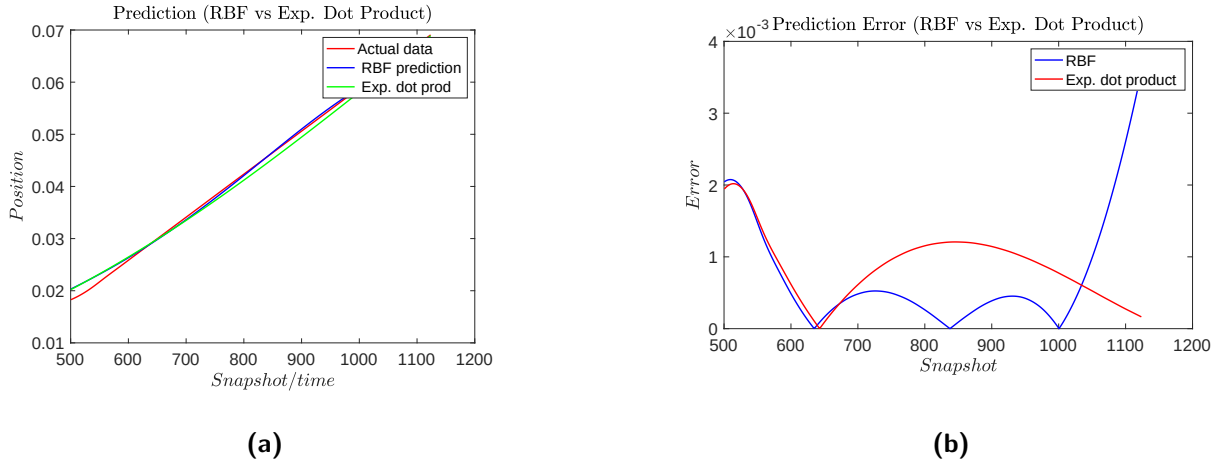
(a) **(b)**

Figure 11. Experiment 1: (a) Predicted pattern over $[2\pi, 4\pi]$ using trained model with RBF (blue) and exp. dot product kernel (red). (b) Log plot of associated prediction error.



(a) **(b)**

Figure 12. Experiment 2: (a) EMPS joint position measurements: Actual position data (red) versus approximation (blue) over training interval (to the left of gray line). Predicted time series data using trained model (to the right of gray line). (b) Associated absolute pointwise error over prediction interval using RBF.



(a) **(b)**

Figure 13. Experiment 2: (a) EMPS joint position measurements: Actual position data (red) versus predictions using RBF (blue) and exponential dot product kernel(green). (b) Associated approximation errors over prediction interval for RBF (blue) and exp. dot product kernel (red).

CHAPTER 4:
KERNELIZED FOURIER TRANSFORM

4.1 Introduction

As we alluded to in Chapters 2 and 3, given the Hilbert spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$, it is known that, if $M_\varphi : \mathcal{H} \to \tilde{\mathcal{H}}$ is a bounded multiplication operator, then it admits a well-defined adjoint $M_\varphi^* : \tilde{\mathcal{H}} \to \mathcal{H}$. In addition, if M_φ is densely defined, M_φ^* is in turn densely defined. Moreover, the occupation kernel Γ_γ associated with the system trajectory γ resides in the domain of the adjoint multiplication operator; i.e., $\Gamma_\gamma \in D(M_\varphi^*) \subset \tilde{\mathcal{H}}$. For a fixed symbol φ, the action $M_\varphi^*\Gamma_\gamma$ of the adjoint operator on the occupation kernel is a well-defined kernelized-transform over the underlying set $\mathcal{X}$ of the Hilbert space. This integral transform shares theoretical properties with the Fourier transform and Hilbert Schmidt integral transforms in general. $M_\varphi^*\Gamma_\gamma$ can also be viewed as an operator acting on the set S of the symbols of densely defined multiplication operator $S \mapsto \left[M_\varphi^*\right]\Gamma_\gamma$. Where appropriate, we shall conveniently interchange between a compact subspace $\mathcal{X} \subset \mathbb{R}^n$ (equivalently $\mathcal{X} \subset \mathbb{C}^n$) and S for our analysis. We shall assume that the underlying Hilbert spaces are separable unless otherwise stated. Elements of the class S of symbols of densely defined multiplication operators shall be simply referred to as "symbols" unless there is ambiguity. At the very least, every symbol is continuous on its domain. For all integrals in the plane or on the unit disk, the mean modulus $\int |f| ds$ and $\int |f| dxdy$ shall be though of as length or area respectively in the the Lesbesgue sense. Similar notations will be $\int |f| d\mu(y)$ and $\int\int |f| dA(x,y)$, as Lebesgue area measure where appropriated. Most of the analysis in this chapter is done in the Hardy space over the units disk. Otherwise, all domains shall be considered to be simply connected.

This chapter investigates the developed transform in light of similar integral transforms. It studies function-operator theoretic properties of the transform in the Hardy space. In the first section we look at the transform as a function on the set $\mathcal{X}$, and later as an operator Λ on the set S of densely defined symbols over L^2 and gradually progress to $\mathcal{H}^2$. Generally, the kernel function for the integral

operator differs from that of an RKHS, while the later can occur as an integral kernel, although there are distinctions worth noting.

4.2 Background

In addition to literature from the previous chapters, supplemental information and theorems are presented in this section which we shall periodically reference.

Definition 4.2.1. (Hilbert-Schmidt Operator [18]) A bounded linear operator K on a separable Hilbert space $\mathcal{H}$ is said to be Hilbert-Schmidt if for any orthonormal basis of $\mathcal{H}$, $\{e_i \mid i \in \mathbb{N}\}$

$$\sum_{i=1}^{\infty} \|Ke_i\|^2 < \infty.$$

If K is a Hilbert-Schmidt operator, then

$$\|K\|_{HS} := \sqrt{\sum_{i=1}^{\infty} \|Ke_i\|^2}$$

is called the Hilbert-Schmidt norm of K.

Definition 4.2.2. The integral operator $K : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ is defined by

$$K\square := \int_{\mathcal{X}} k(x,y)\square(y)dy \tag{4.1}$$

where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is called the kernel of the integral operator

Theorem 4.2.1. *[17] Let $\mathcal{X} \subset \mathbb{R}^n$, and $k(x,y) \in L^2(\mathcal{X} \times \mathcal{X})$. Then for every $f \in L^2(\mathcal{X})$, then the integral operator*

$$(Tf)(x) = \int_{\mathcal{X}} k(x,y)f(y)dy \tag{4.2}$$

is Hilbert-Schmidt and its Hilbert-Schmidt norm is

$$\|T\|_{HS} = \left(\int_{\mathcal{X}\times\mathcal{X}} |k(x,y)|^2 dxdy \right)^{1/2} \tag{4.3}$$

Proof. Refer to [17] □

Let k be a continuous kernel function on the product space $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$, if k is Hilbert-Schmidt, that is

$$\int_{\mathcal{X}} \int_{\mathcal{X}} |k(x,y)|^2 dxdy = C < \infty, \tag{4.4}$$

then the associated integral operator

$$Kf = \int_{\mathcal{X}} k(x,y) f(y) dy$$

is continuous (hence bounded) and compact in the norm topology of $\mathcal{B}(\mathcal{H})$; the Banach space of all linear operators on $\mathcal{H}$. Let $\mathcal{X}$ be a measure space and $L^2(\mathcal{X})$ indicate the Hilbert space of real-valued square-integrable functions on $\mathcal{X}$. suppose $k \in L^2(\mathcal{X} \times \mathcal{X})$ and consider the corresponding Hilbert-Schmidt operator

$$\begin{aligned} K: \quad & L^2(\mathcal{X}) \to L^2(\mathcal{X}) \\ & Kf(x) = \int_{\mathcal{X}} k(x,y) f(y) dy \end{aligned} \tag{4.5}$$

If k is symmetric; i.e., $k(x,y) = k(y,x)$, then by Fubini's theorem

$$\langle f, Kg \rangle = \int_{\mathcal{X}} f(x) \int_{\mathcal{X}} k(x,y) g(y) dy dx = \int_{\mathcal{X}} \int_{\mathcal{X}} k(y,x) f(x) dx g(y) dy = \langle Kf, g \rangle$$

This indicates that, symmetric kernels k correspond to self-adjoint Hilbert-Schmidt operators. Lemma (4.2.2) formulates this in a more general context,

Lemma 4.2.2. *Suppose $(\mathcal{X}, \Omega, \mu)$ is a σ-finite measure space, k a kernel function on $\mathcal{X} \times \mathcal{X}$ for some integral operator. Define the operator $L : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ by*

$$(L)f = \int_{\mathcal{X}} k(x,y) f(y) dy \tag{4.6}$$

Then L is bounded and has unique adjoint operator $L^* : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ given by

$$L^* f(x) = \int_X \overline{k(y,x)} f(y) d\mu(y).$$

Proof. Boundedness follows immediately. To find L^*, we let $A : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ be the integral operator with kernel $\overline{k(y,x)}$, i.e.,

$$Af(x) = \int_X \overline{k(y,x)} f(y) d\mu(y).$$

Then, given any f and $g \in L^2(\mathcal{X})$, we have

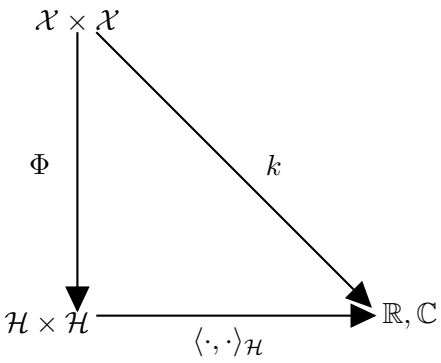
$$\begin{aligned}
\langle f, L^* g \rangle = \langle Lf, g \rangle &= \int_{\mathcal{X}} Lf(x) \overline{g(x)} d\mu(x) \\
&= \int_{\mathcal{X}} \int_X k(x,y) f(y) d\mu(y) \overline{g(x)} d\mu(x) \\
&= \int_{\mathcal{X}} f(y) \int_X k(x,y) \overline{g(x)} d\mu(x) d\mu(y) \\
&= \int_{\mathcal{X}} f(y) \overline{\int_X \overline{k(x,y)} g(x) d\mu(x)} d\mu(y) \\
&= \int_{\mathcal{X}} f(y) \overline{Ag(y)} d\mu(y) \\
&= \langle f, Ag \rangle.
\end{aligned}$$

By uniqueness of the adjoint, we have $L^* = A$. □

Additionally, a continuous kernel k is said to satisfy the *Mercer condition* if an only if

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k(x,y) f(x) f(y) dx dy \geq 0 \quad \text{fo all} \quad f \in L^2(\mathcal{X}) \tag{4.7}$$

In relation to RKHSs, recall that given a measurable space $\mathcal{X}$, the function $k : L^2(\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ is a kernel if and only if there is some feature map into a separable Hilbert space; $\Psi : \mathcal{X} \to \mathcal{H}$ such that $k(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$. This equivalently agrees that k is a kernel if an only if the following diagram commutes



Theorem 4.2.3. *[38] Let $K(x, y)$ be reproducing kernel for the RKHS $\mathcal{H}$ on $\mathcal{X}$. If $\{e_s : s \in I\}$ is an orthonormal basis for $\mathcal{H}$, then $K(x, y) = \sum_{s \in S} \overline{e_s(y)} e_s(x)$, where the series converges pointwise.*

Proof. See [38]. □

Consequentially, given a finite dimensional inner product space $(V, \langle \cdot, \cdot \rangle_V)$, the space of linear maps $V \to V$ $(\mathcal{B}(V))$, is isomorphic to the bilinear functionals $V \times V \to \mathbb{R}$. The subcollection of self adjoint maps of $\mathcal{B}(V)$ is isomorphic to the symmetric linear functionals. This means an inner product on V should be of the form $(u, w) \mapsto \langle u, Lw \rangle_V$ for some self adjoint operator $L : V \to V$. Denote by $\langle \cdot, \cdot \rangle_L$ the map

$$\langle u, w \rangle_L \to \langle u, Tw \rangle_V \tag{4.8}$$

This produces a characterization of positive linear operators as stated in proposition (4.2.4) below

Proposition 4.2.4. *The self adjoint linear map $L : V \to V$ is positive semi-definite if and only if $\langle \cdot, \cdot \rangle_L$ as defined in* (4.8) *is an inner product.*

Proof. We recall that if $L : \mathcal{H} \to \mathcal{H}$ is a bounded, self-adjoint linear map, then L is positive semi-definite if and only if $\langle L\varphi, \varphi \rangle_{\mathcal{H}} \geq 0$. Let L be a self adjoint linear map, then L has the decomposition $L = UDU^*$ for some unitary matrix U and a diagonal matrix D, inferred from spectral theory. If L is positive definite, then D admits a unique root, so that

$$\langle v, v \rangle_L = \langle v, UDU^* v \rangle_V = \left\langle \sqrt{D}U^* v, \sqrt{D}U^* v \right\rangle_V \geq 0. \tag{4.9}$$

$\sqrt{D}U^*$ is full rank and thus $\langle v, v \rangle_L = 0$ if and only if $v = 0$ So $\langle \cdot, \cdot \rangle_L$ is an inner product

On the other hand, suppose g is an eigenvector of L with eigenvalue λ, and if $\langle \cdot, \cdot \rangle_L$ is an inner product,

then

$$\lambda \langle g, g \rangle_V = \langle g, \lambda g \rangle_V = \langle g, Lg \rangle_V = \langle g, g \rangle_L > 0 \tag{4.10}$$

Moreover, $\lambda > 0$ since $\langle \cdot, \cdot \rangle_L$ is an inner product. □

There is a natural way of determining those functions on a product space $\mathcal{X} \times \mathcal{X}$ that meet the criteria of a kernel. Mercer's theorem states that a symmetric, and positive-definite matrix can be represented as a sum of a convergent sequence of product functions thereby generalizing the result that any such matrix is a Gram matrix. That is, heuristically, the theorem imposes a condition on functions to determine those that meet the criteria of a kernel function.

Theorem 4.2.5. *(Mercer's Theorem [61]) Let $(\mathcal{X}, \mu)$ be a finite-measure space, and suppose $k \in L^\infty\left(\mathcal{X}^2\right)$ is a symmetric real-valued function such that the integral operator*

$$\begin{aligned} L_k : L^2(\mathcal{X}) &\to L^2(\mathcal{X}) \\ f \mapsto (L_k f)(x) &= \int_{\mathcal{X}} k\left(x, x'\right) f\left(x'\right) d\mu\left(x'\right) \end{aligned}$$

is positive definite; that is, for all $f \in L^2(\mathcal{X})$, we have

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') f(x) f(x') d\mu(x) d\mu(x') \geq 0.$$

Let $\Psi_j \in L^2(\mathcal{X})$ be the normalized orthogonal eigenfunctions of L_k associated with the eigenvalues $\lambda_j > 0$, sorted in non-increasing order.

Then

i $\left(\lambda_j\right)_j \in \ell_1$,

ii $k(x, x') = \sum_{j=1}^{N_{\mathcal{X}}} \lambda_j \Psi_j(x) \Psi_j\left(x'\right)$ *holds for almost all (x, x'). Either $N_{\mathcal{X}} \in \mathbb{N}$, or $N_{\mathcal{X}} = \infty$; in the latter case, the series converges absolutely and uniformly for almost all (x, x')*

It is implied from the Mercer's theorem that, if k is Mercer kernel, it is possible to construct a mapping ϕ into a space where k is the reproducing kernel.

4.3 Occupation Kernel and the Kernelized Transform

For the feasibility of some of the theorems to come, we assume the trajectory $\gamma(t)$ lies in the interior of $\mathbb{D}$ such that we can find a concentric disk $\mathbb{D}_R$ of radius R inside the unit disk so that $\gamma(t) \in \mathbb{D}_R \subset \mathbb{D}$ with $1 - \eta < R < 1$ for some $\eta > 0$.

Proposition 4.3.1. *Let $\mathcal{H}$ be an RKHS over $\mathcal{X}$(presently $\mathbb{D}$). Let $\gamma : [0, T] \to \mathcal{X}$ be a continuous trajectory of finite length. The functional defined on $\mathcal{H}$ by $g \longmapsto \int_0^T g(\gamma(t))dt$ is bounded over $\mathcal{H}$, and may be represented as $\int_0^T g(\gamma(t))dt = \langle g, \Gamma_\gamma \rangle_{\mathcal{H}}$ for some $\Gamma_\gamma \in \mathcal{H}$ by the Riesz representation theorem. The function $\Gamma_\gamma \in \mathcal{H}$ is called occupation kernel corresponding to γ.*

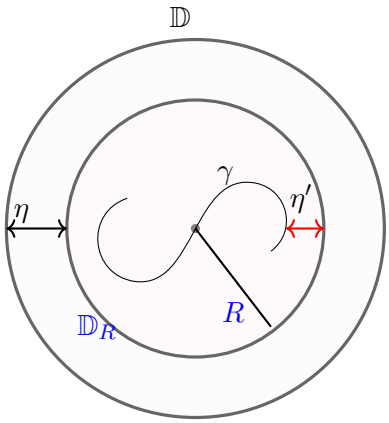


Figure 14. Layout of a trajectory inside the unit disk

Proof. We demonstrate boundedness of the functional for the case $\mathcal{H} = \mathcal{H}^2$, the property can be generalized. We suppose that γ lies in the interior of $\mathbb{D}$ such that we can construct a concentric disk of radius $R > 0$, with $\gamma \in \mathbb{D}_R \subset \mathbb{D}$.

Let $\Xi : \mathcal{H}^2 \to \mathbb{C}$ be the map $\mathcal{H}^2 \ni g \mapsto \int_0^T g(\gamma(t))dt$. Fix $1 - \eta < R < 1$, and let $\eta' := dist(\partial\mathbb{D}_R, \gamma)$; note that $0 < \eta' < \eta$. Then, by the Cauchy integral formula, since $\gamma(t) \in \mathbb{D}_R$, for any $0 < t < T$, we

have that

$$\begin{aligned}|\Xi(g)| &= \left|\int_0^T g(\gamma(t))dt\right| \\ &= \left|\int_0^T \frac{1}{2\pi i}\int_{|\zeta|=R}\frac{g(\zeta)}{\zeta-\gamma(t)}d\zeta dt\right| \\ &\leq \frac{1}{2\pi}\int_{|\zeta|=R}|g(\zeta)|\int_0^T \frac{dt}{|\zeta-\gamma(t)|}|d\zeta|.\end{aligned}$$

Since $\eta' = \text{dist}(\partial\mathbb{D}_R, \gamma) > 0$, we have that $|\zeta - \gamma(t)| > \eta'$. Therefore

$$\begin{aligned}|\Xi(g)| &\leq \frac{T}{2\pi\eta'}\int_{|\zeta|=R}|g(\zeta)||d\zeta| \\ (Cauchy-Schwarz) &\leq \frac{\sqrt{2\pi R}T}{2\pi\eta'}\left(\int_{|\zeta|=R}|g(\zeta)|^2|d\zeta|\right)^{1/2}.\end{aligned}$$

Taking the limit as $R \to 1$, we find that

$$|\Xi(g)| \leq C||g||_{\mathcal{H}^2(\mathbb{D})},$$

where the constant $C = C(\eta, T)$. Thus, Ξ is a bounded linear functional (hence continuous). □

The function Γ_γ, as we saw in the previous chapter, may be expressed in terms of the reproducing kernel of $\mathcal{H}$. This is a direct consequence of the reproducing property of the kernel function k. A formal statement of is given in proposition 4.3.2 below.

Proposition 4.3.2. *[48] Let $\mathcal{H}$ be a RKHS consisting of continuous functions over a compact set $\mathcal{X}$ and let $\gamma : [0, T] \to \mathcal{X}$ be a continuous trajectory as defined above. The occupation kernel corresponding to γ, Γ_γ in $\mathcal{H}$, may be expressed as*

$$\Gamma_\gamma(x) = \int_0^T k(x, \gamma(t))dt$$

Proof. Notice that $\Gamma_\gamma(x) = \langle \Gamma_\gamma, k(\cdot, x)\rangle_{\mathcal{H}}$ by the reproducing property of k. Consequently,

$$\begin{aligned}\Gamma_\gamma(x) &= \langle \Gamma_\gamma, k(\cdot, x)\rangle_{\mathcal{H}} = \overline{\langle k(\cdot, x), \Gamma_\gamma\rangle}_{\mathcal{H}} \\ &= \int_0^T \overline{k(\gamma(t), x)}dt = \int_0^T k(x, \gamma(t))dt, \quad \text{by definition of } \Gamma_\gamma.\end{aligned}$$

which establishes the result

$$\Gamma_\gamma(x) = \int_0^T k(x, \gamma(t))dt. \tag{4.11}$$

□

Proposition 4.3.3. *Let $\mathcal{H}$ be a RKHS of continuous functions over a compact set $\mathcal{X} \subset \mathbb{C}^n$, and suppose $\varphi : \mathbb{C} \to \mathbb{C}$ is Lipschitz continuous. If $\gamma : [0, T] \to \mathcal{X}$ is a trajectory as defined in proposition* (4.3.1)*, and $M_\varphi : \mathcal{H} \to \mathcal{H}$, then $\Gamma_\gamma \in D(M_\varphi^*)$ and*

$$\left[M_\varphi^*\right] \Gamma_\gamma = \int_0^T \overline{\varphi(\gamma(t))} k(\cdot, \gamma(t))dt. \tag{4.12}$$

Proof. Let $\mathcal{H}$ be an RKHS, and Γ_γ the occupation kernel associated with the continuous trajectory $\gamma : [0, T] \to \mathcal{X}$.

To show that $\Gamma_\gamma \in D\left(M_\varphi^*\right)$, it is enough to demonstrate that the linear functional

$$g \mapsto \langle M_\varphi g, \Gamma_\gamma \rangle_{\mathcal{H}} \text{ is bounded for all } g \in D(M_\varphi)$$

Consider the map E defined by

$$E : \mathcal{H} \longrightarrow X$$
$$g \mapsto \langle M_\varphi g, \Gamma_\gamma \rangle_{\mathcal{H}}.$$

We get that $|E(g)| = \left|\langle M_\varphi g, \Gamma_\gamma \rangle_{\mathcal{H}}\right|$, where

$$\left|\langle M_\varphi g, \Gamma_\gamma \rangle_{\mathcal{H}}\right| = \left|\langle \varphi g, \Gamma_\gamma \rangle_{\mathcal{H}}\right| = \left|\int_0^T (\varphi g)(\gamma(t))dt\right| = \left|\int_0^T \varphi(\gamma(t)) g(\gamma(t))dt\right| \leq \int_0^T |\varphi(\gamma(t))||g(\gamma(t))|\, dt.$$

But $\varphi(\gamma(t))$ is continuous on $[0, T]$ and that the image of $\varphi(\gamma(t))$ is compact and bounded, so

$$\int_0^T |\varphi(\gamma(t))||g(\gamma(t))|\, dt \leq \sup_{\gamma(t)} |\varphi(\gamma(t))| \int_0^T |g(\gamma(t))|\, dt \tag{4.13}$$

Since $g \in \mathcal{H}$ and γ is continuous, the function $g(\gamma(t))$ is integrable and the quantity on the right hand side of equation (4.13) is bounded. It follows that $g \mapsto \langle M_\varphi g, \Gamma_\gamma \rangle$ is bounded as desired.

To obtain the closed-form of the interaction $\left[M_\varphi^*\right]\Gamma_\gamma$, we follow from the definition of Γ_γ.

$$\begin{aligned}\left[M_\varphi^*\right]\Gamma_\gamma(z) &= \left\langle M_\varphi^*\Gamma_\gamma, k_z\right\rangle = \overline{\langle M_\varphi k_z, \Gamma_\gamma\rangle} = \int_0^T \overline{(M_\varphi k_z)(\gamma(t))}dt \\ &= \int_0^T \overline{\varphi(\gamma(t))k_z(\gamma(t))}dt = \int_0^T \overline{\varphi(\gamma(t))}k_{\gamma(t)}(z)dt\end{aligned} \tag{4.14}$$

as desired. □

Perhaps, it is worth mentioning that under the setting of Proposition 4.3.3, the action of the multiplication operator on the occupation kernel $M_\varphi\Gamma_\gamma$ has the form

$$M_\varphi\Gamma_\gamma(z) = \varphi(z)\Gamma_\gamma(z) \tag{4.15}$$

By the reproducing property of k, also noting that k is an eigenfunction of the adjoint operator M_φ^*, we get that

$$\begin{aligned}M_\varphi\Gamma_\gamma(z) &= \langle M_\varphi\Gamma_\gamma, k_z\rangle = \langle \Gamma_\gamma, M_\varphi k_z\rangle = \overline{\left\langle M_\varphi^* k_z, \Gamma_\gamma\right\rangle} \\ &= \overline{\left\langle \overline{\varphi(z)}k_z, \Gamma_\gamma\right\rangle} = \overline{\int_0^T \overline{\varphi(z)}k(\gamma(t), z)dt} = \int_0^T \varphi(z)k(z,\gamma(t))dt = \varphi(z)\int_0^T k(z,\gamma(t))dt = \varphi(z)\Gamma_\gamma(z).\end{aligned}$$

4.4 Function - Theoretic Properties of $\left[M_\varphi^*\right]\Gamma_\gamma$

In a typical *inverse problem (in the broad sense)*, one wishes to recover some unknown signal, starting with only partial knowledge about the signal. For instance, in *tomography*, the theoretical question remains; is there a way to recover an unknown function f, if one knows values of its integral along all possible cross-sections in some target domain [12]. The *Radon transform* for example, "measures" integral of f along various cross-sections. In practice, only a discrete set of cross-sections are feasible. The approach therefore is an approximation routine that involves filtering out noise and undesirable components captured by the Radon transform to obtain the target function. Similar to this method, we ask the question: suppose we have enough data on the quantity $\int \overline{\varphi(\gamma(t))}k(z,\gamma(t))dt$, is it possible to recover the symbol φ over its domain of definition? Additionally, one would want to know the properties

of such a function. In what follows, we attempt to study function and operator-theoretic properties of $M_\varphi^*\Gamma_\gamma$ that may offer useful insight in the recovery of the often unknown symbol φ. We begin by investigating some properties of the function over a Hilbert space.

By far, we have not precisely characterized the composition of the set S comprising of symbols of densely defined multiplication operators on $\mathcal{H}^2$. We recall from Chapter 2 that the multiplier algebra of $\mathcal{H}^2(\mathbb{D})$ is $\mathcal{H}^\infty(\mathbb{D})$. Sarason outlines precisely the class of multipliers of the the Hardy space with a dense multiplicative domain in $\mathcal{H}^2$ [57]. Let φ be a holomorphic function in $\mathbb{D}$ and let $M_\varphi : \mathcal{H}^2 \to \mathcal{H}^2$. Clearly, M_φ is a closed operator. If $D(M_\varphi) \neq \{0\}$, then φ is a member of the Nevanlinna class (N), where the Nevanlinna class consists of all functions $\{f : \mathbb{D} \to \mathbb{C},\ \text{holomorphic}\ : f = a/b; a, b \in \mathcal{H}^\infty\}$.

Lemma 4.4.1. *[57] If $D(M_\varphi)$ is dense in $\mathcal{H}^2$, then φ is in the Smirnov class N^+.*

Proof. See [57] □

The Smirnov class N^+ on the other hand consists of holomorphic functions in $\mathbb{D}$ that are a quotient of two functions in $\mathcal{H}^\infty$ such that the denominator is an *outer function*. i.e.,

$$N^+ = \{f \in \mathcal{H}^\infty : f = \frac{b}{a}; a, b \in \mathcal{H}^\infty, \text{a is outer}\}.$$

An outer function is a function $f \in \mathcal{H}^1$ which can be written in the form

$$f\left(re^{i\theta}\right) = \alpha \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{it} + re^{i\theta}}{e^{it} - re^{i\theta}} k\left(e^{it}\right) dt\right), \tag{4.16}$$

for an integrable function k, $e^{i\theta} \in \mathbb{D}$, and $|\alpha| = 1$.

In a converse of Lemma 4.4.1, Sarason in [57] further characterized the domain, $D(M_\varphi)$, corresponding to all such symbols $\varphi \in N^+$.

Proposition 4.4.2. *[57] Let φ be an nonzero function in N^+ with representation $\varphi = a/b$. Then $D(M_\varphi) = a\mathcal{H}^2$.*

Therefore, the class of symbols of densely defined multiplication operators on the Hardy space is precisely the Smirnov class N^+. However, for the rest of this chapter, we shall focus on the subset $S = \mathcal{H}^\infty \subset N^+$ unless otherwise specified.

There have been numerous discoveries in recent years on the classification of symbols of the the multiplication and Toeplitz operators. For an in depth review of the subject, we refer the reader to the references [56, 57, 63].

Fix $\varphi \in S$ and a system trajectory $\gamma : [a, b] \to \mathbb{D}$. The adjoint operator $M_\varphi^* : \mathcal{H}^2 \to \mathcal{H}^2$ acts on $\mathcal{H}^2$ as multiplication by the conjugate function $\overline{\varphi}$ preceeded by the analytic Szego projection onto $\mathcal{H}^2$:

$$(M_\varphi^*)[g](z) = \mathbb{P}_{\mathcal{H}^2(\mathbb{D})}\left[\overline{\varphi(\zeta)}g(\zeta)\right](z). \tag{4.17}$$

Here, $\mathbb{P}_{\mathcal{H}^2(\mathbb{D})} : L^2(\mathbb{T}) \to \mathcal{H}^2(\mathbb{D})|_{\mathbb{T}}$ is the canonical projection onto the Hardy space defined via integration against the reproducing kernel

$$\mathbb{P}_{\mathcal{H}^2(\mathbb{D})}[h](z) := \frac{1}{2\pi}\int_0^{2\pi} h(\zeta)k_z(\zeta)d\zeta = \langle h, k_z\rangle, \qquad h \in L^2(\mathbb{T}). \tag{4.18}$$

In particular, the action of M_φ^* on the occupation kernel belongs to $\mathcal{H}^2$, $\left[M_f^*\right]\Gamma_\gamma(z) \in \mathcal{H}^2(\mathbb{D})$.

We shall represent the function $\left[M_\varphi^*\right]\Gamma_\gamma$ in (4.33) by $T_{\varphi,\gamma}$, that is

$$\begin{aligned} T_{\varphi,\gamma} &: \mathbb{D} \to \mathbb{C} \\ &x \mapsto \int_0^T \overline{\varphi(\gamma(t))}k(x, \gamma(t))dt \end{aligned} \tag{4.19}$$

and for such choices of φ and γ, $T_{\varphi,\gamma}$ is well defined. The conjugation of the integrand is dropped when we shift the focus to real-valued symbols. Given $\varphi \in \mathcal{H}^\infty$ as indicated, and the reproducing kernel k of the Hardy space, the function $\varphi(z)k(z, w)$ is analytic in z being the product of two analytic functions. Assuming the function $T_{\varphi,\gamma}$ is analytic, then using an idea from the complex analog of fundamental theorem of calculus, we may construct a primitive F with $F' = T_{\varphi,\gamma}$ as below. Taking Ω to be a smooth curve properly contained in the interior of the disk and parameterized by $\xi(t), \quad 0 \le t \le T$, if $\gamma(t) \in \Omega$, then we may write

$$F(\xi(b)) - F(\xi(a)) = \int_\Omega \left[\int_0^T \varphi(\gamma(t))k(z, \gamma(t))dt\right] dz \tag{4.20}$$

where $F' = \int_0^T \varphi(\gamma(t))k(z,\gamma(t))dt$. By Cauchy's integral formula, we know that analytic functions have derivatives of all orders, and so F'' exists, thus $F' = T_{\varphi,\gamma}$ is analytic. However, analyticity of $T_{\varphi,\gamma}$ is not obvious for a general RKHS.

Take for example $k(z,w) = e^{itz}$ and $\varphi \in L^2(0,T)$, then the function

$$M_\varphi^* \Gamma_\gamma(z) = \int_0^T \varphi(w) e^{iwz} dw$$

is entire and the exponential type. It satisfies the growth conditions [53]

$$|M_\varphi^* \Gamma_\gamma(z)| \leq \int_0^T |\varphi(t)| e^{-ty} dt \leq e^{T|y|} \int_0^T |\varphi(t)| dt. \tag{4.21}$$

Suppose $\int_0^T |\varphi(t)| dt = C < \infty$. Then equation (4.21) implies

$$|M_\varphi^* \Gamma_\gamma(z)| \leq C e^{T|z|}. \tag{4.22}$$

In the case of the Hardy space, a similar argument can be made. We observe that the Szego kernel which is skew symmetric is most importantly holomorphic in the first variable and antiholomorphic in the second. As such, we can write

$$|T_{\varphi,\gamma}(z)| = \left| \int_0^T \overline{\varphi(\gamma(t))} k(z,\gamma(t)) dt \right| \leq \|\varphi(\gamma)\|_\infty \int_0^T |k(z,\gamma(t))dt| = \|\varphi(\gamma)\|_\infty \int_0^T \sum_{n=0}^\infty \left[z\overline{\gamma(t)} \right]^n dt \tag{4.23}$$

where the series on the right is convergent per the hypothesis. Therefore, we may compute the $\mathcal{H}^2$ norm accordingly

$$\int_\Omega |T_{\varphi,\gamma}(z)|^2 dz \leq \|\varphi(\gamma)\|_\infty \int_\Omega \left| \int_0^T \sum_{n=0}^\infty \overline{\gamma(t)}^n z^n dt \right|^2 dz. \tag{4.24}$$

For an arbitrary RKHS $\mathcal{H}$ over the set $\mathcal{X}$, if the reproducing kernel k is uniformly bounded, a similar argument can be made. For $\varphi \in \mathcal{H}^\infty$, the function $x \mapsto T_{\varphi,\gamma}$ is bounded with respect to the norm on $\mathcal{H}$. The conditions on the symbol φ and kernel could be relaxed, the property still holds for continuous symbols. However, since the ultimate goal involves symbols S of densely defined multiplication, we will mostly cover elements of $\mathcal{H}^\infty$ space.

As mentioned, the map $T_{\varphi,\gamma}$ may be viewed as either a function on the set $\mathbb{D}$ by holding quantities φ and γ constant, or as a map on the collection S of symbols of densely-defined multiplication operators M_φ.

$$T_{\varphi,\gamma} : \mathcal{X} \to \mathbb{C} \tag{4.25a}$$

$$z \mapsto M_\varphi^* \Gamma_\gamma(z) = T_{\varphi,\gamma}(z) = \int_0^T k(z,\gamma(t))\overline{\varphi(\gamma(t))}dt, \quad \text{for fixed} \quad \varphi,\gamma \tag{4.25b}$$

$$\Lambda : S \to \mathcal{H}^2 \tag{4.25c}$$

$$[\Lambda](\varphi) \mapsto T_{\varphi,\gamma}(z) = \int_0^T k(z,\gamma(t))\overline{\varphi(\gamma(t)}dt, \quad \forall z \in \mathcal{X}. \tag{4.25d}$$

Depending on how one chooses to view this map, different properties about the action $\left[M_\varphi^*\right]\Gamma_\gamma$ may be learned. It is important to note that the collection S of symbols under consideration coupled with the scalar product $\langle \cdot, \cdot \rangle_S$ defines an inner product space. Additionally, the map $\Lambda : S \mapsto T_{\varphi,\gamma}$ is a bounded linear map that represents a one-to-one correspondence between S and the kernelized transform. Given $\psi, \varphi \in S$ such that $T_{\varphi,\gamma}(x) = T_{\psi,\gamma}$ for all $x \in \mathcal{X}$, we get that $\psi = \varphi$ almost everywhere on $\gamma([0,T])$ in the L^2 sense. We posit that for any open contour γ, $ker(\Lambda) = \{\psi \in S : \Lambda\varphi = 0, \forall x \in \mathcal{X}, \quad \& \quad \forall\gamma([0,T])\} = \{0\}$, and is injective

Proposition 4.4.3. *Suppose $\gamma(t)$ is an open contour, that is $\mathbb{D}\backslash\gamma$ is connected, then the operator*

$$\begin{aligned} &\Lambda : \mathcal{H}^\infty \to \mathcal{H}^\infty \subset \mathcal{H}^2 \\ &g \mapsto \int_0^T \overline{g(\gamma(t))}k(z,\gamma(t))dt \end{aligned} \tag{4.26}$$

is injective.

Proof. Suppose not, so that $(\Lambda g)(z) = 0$ for all $z \in \mathbb{D}$ yet $g \neq 0$ i.e.

$$\int_\gamma \overline{g(\zeta)}k(z,\zeta)\frac{dt}{\gamma'} = \int_\gamma \frac{\overline{g(\zeta)}}{1-z\bar{\zeta}}\frac{d\zeta}{\gamma'} = 0. \tag{4.27}$$

Recall from Cauchy-type integrals that if

$$\Phi = \int_{\Gamma} \frac{\varphi(\zeta)}{\zeta - z} d\zeta, \tag{4.28}$$

then Φ is holomorphic in $\mathbb{C}\backslash\Gamma$. With a slight manipulation, the right hand side of equation (4.27) maybe written as a Cauchy-type integral where the integration happen over the conjugate of the path γ. i.e.

$$(\Lambda g)(z) = \int_{1/\overline{\gamma}} \frac{\Psi(\zeta)}{\zeta - z} d\zeta, \tag{4.29}$$

where Ψ depend on $\overline{g(\gamma(t))}/\gamma'$. $(\Lambda g)(z)$ is analytic up to the boundary of $1/\overline{\gamma}$. Implying $\Lambda g = 0$ in $\mathbb{C}\backslash\frac{1}{\overline{\gamma}}$ by the identity principle. Therefore by Sokhotski–Plemelj formula, $2\pi i\Psi = (\Lambda g)_+ - (\Lambda g)_-$, where $(\Lambda g)_\pm = \lim_{z \to 1/\overline{\gamma}^\pm} (\Lambda g)(z)$. But $\Lambda g = 0$ on $\mathbb{C}\backslash 1/\overline{\gamma}'$. Hence $\Psi(\zeta) = 0$ for all $t \in [0, T]$. Since $\gamma' \neq 0$, and g is holomorphic, $g \equiv 0$ by the Identity Theorem. □

For the first part of our investigation, we focus on function-theoretic properties of $T_{\varphi,\gamma}$ as a function on $\mathbb{D}$. There are resemblances between the transform in (4.25b) and Fourier and Fredholm integral operators. For instance, we observe that, if the path $\gamma(t)$ is a smooth rectifiable curve, appropriate parameterization restructures (4.25d) as a normalized Hilbert-Schmidt integral operator provided the kernel function $k(x, y) \in L^2$. Hilbert -Schmidt operators has a natural Hilbert space structure and it contains the trace class operators [16] as well as compact, hence, such a connection provides useful insight about our transform. We shall investigate properties like continuity, and analyticity of Λ . We review a few underlying notions about the kernel function and integral operators which shall serve as basis for the arguments to come. Discussions involving the Fredholm operator is deferred to Chapter 5.

We recall that if L is a linear map between the normed linear spaces $\mathcal{X}$ and $\mathcal{Y}$, we say L is bounded if maps bounded subset of $\mathcal{X}$ into bounded sets of $\mathcal{Y}$, or equivalently, if there exists $C < \infty$ such that $\|Lx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$. If certain growth constraints are imposed on the kernel function k, the function $T_{\varphi,\gamma}$ is bounded in $\mathbb{D}$. Suppose $\mathcal{X} = \mathbb{D}$ and $k : \mathbb{D} \times \mathbb{D} \to \mathbb{C}$ satisfy

$$\sup_{z\in\mathbb{D}} \int_{\mathcal{X}} |k(z, w)|dw < \infty \text{ and } \sup_{w\in\mathbb{D}} \int_{\mathcal{X}} |k(z, w)|dz < \infty.$$

Then the function $T_{\varphi,\gamma}$ defined by $T_{\varphi,\gamma}(z) = \int_0^T \overline{\varphi(\gamma(t))} k(z,\gamma(t))dt$ is bounded since

$$|T_{\varphi,\gamma}(z)| = \left| \int_0^T \overline{\varphi(\gamma(t))} k(z,\gamma(t))dt \right| \leq \int_0^T |\varphi(\gamma t) k(z,\gamma(t))|\, dt \leq \|\varphi\|_{\mathcal{H}^\infty} \int_0^T |k(z,\gamma(t))|\, dt. \tag{4.30}$$

A continuous kernel function is termed *Hilbert-Schmidt* if it is bounded in both variables. Specifically, given the kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we say k is Hilbert-Schmidt if

$$\int_{\mathcal{X}} \int_{\mathcal{X}} |k(x,y)|^2 dxdy = C < \infty. \tag{4.31}$$

For the nature of integral in $T_{\varphi,\gamma}$, the Hillbert-Schmidt property is perhaps a strong condition that guarantees continuity for a wide class of symbols. For instance it guarantee boundedness of the function in equation (4.19) for merely continuous symbols.

4.5 The Operator $\Lambda : S \to T_{\varphi,\gamma}$

Now consider the operator Λ defined as

$$\begin{aligned} \Lambda :& S \to \mathcal{H}^2 \\ & [\Lambda] : \varphi \mapsto T_{\varphi,\gamma}(z) = \int_0^T k(z,\gamma(t)) \overline{\varphi(\gamma(t)} dt, \quad \forall z \in \mathcal{X}. \end{aligned} \tag{4.32}$$

Given the set S comprised of $\mathcal{H}^\infty$ functions, the transform Λ is bounded if $\Lambda(\varphi) \leq M\|\varphi\|$ for all $\varphi \in S$. It is clear that, for any "nice" symbol, convergence of the integral to the right of equation (4.32) depends heavily on the kernel function. For the obvious cases, if k is continuous and an L^∞ function, then the integral operator inherits boundedness directly as discussed in the case of the function above. On the other hand, if $k(\cdot,\cdot)$ is continuous, symmetric and bounded in the first variable, Λ is bounded as we will show shortly. While these conditions appear desirable, the most frequently used kernels for system identification routines may not possess them. For instance, the Szego kernel, though analytic in the first variable, is antiholomorphic in the second and antisymmetric; unless the region is carefully defined the integral in (4.32) may not be defined.

For the Hardy space, there are possible limitations on the operator Λ should the trajectory γ be made arbitrarily close to the boundary $\partial\mathbb{D}$. The presence of poles of the $\mathcal{H}^2$ kernel along the boundary induces rapid growth in the integrand rendering the integral in (4.32) undefined. As such, we assume trajectories reside in the interior of the disk reasonably spaced from the boundary as indicated in the diagram from proposition (4.3.1).

Proposition 4.5.1. *Let $\mathcal{X}$ be the unit disk in the complex plane. Given $\gamma(t) : [0, T] \to \mathcal{X}$, a smooth trajectory, such that $\gamma(t) \subset int(\mathbb{D})$, the operator Λ defined above in (4.32) maps $\mathcal{H}^\infty \to \mathcal{H}^2$ and is bounded.*

Proof. For any $z \in \mathbb{D}$, Λ has a bound. That is, by setting $\zeta = \gamma(t)$, we get that the operator

$$[\Lambda](\varphi) = \int_0^T \overline{\varphi(\gamma(t))} k(\cdot, \gamma(t)) dt \tag{4.33}$$

is a normalized Hilbert-Schmidt integral transform where the measure is normalized by the factor $1/\gamma'$. That is

$$[\Lambda]\varphi(z) = \int_0^T \overline{\varphi(\gamma(t))} k(z, \gamma(t)) dt = \int_\gamma \left(\frac{\overline{\varphi(\zeta)}}{\gamma'} \right) k(z, \zeta) d\zeta, \tag{4.34}$$

Suppose that γ' is bounded away from 0; $|\gamma'| > m > 0$, the function $\overline{\varphi(\zeta)}/\gamma'$ is continuous, and thus for the Szego kernel,

$$\int_\gamma \left(\frac{\overline{\varphi(\zeta)}}{\gamma'} \right) k(z, \zeta) d\zeta = \int_\gamma \frac{\Psi(\zeta)}{z - \zeta} d\zeta, \tag{4.35}$$

where $\Psi(\zeta) = \Psi(\zeta, \frac{\overline{\varphi(\zeta)}}{\gamma'})$. The integral is well-defined where $z \neq 1/\zeta$ and is analytic in the whole of $\mathbb{D}$ except possible along the contour γ. Now suppose we define $B(z) = \int_\gamma \frac{\Psi(\zeta)}{1 - z\bar{\zeta}} d\zeta$, for $z_0 \in \gamma$, let $B_+(z_0) = \lim_{\zeta \to z_0^+} B(z)$ be the limiting value as Ψ tends to the contour from inside and $B_-(z_0) = \lim_{\zeta \to z_o^-} B(z)$. Then by the *Sokhotski–Plemelj theorem* [27], $B_+(z_0) - B_-(z_0) = 2\pi i \Psi(z_0)$. However, $B_+(z_0)$ and $B_-(z_0)$ are well-defined, holomorphic and their limits agree on the boundary γ. Thus the function $B(z)$ extends analytically to the boundary γ. Boundedness follows immediately since γ lies completely in some compact set $\overline{\mathbb{D}_R} \subset \mathbb{D}$. □

For the Hardy space on the unit disk, we may estimate this bound as below.

Without loss of generality, let the set $[a,b]=[0,T]$, therefore

$$|\Lambda(\varphi)| = \left| \int_a^b \overline{\varphi(\gamma(t))} k(z,\gamma(t)) dt \right| \leq \int_a^b \left| \overline{\varphi(\gamma(t))} k(z,\gamma(t)) \right| dt \tag{4.36a}$$

$$\leq \|\varphi\|_{\mathcal{H}^\infty} \int_a^b |k(z,\gamma(t))|\, dt = \|\varphi\|_{\mathcal{H}^\infty} \int_a^b \left| \frac{1}{1-z\overline{\gamma(t)}} \right|, \tag{4.36b}$$

where $k(z,\gamma(t)) = \dfrac{1}{1-z\overline{\gamma(t)}} = \sum_{n=0}^{\infty} \left(z\overline{\gamma(t)}\right)^n$, $z, \gamma(t)$ is in the interior of $\mathbb{D}$ which guarantees convergence in the series on the right. Therefore

$$|\Lambda(\varphi)| \leq \|\varphi\|_{\mathcal{H}^\infty} \int_a^b \left| \sum_{n=0}^{\infty} (z\gamma(t))^n \right| dt \leq \|\varphi\|_{\mathcal{H}^\infty} \int_a^b \sum_{n=0}^{\infty} |(z\gamma(t))^n|\, dt = \frac{\rho \|\varphi\|_{\mathcal{H}^\infty}}{m(1-|z|)},$$

where ρ is a constant depending on the measure of $[a,b]$ and $m \leq \gamma'([a,b]) \leq M$ for some $m, M \in \mathbb{R}$. In fact, this results can be refined further. Specifically, given a continuous symbol $\varphi \in \mathcal{C}(\mathbb{D})$ and a smooth rectifiable curve γ of finite length, the operator $\Lambda : \mathcal{C}(\mathbb{D}) \to \mathcal{H}^2$ is bounded. The image of the compact set $[a,b]$ is compact under γ, and since φ is continuous, the results follows.

Recall that, every Hilbert-Schmidt operator is compact [18]. That is, given that K is Hilbert-Schmidt operator on $\mathcal{H}$, $\{e_i | i \in \mathbb{N}\}$ an orthonormal basis collection for $\mathcal{H}$, and P_N a projection of K onto the finite dimensional space spanned by $\{e_1, e_2, \cdots, e_N\}$. Then $P_N K$ is a finite rank operator with $P_N K \to K$ uniformly as $N \to \infty$. The result therefore follows immediately. A formal statement regarding this is given by the following theorem.

Theorem 4.5.2. *[45] Suppose $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a Hilbert-Schmidt kernel, then the integral operator $K \in \mathcal{B}(L^2(\mathcal{X}))$, defined by*

$$K(f)(x) = \int_{\mathcal{X}} k(x,y) f(y) dy$$

is compact,

where $\mathcal{B}(L^2(\mathcal{X}))$ is the space of all bounded linear operators on $L^2(\mathcal{X})$.

Proof. Let $\{\phi_i(x)\}$ be an orthonormal basis for $L^2(\mathcal{X})$. One can show that $\{\phi_i(x)\phi_j(y)\}$ is a basis for $L^2(\mathcal{X}\times\mathcal{X})$. Expanding k with respect to this basis gives

$$k(x,y)=\sum_{i,j=1}^{\infty}k_{ij}\phi_i(x)\phi_j(y), \tag{4.37}$$

where convergence of (4.37) is in the $L^2(\mathcal{X}\times\mathcal{X})$ norm and

$$k_{ij}=\int_{\mathcal{X}}\int_{\mathcal{X}}k(x,y)\phi_i(x)\phi_j(y)dxdy. \tag{4.38}$$

Furthermore, we have that

$$\int_{\mathcal{X}}\int_{\mathcal{X}}|k(x,y)|^2dxdy=\sum_{i,j=1}^{\infty}|k_{ij}|. \tag{4.39}$$

For $n\in\mathbb{N}$, define the operator $K_n\in\mathcal{B}(L^2(\mathcal{X}))$ by

$$K_n(f)(x):=\int_{\mathcal{X}}k_n(x,y)f(y)dy, \tag{4.40}$$

where

$$k_n(x,y)=\sum_{i,j=1}^{n}k_{ij}\phi_i(x)\phi_j(y).$$

We refer to k_n and K_n as separable kernels and separable operators respectively. Notice that the separable operators have finite rank and thus are compact.

Applying lemma (4.2.2), we get that

$$\|K-K_n\|^2\leq\int_{\mathcal{X}}\int_{\mathcal{X}}|k(x,y)-k_n(x,y)|^2dxdy \tag{4.41}$$

and using the expression in equation (4.39) gives

$$\lim_{n\to\infty}\int_{\mathcal{X}}\int_{\mathcal{X}}|k(x,y)-k_n(x,y)|^2dxdy=\lim_{n\to\infty}\sum_{i,j=n+1}^{\infty}|k_{ij}|=0, \tag{4.42}$$

implying that K_n converges to K in the operator norm. Recall that, if $A_n \in \mathcal{B}(X,Y)$ is a sequence of compact operators such that A_n converges in the operator norm to an operator A, then A is compact. Hence, K is compact. □

Take $(\mathcal{X}, \Omega, \mu)$ to be a σ-finite measure space. Let φ be an element of $\mathcal{H}^\infty$ and k, a Hilbert-Schmidt kernel, then given γ, a smooth trajectory of finite length, the operator $\Lambda : \mathcal{H}^\infty \to \mathcal{H}^2$ defined in (4.32) is compact.

It suffices to show that the map $[\Lambda(\varphi)](z) = \int_0^T \overline{\varphi(\gamma(t))} k(z, \gamma(t))$ is a Hilbert- Schmidt operator. By virtue of the Hilbert-Schmidt kernel k, Λ is bounded over the collection of symbols S for a fixed trajectory as shown in Proposition 4.5.1.

Given a bounded metric space $\mathcal{X}$, recall that $\mathcal{X}$ is sequentially compact provided every sequence of points in $\mathcal{X}$ has a convergent subsequence. Similarly, the operator $\mathbf{A}$ acting on a metric space $\mathcal{X}$ is said to be compact if for any bounded sequence $\{z_n\}$, the collection $\{\mathbf{A}z_n\}_{n=1}$ has a convergent subsequence.

Definition 4.5.1. Let $\mathcal{X}, \mathcal{Y}$ be two Banach spaces. The operator $K : \mathcal{X} \to \mathcal{Y}$ is compact if for all bounded sequences in $\mathcal{X}$, there is a subsequence $\{n_k\}_k$ such that $\{Kx_{n_k}\}$ is convergent.

One way to establish compactness of Λ is to show this property. To do so, we resort to an analog of the Bolzano-Weierstrass theorem of the Euclidean space. Recall that, in the Euclidean space $\mathbb{R}^n$, the Bolzano-Weierstrass theorem states that every bounded sequence $\{x_n\}_n$ in $\mathbb{R}^n$, has a convergent subsequence $\{x_{nk}\}_{nk} \subset \mathbb{R}^n$. Here, we employ the notion of a normal family.

Definition 4.5.2. [60] The collection $\mathcal{F} = \{f_\alpha\}_{\alpha \in I}$ of holomorphic functions on a domain Ω is called a *normal family* if every sequence in $\mathcal{F}$ has a uniformly convergent subsequence.

Therefore, it is enough to show that $\{M^*_{\varphi_n} \Gamma_\gamma\}$ is a normal family for any sequence $\{\varphi_n\} \subset \mathcal{H}^\infty$. According Montel's theorem, locally uniformly bounded families are normal.

Theorem 4.5.3. *(Montel's theorem [60, chapter 3]) Let Ω be open in $\mathbb{C}$, a family $\mathcal{F}$ of holomorphic functions on Ω is said to be normal if every sequecce in $\mathcal{F}$ has a subsequence that converges uniformly on every compact subset of Ω.*

Equivalently, $\mathcal{F}$ is normal if and only if it is locally uniformly bounded.

Theorem 4.5.4. *Given a smooth trajectory $\gamma : [0,T] \to \mathbb{D}$ of finite length, such that $dist(\partial\mathbb{D}, \gamma) = \eta > 0$, the the operator Λ defined as*

$$\begin{aligned} \Lambda : \mathcal{H}^\infty &\to \mathcal{H}^\infty \subset \mathcal{H}^2 \\ \varphi &\mapsto \int_0^T \overline{\varphi(\gamma(t))} k(z, \gamma(t)dt) \end{aligned} \tag{4.43}$$

is compact.

Proof. By Montel's theorem, it is enough to show that $\{\Lambda\varphi_n\}_{n\in\mathbb{N}}$ is locally uniformly bounded whenever $\{\varphi_n\}$ is bounded. We have that

$$\begin{aligned} |\Lambda\varphi_n(z)| &= \left| \int_0^T \varphi_n(\gamma(t)) k(z, \gamma(t)) dt \right| \\ &\leq \int_0^T |\varphi_n(\gamma(t))|\, k(z, \gamma(t))\, ||\, dt \\ &\leq \frac{MT}{\eta}, \end{aligned} \tag{4.44}$$

independently of n, where $|\varphi_n| < M$ for all n.

This is because, for $z \in \mathbb{D}$ say $z = re^{i\theta}$ and for any t, $\gamma(t) = \rho e^{i\psi}$, by the reverse triangle inequality, we have $|1 - z\gamma(t)| \geq |1 - |z||\gamma(t)|| = |1 - |r||\rho|| \geq |1 - (1-\eta)| > \eta$. Hence $|k(z, \gamma(t))| = \frac{1}{|1-z\overline{\gamma(t)}|} < \frac{1}{\eta}$ Therefore, by taking supremum over n, we get $\|\Lambda\varphi_n\| \leq M$. Therefore Λ is compact. □

Since we lack an inner product structure on $\mathcal{H}^\infty$, consider Λ as a map from $\mathcal{H}^2$ to $\mathcal{H}^2$, so that $D(\Lambda) = \{\varphi \in \mathcal{H}^2 : \Lambda(\varphi) \in \mathcal{H}^2\}$. Suppose $D(\Lambda)$ is not trivial, we can discuss Hilbert-Schmidt properties of Λ under this definition in the most restrictive sense. The closed operator Λ admits an adjoint $\Lambda^* : \mathcal{H}^2 \to \mathcal{H}^2$. Λ^* is in an integral transform with a kernel

$$k^*(x, y) = \overline{k(y, x)} \tag{4.45}$$

That is

$$\begin{aligned}\langle \Lambda\varphi, g\rangle &= \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \varphi(\gamma(t))k(k(x,\gamma(t)))dt\right] \overline{g(x)}dx \\ &= \int_{\mathcal{X}} \varphi(\gamma(t)) \overline{\left[\int_{\mathcal{X}} g(x)\overline{k(x,\gamma(t))}dx\right]} dt \\ &= \langle \varphi, \Lambda g\rangle\end{aligned} \tag{4.46}$$

Proposition 4.5.5. *Let $\Lambda : \mathcal{H}^2 \to \mathcal{H}^2$ such that $D(\Lambda)$ is non-trivial, then Λ a Hilbert-Schmidt operator.*

Proof. Recall that a bounded linear operator K on a separable Hilbert space $\mathcal{H}$ is said to be Hilbert-Schmidt if for any orthonormal basis of $\mathcal{H}$, $\{e_i \mid i \in \mathbb{N}\}$, we have

$$\sum_{i=1}^{\infty} \|Ke_i\|^2 < \infty.$$

with a corresponding Hilbert-Schmidt operator norm

$$\|K\|_{HS}^2 := \sum_{i=0}^{\infty} \langle Ke_i, Ke_i\rangle = \sum_{i=1}^{\infty} \|Ke_i\|^2$$

take the orthonormal basis for the Hardy space $e_n = z^n, n \geq 0$. then we wish to show that

$$\|\Lambda\|_{HS}^2 = \left(\sum_{n=0}^{\infty} \langle \Lambda e_n, \Lambda e_n\rangle\right) < \infty \tag{4.47}$$

For any any orthonormal basis, we deduce from direct calculation that $|\Lambda e_n(z)| \leq C\varepsilon^n$ for some constant C and $0 < \varepsilon < 1$. That is

$$\begin{aligned}|\Lambda e_n(z)| = |M_{e_n}^* \Gamma_\gamma(z)| &= \left|\int_0^T \overline{(\gamma(t))^n} k(z,\gamma(t))dt\right| \\ &\leq \int_0^T |\gamma(t)|^n \, |k(z,\gamma(t))| \, dt \leq \frac{T}{\eta}(1-\eta)\end{aligned} \tag{4.48}$$

since $|\gamma(t)|< 1-\eta$. Therefore

$$\begin{aligned}\langle \Lambda e_n, \Lambda e_n\rangle = \left\langle M^*_{e_n}\Gamma_\gamma, M^*_{e_n}\Gamma_\gamma\right\rangle &= \lim_{r\to 1}\int_0^{2\pi}\left|M^*_{e_n}\Gamma_\gamma(re^{i\theta})\right|^2\frac{d\theta}{2\pi}\\ &\leq C^2\varepsilon^{2n}\int_0^{2\pi}\frac{d\theta}{2\pi}\leq C^2\varepsilon^{2n}\end{aligned} \tag{4.49}$$

This yields

$$\|\Lambda\|^2_{HS}=\sum_{n=0}^{\infty}C^2\varepsilon^{2n}<\infty. \tag{4.50}$$

□

Recall that, every linear contraction map on a complete metric space has a unique fixed point, which is a key component in the study of stability theory of dynamical systems. We notice in Theorem 4.5.7 below that by imposing a bound on the kernel, the operator Λ is a contraction map.

Definition 4.5.3. (Contraction map [18]) Let $T:\mathcal{X}\to\mathcal{X}$ be a linear map on the metric space $\mathcal{X}$. T is a contraction map if $\|Tx-Ty\|\leq q\|x-y\|$ whenever $0\leq q<1$.

Theorem 4.5.6. *[18] If T is contraction, then there exists a unique $x^*\in\mathcal{X}$ such that $Tx^*=x^*$. Moreover, $T^nx_0=x^*$ for all $x_0\in\mathcal{X}$*

Theorem 4.5.7. *Let $\mathcal{X}$ be the unit disk, $\gamma:[0,T]\to\mathbb{D}$ be a smooth trajectory residing entirely in the interior of $\mathbb{D}$ so that $dist(\partial\mathbb{D},\gamma)=\eta>0$. Let the kernel k be such that*

$$\int_{\mathcal{X}}\int_{\mathcal{X}}|k(x,y)|^2\,dydx=C<1.$$

Then the map $\Lambda:\mathcal{H}^\infty\to\mathcal{H}^\infty\subset\mathcal{H}^2$ is a contraction map in the $\|\cdot\|_\infty$ norm if $T<\eta$.

Proof. By definition

$$(\Lambda\varphi)(z)=\int_0^T k(z,\gamma(t))\overline{\varphi(\gamma(t))}dt$$

Let $\varphi_1, \varphi_2 \in \mathcal{H}^\infty$ be two continuous symbols, we get that

$$\begin{aligned}
|\Lambda(\varphi_1 - \varphi_2)| = |\lambda(\varphi_1) - \Lambda(\varphi_2)| &= \left| \int_0^T k(z, \gamma(t))\overline{\varphi_1(\gamma(t))} - k(z, \gamma(t))\overline{\varphi_2(\gamma(t))} dt \right| \\
&\leq \int_0^T |k(z, \gamma(t))| \left[\overline{\varphi_1(\gamma(t)) - \varphi_2(\gamma(t))} \right] dt \\
&\leq \|\varphi_1 - \varphi_2\|_\infty \int_0^T |k(z, \gamma(t))| dt \\
&\leq \frac{T}{\eta} \|\varphi_1 - \varphi_2\|_\infty^2.
\end{aligned}$$

□

4.6 Paley-Wiener Type Theorems

Our discussion in this section is aimed at exploring connections, where possible, between the Paley-Wiener theorem and the induced kernelized transform $M_\varphi^* \Gamma_\gamma$. At the center of they Payley-Wiener theorem is the Fourier transform. The Paley-Wiener theorem utilizes the holomorphic Fourier transform on the space of square-integrable functions that are supported on the real line $(L^2(\mathbb{R}))$. The Hilbert space $L^2(\mathbb{R})$ remains a classic space to examine the Fourier transform as it is the setting for more interesting theories of the transform. A unique property of $L^2(\mathbb{R})$ is that it is its own dual. Let $f \in L^1(\mathbb{R})$, with norm $\|f\|_{L^1}$ given as

$$\|f\|_{L^1(\mathbb{R})} = \int_{-\infty}^{\infty} |f(x)| dx. \tag{4.51}$$

We denote by $\hat{f}(w) = \mathcal{F}(f(x))$, the Fourier transform of the integrable function $f(x)$ defined by

$$\hat{f}(w) = (\mathcal{F}f)(w) := \int_{-\infty}^{\infty} f(t) e^{-iwt} dt \quad \text{for all} \quad x \in \mathbb{R}. \tag{4.52}$$

Intuitively, for each w, $\hat{f}(w)$ captures the component of f that has the frequency $w/2\pi$ [12, 28, 53]. For any such integrable function, the integral in equation (4.52) is well defined for every real t and $\hat{f}$ is uniformly continuous [21]. Given $f \in L^2(\mathbb{R})$, the integral in $(\ 4.52)$ converges and the transform is

bounded. The inverse Fourier transform for $f \in L^1$ is defined as

$$f(t) = \left(\mathcal{F}^{-1} f\right)(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w) e^{iwt} dt. \tag{4.53}$$

Theorem 1. Let $f \in L^1$, then $\hat{f}$ satisfies the following

1. $\hat{f} \in L^\infty(\mathbb{R})$ with $\|\hat{f}\|_\infty \leq \|f\|_{L^1}$.
2. $\hat{f}$ is uniformly continuous on $\mathbb{R}$.
3. If f' exists and $f' \in L^1$, then $\hat{f}'(w) = iw\hat{f}(w)$.
4. Let $f, h \in L^1$, then their convolution $f \star h \in L^1$ and $\widehat{(f \star h)}(w) = \hat{f}(w)\hat{h}(w)$.
5. $\hat{f}(w) \to 0$ as $w \to \pm\infty$.

The two theorems of R. Paley and N. Wiener collectively referred to as *Paley-Wiener theorems* outlines conditions under which a function may be extended to a holomorphic function in some specified region. Lemma 4.6.1 identifies an important relationship between an $L^2(0, \infty)$ functions and its Fourier inverse on the upper half plane.

Lemma 4.6.1. *Let $F \in L^2(0, \infty)$, then f defined by*

$$f(z) = \int_0^\infty F(t) e^{2\pi z t} dt \tag{4.54}$$

is analytic in $\Pi^+ = \{z = x + iy : y > 0\}$.

Theorem 4.6.2. *(Paley-Wiener) [53] Suppose $f \in H(\Pi^+)$ (Holomorphic in Π^+) and*

$$\sup_{0<y<\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} |f(x + iy)|^2 dx = C < \infty. \tag{4.55}$$

Then there exists an $F \in L^2(0, \infty)$ such that

$$f(z) = \int_0^\infty F(t) e^{itz} dt \quad \left(z \in \Pi^+\right) \tag{4.56}$$

and

$$\int_0^\infty |F(t)|^2 dt = C. \tag{4.57}$$

Theorem 4.6.3. *[53] Suppose A and C are positive constants and f is an entire function such that*

$$|f(z)| \leq Ce^{A|z|} \tag{4.58}$$

for all z, and

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty. \tag{4.59}$$

Then there exists an $F \in L^2(-A, A)$ such that

$$f(z) = \int_{-A}^{A} F(t)e^{itz} dt \tag{4.60}$$

for all z.

From Theorem 4.6.2, the function F sought after has to have the property that $f(x+iy)$ is the Fourier transform of $F(t)e^{-yt}$. Suppose an inverse exist, then the desired F is of the form

$$F(t) = e^{ty}\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x+iy)e^{-itx}dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(z)e^{-itz}dz, \tag{4.61}$$

where the last integral is over $\{z = x + iy | y > 0\}$. As noted in the previous sections, the function $T_{\varphi,\gamma}$ is holomorphic in $\mathbb{D}$, and in fact $T_{\varphi,\gamma} \in \mathcal{H}^2$. To satisfy the hypothesis of the Paley-Wiener theorem, we construct an analytic extension of $T_{\varphi,\gamma}$ to the upper half plane so that $T_{\varphi,\gamma} \in \mathcal{H}^2(\mathbb{H}_+)$ (*Hardy space of the upper half plane*). This space is comprised of holomorpic functions f on $\mathbb{H}_+$ with a bounded norm.

$$\begin{aligned} \mathcal{H}^2(\mathbb{H}_+) &= \left\{ f \text{ analytic in } \mathbb{H}_+ : \sup_{y>0} \int |f(x+iy)|^2 dx < \infty \right\} \\ \|f\|_{\mathcal{H}^2(\mathbb{H}_+)} &= \sup_{y>0} \left(\int_{-\infty}^{\infty} |f(x+iy)|^2 dx \right)^{1/2}. \end{aligned} \tag{4.62}$$

The *Cayley transform* conformally maps the unit disk to the upper half plane. This establishes an isometric isomorphism between $\mathcal{H}^2(\mathbb{D})$ and $\mathcal{H}^2(\mathbb{H}_+)$. The Cayley map is defined as

$$\eta(z) = i\frac{1+z}{1-z}, \quad z \in \mathbb{D}, \quad z \neq 1 \tag{4.63}$$

with inverse

$$\eta^{-1}(z) = \frac{z+i}{z-i}, \quad z \in \mathbb{H}_+ \tag{4.64}$$

η maps the disk to the upper half plane and η^{-1} maps the upper half plane to the disk. The linear transform

$$H_f : \mathcal{H}^2(\mathbb{D}) \to \mathcal{H}^2(\mathbb{H}_+) \tag{4.65}$$

defined by

$$[H_f](z) = \frac{1-z}{\sqrt{\pi}} f(\eta^{-1}(z)) \tag{4.66}$$

maps a holomorphic function in the disk to a holomorphic function in the upper half plane so that given $T_{\varphi,\gamma} \in \mathcal{H}^2(\mathbb{D})$

$$\widetilde{T}_{\varphi,\gamma} = \left[H_{T_{\varphi,\gamma}}\right](z) = \frac{1-z}{\sqrt{\pi}} T_{\varphi,\gamma}(\eta^{-1}(z)) = \frac{1-z}{\sqrt{\pi}} \int_0^T \overline{\varphi(\gamma(t))} k(\eta^{-1}(z), \gamma(t)) dt \in \mathcal{H}^2(\mathbb{H}_+), \tag{4.67}$$

where

$$\|T_{\varphi,\gamma}\|_{\mathcal{H}^2(\mathbb{D})} = \|\widetilde{T}_{\varphi,\gamma}\|_{\mathcal{H}^2(\mathbb{H}_+)}. \tag{4.68}$$

By the Paley-wiener theorem,

$$\begin{aligned}
h(\zeta) &:= \int_{-\infty}^{\infty} \tilde{T}_{\varphi,\gamma}(z) e^{-2i\pi z\zeta} dz \\
&= \int_{-\infty}^{\infty} \frac{1-z}{\sqrt{\pi}} \left[\int_0^T \overline{\varphi(\gamma(t))} k(\eta^{-1}(z), \gamma(t)) dt \right] e^{-2i\pi z\zeta} dz \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \int_0^T (1-z) \overline{\varphi(\gamma(t))} k(\eta^{-1}(z), \gamma(t)) e^{-2i\pi z\zeta} dt dz \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \int_0^T (1-z) \frac{\overline{\varphi(\gamma(t))}}{(1-\eta^{-1}(z)\overline{\gamma(t)})} e^{-2i\pi z\zeta} dt dz \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \int_0^T \frac{(1-z)\overline{\varphi(\gamma(t))}(z-i)}{((z-i)-(z+i)\overline{\gamma(t)})} e^{-2i\pi z\zeta} dt dz \\
&= \frac{1}{\sqrt{\pi}} \int_0^T \overline{\varphi(\gamma(t))} \left[\int_{-\infty}^{\infty} \frac{(1-z)(z-i)e^{-2i\pi z\zeta}}{((z-i)-(z+i)\overline{\gamma(t)})} dz \right] dt,
\end{aligned} \tag{4.69}$$

and thus, $h(\zeta)$ is an analytic function. The integration against the the reproducing Szego kernel defines a projection of h onto $\mathcal{H}^2(\mathbb{D})$. Equivalently the map

$$H_h^* : \mathcal{H}^2(\mathbb{H}_+) \to \mathcal{H}^2(\mathbb{D}) \tag{4.70}$$

given by

$$\left[H_h^*\right](w) := \frac{\sqrt{\pi}}{1-z} h(\eta(w)) \tag{4.71}$$

is an isomorphism between the spaces $\mathcal{H}^2(\mathbb{H}_+)$ and $\mathcal{H}^2(\mathbb{D})$ so that

$$\begin{aligned}
\left[H_h^*\right](w) &= \frac{\sqrt{\pi}}{1-z} \frac{1}{\sqrt{\pi}} \int_0^T \overline{\varphi(\gamma(t))} \left[\int_{-\infty}^{\infty} \frac{(1-z)(z-i)e^{-2i\pi z\eta(w)}}{((z-i)-(z+i)\overline{\gamma(t)})} dz \right] dt \\
&= \frac{1}{1-z} \int_0^T \overline{\varphi(\gamma(t))} \left[\int_{-\infty}^{\infty} \frac{(1-z)(z-i)e^{-2i\pi z\eta(w)}}{((z-i)-(z+i)\overline{\gamma(t)})} dz \right] dt \\
&= \frac{1}{1-z} \int_0^T \overline{\varphi(\gamma(t))} \left[\int_{-\infty}^{\infty} \frac{(1-z)(z-i)e^{-2i\pi z\frac{i(1+w)}{1-w}}}{((z-i)-(z+i)\overline{\gamma(t)})} dz \right] dt.
\end{aligned} \tag{4.72}$$

Recovery of holomophic symbols along closed trajectories

Suppose we consider a sequence of closed trajectories in the form of concentric circles of specified radii, that is, let $\{r_n\}$ be a sequence of positive real numbers $r_n \nearrow 1$, and take the corresponding sequence of trajectories

$$\Omega = \left\{\gamma_n(t) := r_n e^{it}, 0 \leq t \leq 2\pi\right\}. \tag{4.73}$$

Suppose also that we can access the values $\{(\Lambda\varphi)_{\gamma_n}(z)\}$ for a holomorphic symbol φ, then we may recover the function φ as follows,

Proposition 4.6.4. *With the notation above, for $z \in \mathbb{D}_n = \{z : |z| < r_n\}$ such that $|z/r_n^2| < r_n$, then the symbol φ is expressed as*

$$\varphi(z) = \lim_{n\to\infty} \frac{1}{2\pi}(\Lambda\varphi)_{\gamma_n}(z/r_n^2), \tag{4.74}$$

where

$$\varphi_n(z) = \frac{1}{2\pi}(\Lambda\varphi)_{\gamma_n}(z/r_n^2).$$

and $(\Lambda\varphi)_{\gamma_n}(z)$ is the value of $\Lambda\varphi$ along the trajectory γ_n.

Proof. Fix $z \in \mathbb{D}_n$. By definition, for the trajectory $\gamma_n(t) = r_n e^{it}$

$$\begin{aligned}
[\Lambda]\varphi_n(z) &= \int_0^{2\pi} \varphi_n(r_n e^{it}) k_z(r_n e^{it}) dt \\
&= \int_0^{2\pi} \frac{\varphi(r_n e^{it})}{1 - z r_n e^{-it}} dt \\
(\zeta = r_n e^{i\theta}) &= \int_{|\zeta|=r_n} \frac{\varphi_n(\zeta)}{1 - z\bar{\zeta}} \frac{d\zeta}{i\zeta}, \qquad (k_z(w) = 1/(1 - w\overline{z})) \\
&= -i \int_{|\zeta|=r_n} \frac{\varphi_n(\zeta)}{\zeta - r_n^2 z} d\zeta \qquad \text{by Cauchy's integral formula} \\
&= 2\pi\varphi_n(r_n^2 z).
\end{aligned}$$

Therefore $[\Lambda]\varphi_n(z) = 2\pi\varphi(r_n^2 z)$, hence $\varphi_n(z) = \frac{1}{2\pi}(\Lambda\varphi_n)(z/r_n^2)$.

Precisely,

$$\varphi(z) = \lim_{n\to\infty} \varphi_n(z) = \lim_{n\to\infty} \frac{1}{2\pi}(\Lambda\varphi)_{\gamma_n}(z/r_n^2). \tag{4.75}$$

□

CHAPTER 5:

KERNEL RECONSTRUCTION FOR FREDHOLM OPERATORS

5.1 Fredholm Integral Equations For Real-Valued Functions

Throughout this section, we use the notations ($a < b, a, b \in \mathbb{R}$):

$$\mathcal{H} = L^2([a,b], dx), \quad (f,g) \equiv \int_a^b \overline{f(x)} g(x) dx, \quad ||f||_{\mathcal{H}}^2 \equiv (f,f), \quad \forall f, g \in \mathcal{H}.$$

Definition 5.1.1. A continuous kernel on $\mathcal{H}$ is a function $\kappa \in C(\Omega), \kappa : \Omega \to \mathbb{C}$, where $\Omega \equiv [a,b] \times [a,b]$. The associated integral operator $K \in L(\mathcal{H}, \mathcal{H})$ is defined by $K(f)(x) = \int_a^b \kappa(x,t) f(t) dt$.

Theorem 5.1.1. *The integral operator K is bounded, as $\forall f \in C[a,b]$,*

$$||K(f)||_{\mathcal{H}} \leq (b-a) \sup_{\Omega} |\kappa(x,t)| \sup_{[a,b]} |f(x)|.$$

Moreover, $K(f) \in C[a,b], \ \forall f \in C[a,b]$.

Definition 5.1.2. Let $f, g \in C[a,b]$ and $\kappa \in C(\Omega), \kappa : \Omega \to \mathbb{C}$ a continuous kernel on $\mathcal{H}$. For $\lambda \in \mathbb{C}^*$, the integral equation of Fredholm type with data g, κ is given by

$$f(x) = g(x) + \lambda \int_a^b \kappa(x,t) f(t) dt. \tag{5.1}$$

A solution for (5.1) is any pair (f, λ) satisfying the equation.

Theorem 5.1.2 (Preliminary result)**.** *For all λ in the disk*

$$\mathcal{D} = \left\{ \lambda \in \mathbb{C} || \lambda | < \frac{1}{(b-a) \sup_{[a,b]} |\kappa|} \right\},$$

Eq. (5.1) *has the unique solution* $f_\infty \in C[a,b]$ *given by*

$$f_\infty(x) = g(x) + \lambda \int_a^b R(x,t;\lambda) g(t) dt, \tag{5.2}$$

where

$$R(x,t;\lambda) \equiv \sum_{n=0}^{\infty} \kappa_{n+1}(x,t)\lambda^n,$$

and

$$\kappa_n(x,t) \equiv \int_a^b \kappa(x,u)\kappa_{n-1}(u,t)du, n \geq 2, \quad \kappa_1 = \kappa.$$

Remark 5.1. The function $R(x,t;\lambda)$ is the resolvent associated to the kernel $\kappa(x,t)$. As a function of λ, it is analytic inside the disk $\mathcal{D}$.

Remark 5.2. The resolvent $R(x,t;\lambda)$ and the kernel $\kappa(x,t)$ satisfy the identities

$$R(x,t;\lambda) = \kappa(x,t) + \lambda \int_a^b \kappa(x,u) R(u,t;\lambda) du \tag{5.3}$$

$$R(x,t;\lambda) = \kappa(x,t) + \lambda \int_a^b R(x,u;\lambda) \kappa(u,t) du. \tag{5.4}$$

The first important observation made by Fredholm extends the range of applicability of Theorem (5.1.2):

Theorem 5.1.3. *Assume there exists a function* $R(x,t;\lambda) : \Omega \times \mathbb{C}$ *such that, for some* λ_0*, it satisfies the identities* (5.3), (5.4)*. Then the equation* (5.1) *has a unique, continuous solution for* $\lambda = \lambda_0$*, and it is given by* (5.2)*.*

The second (very) important observation of Fredholm is that the resolvent can be obtained constructively:

Definition 5.1.3. Let

$$d_n = \int_a^b \underset{n \text{ times}}{\cdots} \int_a^b \det[\kappa(t_i,t_j)]_{1\leq i,j\leq n} dt_1 dt_2 \dots dt_n,$$

$$d_n(x,t) = \int_a^b \underset{n \text{ times}}{\cdots} \int_a^b \det[\kappa(x_i,t_j)]\Big|_{x_0=x, t_0=t, x_i=t_i, i\geq 1}^{0\leq i,j\leq n} dt_1 dt_2 \dots dt_n,$$

and define the functions

$$D(\lambda) = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} d_n \tag{5.5}$$

$$D(x,t;\lambda) = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} d_n(x,t). \tag{5.6}$$

Remark 5.3. The functions $D(x,t;\lambda), D(\lambda)$ are entire with respect to λ.

Theorem 5.1.4. *Any function $R(x,t;\lambda)$ satisfying* (5.3), (5.4) *is given by*

$$R(x,t;\lambda)D(\lambda) = D(x,t;\lambda). \tag{5.7}$$

Corollary 5.1.5. *For all $\lambda \in \mathbb{C}$ such that $D(\lambda) \neq 0$, equation* (5.1) *has a unique, continuous solution given by* (5.2)*, and resolvent* (5.7)*.*

The points $\lambda \in \mathbb{C}$ where the resolvent is not defined (and therefore the previous result does not hold) are the zeros of the (entire) function $D(\lambda)$. They form a countable set, and their only possible accumulation point is $\lambda = \infty$. They are called *eigenvalues* of the kernel κ because of the following:

Lemma 5.1.6. *Assume $\lambda = \lambda_0$ is a zero of order m for $D(\lambda)$. Then it may be a zero of order at most $m-1$ for $D(x,t;\lambda)$, so the resolvent $R(x,t;\lambda)$ has a pole at λ_0.*

Theorem 5.1.7. *Assume $\lambda = \lambda_0$ is a zero for $D(\lambda)$, and that the resolvent has the local Laurent expansion*

$$R(x,t;\lambda) = \frac{b_r(x,t)}{(\lambda-\lambda_0)^r} + O((\lambda-\lambda_0)^{-r+1}).$$

Then the function $b_r(x,t)$ is a solution for the homogenous equation

$$f(x) = \lambda_0 \int_a^b \kappa(x,t) f(t) dt = \lambda_0 K(f)(x). \tag{5.8}$$

The most general situation for equations of Fredholm type is obtained by analyzing the space of solutions of (5.1) when λ is an eigenvalue of the kernel κ. It requires a preliminary step, involving the *adjoint* of K:

Definition 5.1.4. The adjoint of the kernel κ is defined by $\kappa^*(x,t) = \overline{\kappa(t,x)}$. Correspondingly, the integral operators K, K^* satisfy the adjoint condition

$$(f, Kg) = (K^*f, g), \quad \forall f, g \in \mathcal{H}.$$

Theorem 5.1.8. *Assume that λ_0 is an eigenvalue for the kernel κ. The space of solutions for the homogenous equation*

$$V_{\lambda_0} \equiv \{f \in \mathcal{H} | f = \lambda_0 K(f)\}$$

is finite-dimensional. Moreover, $\overline{\lambda_0}$ is an eigenvalue for the adjoint kernel κ^*, and the space of solutions

$$V^*_{\overline{\lambda_0}} \equiv \{f \in \mathcal{H} | f = \overline{\lambda_0} K^*(f)\}$$

is isomorphic to V_{λ_0} (i.e., they have the same dimension).

Theorem 5.1.9. *Let λ_0 be an eigenvalue for the kernel κ, and f a solution for* (5.1)*. Then the function g is orthogonal to the space of eigenvectors $V^*_{\overline{\lambda_0}}$:*

$$(\psi, g) = 0, \quad \forall \psi \in V^*_{\overline{\lambda_0}}.$$

The most general situation is summarized in the following *alternative*:

Theorem 5.1.10. *A given $\lambda \in \mathbb{C}$ can be either an eigenvalue for the kernel κ (i.e., a zero of the entire function $D(\lambda)$), or a regular point for the resolvent $R(x,t;\lambda)$. If it is not an eigenvalue, then* (5.1) *has a unique, continuous solution given by* (5.2)*, and the homogenous equation* (5.8) *has only the trivial solution. If λ is an eigenvalue, then $R(x,t;\lambda)$ has a pole singularity at λ, and* (5.8) *has a finite-dimensional space of non-trivial solutions, V_λ. The adjoint homogenous equation also has a finite-dimensional space of solutions, $V^*_{\overline{\lambda}}$, of the same dimension as V_λ. Then* (5.1) *has solutions for λ if and only if $g \perp V^*_{\overline{\lambda}}$. Let $\{\phi_j\}_{j=1}^n$ be an orthonormal basis for V_λ and $\{\psi_j\}_{j=1}^n$ an orthonormal basis for $V^*_{\overline{\lambda}}$. Then λ is not an eigenvalue for the modified kernel*

$$\widetilde{\kappa}(x,t) = \kappa(x,t) - \sum_{j=1}^{n} \psi(x)\overline{\phi(t)},$$

so the equation $f(x) = g(x) + \lambda \int_a^b \widetilde{\kappa}(x,t) f(t) dt$ *has a unique, continuous solution* $\widetilde{f}$*. Finally, if* $g \perp V_{\overline{\lambda}}^*$*, then* (5.1) *has the solutions*

$$f = \widetilde{f} + \sum_{j=1}^{n} C_j \phi_j, \quad C_j \in \mathbb{C}.$$

5.2 General Theory of the Sturm-Liouville Problem

Let $p(x), q(x),\ f(x), g(x) \in C[a,b]$ be complex-valued functions on $[a,b] \subset \mathbb{R}$, sufficiently smooth, and such that $p(x) > 0$ on $[a,b]$. Take the linear differential operator

$$L \equiv \frac{d}{dx}\left(p(x)\frac{d}{dx}\right) + q(x),$$

and the associated equation

$$L(f) = g, \tag{5.9}$$

subject to the boundary-value (B.V.) conditions

$$\text{(I)}\ \alpha_1 f(a) + \alpha_2 f'(a) = 0, \quad \text{(II)}\ \beta_1 f(b) + \beta_2 f'(b) = 0, \tag{5.10}$$

such that $|\alpha_1| + |\alpha_2| \neq 0, |\beta_1| + |\beta_2| \neq 0$. Then the Sturm-Liouville problem (5.9), (5.10) can be converted to an integral equation of Fredholm type, and the corresponding alternative is discussed below.

5.2.1 No Nontrivial Solutions For the Homogeneous Problem

Assume that the equation $L(f) = 0$, subject to B.V. (5.10), has only the trivial solution $f = 0$. Let $f_{1,2}$ be solutions for the homogenous equation $L(f) = 0$, such that f_1 satisfies condition (I) from (5.10), and f_2 satisfies condition (II) from (5.10). As independent solutions for the homogenous problem, the Wronskian of $f_{1,2}$ does not vanish, and the functions can be chosen such that

$$W[f_1, f_2] = f_1' f_2 - f_2' f_1 = \frac{1}{p(x)}.$$

Then the solution to the problem $L(f) = g$, subject to B.V. (5.10), can be obtained by the method of variation of coefficients, and it is given by

$$f(x) = \int_a^b G(x,t)g(t)dt, \tag{5.11}$$

where

$$G(x,t) = f_1(x)f_2(t), \quad x \leq t, \quad G(x,t) = f_1(t)f_2(x), \quad x > t.$$

Moreover, the homogenous equation $L(f) = \lambda f$ is equivalent to the homogenous integral equation

$$f(x) = \lambda \int_a^b G(x,t)f(t)dt.$$

The kernel $G(x,t)$ is Hermitian (self-adjoint) and has an infinite, countable set of real eigenvalues, whose accumulation point is $\lambda = \infty$.

5.2.2 Non-Trivial Solutions For the Homogenous Equation

If the problem $L(f) = 0$ subject to (5.10) does have a non-trivial solution, it is always possible to find $\lambda_0 \in \mathbb{R}$ such that $L_0 = L + \lambda_0$ falls under the situation described above. Denoting by $\{\mu_k\}$ the spectrum of the kernel G_0 associated to L_0, the eigenvalues of L are given by $\lambda_k = \lambda_0 + \mu_k, k \in \mathbb{N}$.

Remark 5.4. To prepare for the inverse problem associated with this family of operators, we note the useful orthogonal decomposition of the operator L,

$$L = L^{(0)} + \tilde{L} := \widehat{P}L\widehat{P} + (1 - \widehat{P})L(1 - \widehat{P}),$$

where $\widehat{P}$ represents the (finite-rank) projector onto $Ker(L)$, such that

$$[L^{(0)}, \tilde{L}] = 0.$$

5.3 Inverse Problem For Kernels of Sturm-Liouville Operators

Theorem 5.3.1. *Assume that $q < 0$ a.e. on $[a, b]$, then*

$$L(f) = g \Leftrightarrow f = \arg \min_{\gamma \in C^2(a,b)} S[\gamma], \quad (\star)$$

$$S[\gamma] = -\int_a^b \left[p(x)(\gamma'(x))^2 - q(x)\gamma(x)^2 + g(x)\gamma(x) \right] dx,$$

subject to either pure Dirichlet or Neumann B.V. conditions on $\{a, b\}$.

Proof. With pure Dirichlet or Neumann B.V. conditions at $x = a, b$, the variational problem is equivalent to solving the Euler-Lagrange equations for the functional $S[\gamma]$. We first rewrite its integrand in the form

$$\left[p(x)(\gamma'(x))^2 - q(x)\gamma(x)^2 + g(x)\gamma(x) \right] = \Phi'(x) - \gamma(x) \left\{ \left[p(x)\gamma'(x) \right]' + q(x)\gamma(x) - g(x) \right\},$$

where

$$\Phi(x) = \gamma(x)p(x)\gamma'(x),$$

or

$$S[\gamma] = \int_a^b \gamma(x) \left\{ \left[p(x)\gamma'(x) \right]' + q(x)\gamma(x) - g(x) \right\} dx + \Phi(x)\Big|_a^b$$

At fixed B.V. data, the first variation of the functional has no contribution from the term boundary term and it implies for the minimizer γ_*

$$\left[p(x)\gamma_*'(x) \right]' + q(x)\gamma_*(x) - g(x) = 0.$$

Existence of the minimizer is guaranteed by the fact that the integrand is a non-degenerate quadratic form with strictly-positive definite principal symbol. □

Remark 5.5. The variational formulation also allows to carry over the decomposition noted in Remark 5.4 to the decomposition of the general solution for the Sturm-Liouville problem into its homoge-

neous and inhomogeneous parts

$$f[g] = f^{(0)} + \tilde{f}[g], \quad f^{(0)} \in \ker(L),\ \tilde{f}[0] = 0,$$

leading to the variational derivatives

$$-\frac{\delta S}{\delta g(x)} = \tilde{f}(x), \quad -\frac{\delta^2 S}{\delta g(y)\delta g(x)} = \frac{\delta \tilde{f}(x)}{\delta g(y)} = G(x,y),$$

consistent with Eq. (5.11).

5.3.1 Reconstruction For the Equilibrium Distribution Class

We will now pose the kernel reconstruction inverse problem for a convex subset of the Hilbert space $\mathcal{H}$, as the (unique) solution of Problem $(\star)$ on

$$K = \{f \in \mathcal{H} | f \geq 0\, a.e., ||f||_1 \leq 1\}.$$

More precisely, we seek to find $\rho : [0,\infty) \times [a,b], \rho(t,.) \in K \cap C^2(a,b)$, such that

$$\frac{\partial \rho}{\partial t} = L(\rho), \quad \rho(0,.) = \rho_0 \in K \cap C^2(a,b),$$

subject to B.V. conditions (I, II). Reconstructing the kernel for the Sturm-Liouville operator from this family of solutions (functionally dependent on the set of initial-value data) is the goal for the remaining part of this chapter. We will identify classes of kernels for which the reconstruction is possible and characterize the convergence in the sense of approximation theory, with emphasis on applicability for efficient numerical implementations.

Theorem 5.3.2. *For elliptic $\tilde{L}$, with spectrum $\{\lambda_n\}_{n=1}^{\infty} \subset (-\infty, 0)$, $\lambda_n \to -\infty$,*

$$\rho(t,x) = \left[e^{t\tilde{L}}\rho_0\right](x) = \sum_{n=1}^{\infty} c_n \phi_n(x) e^{t\lambda_n},$$

where

$$\tilde{L}\phi_n = \lambda_n \phi_n,\ \phi_n \in C^2(a,b), \quad \sum_{n=1}^{m} c_n \phi_n \xrightarrow{\|\cdot\|_2} \rho_0.$$

Then the Baiocchi transform of $\rho(t,.)$ provides the pointwise reconstruction of the kernel $G(x,y)$ by

$$G(x,y) = \lim_{\epsilon \to 0} \bar{\rho}_y^{(\epsilon)}(x), \tag{5.12}$$

where

$$\bar{\rho}_y^{(\epsilon)}(x) \equiv \lim_{T \to \infty} \int_0^T \rho_y^{(\epsilon)}(t,x)dt,$$

and $\rho_y^{(\epsilon)}$ solves the initial-value problem with normal initial data

$$\frac{\partial \rho_y^{(\epsilon)}}{\partial t} = L(\rho_y^{(\epsilon)}), \quad \rho_y^{(\epsilon)}(0,.) = N(y,\epsilon),\ \epsilon > 0.$$

Proof. Since $\rho, \rho_0 \in K \cap C^2(a,b)$, it implies that $\rho_0 \in L^\infty(a,b) \subset L^2(a,b)$ and there are real coefficients $\{c_n\}_{n=1}^\infty$ such that, in the L^2 norm, as $m \to \infty$,

$$\rho_0 - \sum_{n=1}^{m} c_n \phi_n \longrightarrow \rho^{(0)} \in \ker(L).$$

For $t \geq 0$, the operator $U(t) = \exp(tL)$ is bounded with operator norm $\|U(t)\|_* \leq 1$, so the solution to the initial value problem reads $\rho(t,.) = e^{tL}\rho_0(.)$. By the orthogonal decomposition noted in Remark 5.4, we find by expanding

$$\rho(t,.) = \rho^{(0)} + \sum_{n=1}^{\infty} c_n \phi_n(x) e^{t\lambda_n},$$

and therefore in the limit $t \to \infty$,

$$\lim_{t \to \infty} \rho(t,.) = \rho^{(0)}$$

solves the homogenous Sturm-Liouville problem. But under our assumption the only solution is trivial, which implies that

$$\rho(t,.) = e^{tL} \sum_{n=1}^{\infty} c_n \phi_n(.) = e^{t\tilde{L}} \sum_{n=1}^{\infty} c_n \phi_n(.),$$

proving the first part of the theorem.

For the second part, let us note that $\tilde{L}$ is an invertible self-adjoint, negative-definite elliptic operator, and therefore

$$\int_0^T \rho(t,x)dt = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} c_n \phi_n(x)[e^{T\lambda_n} - 1],$$

which implies that the Baiocchi transform of the solution exists and has the spectral form

$$\overline{\rho}(x) \equiv \lim_{T\to\infty} \int_0^T \rho(t,x)dt = -\sum_{n=1}^{\infty} \frac{1}{\lambda_n} c_n \phi_n(x) = [R_0(L)](\rho_0),$$

where $R_\zeta(L)$ denotes the Fredholm resolvent of the operator L at ζ. Therefore,

$$\overline{\rho}(x) = \int_a^b G(x,y)\rho_0(y)dy.$$

For the case of initial-value data $\rho_y^{(\epsilon)}(0,.) = N(y,\epsilon)$, $\epsilon > 0$, this leads to the family of Baiocchi transforms

$$\overline{\rho}_y^{(\epsilon)} = \int_a^b G(x,z)\rho_y^{(\epsilon)}(z)dz,$$

and in the limit $\epsilon \to 0^+$, to the pointwise evaluation (5.12) and the completion of the proof. □

5.4 Conclusion and Remarks

In this work we have presented two main ideas. Firstly, we have developed a parameter identification method using the Nonlinear Autoregressive (NAR) model architecture over a reproducing kernel Hilbert space. Specifically, we leveraged the interaction between occupation kernels and densely-defined multiplication operators to develop a model that determine system parameter from data measurements. Two classes of symbols of the operator were considered; scalar-valued and vector-valued symbols. Owing to the experimental results presented in this work, we foresee possible generalization and improvements of the current framework. For instance, incorporating noise-damping factors and introducing essential regularizing terms promise a more stabilized learning algorithm that will be suitable for a wide range of dynamics.

Secondly, we have presented a kernel reconstruction for the Fredholm integral operator. Motivated by the general theory of Sturm-Liouville differential equations, we introduced an inverse problem that relates the reconstruction of the kernel of a Fredholm operator to solving an associated Sturm Liouville Problem.

REFERENCES

[1] Jim Agler, John Edward McCarthy, and Linus E Kramer. *Pick interpolation and Hilbert function spaces*. Vol. 44. American Mathematical Soc., 2002.

[2] Nachman Aronszajn. "Theory of reproducing kernels". In: *Transactions of the American mathematical society* 68.3 (1950), pp. 337–404.

[3] SA Billings and MB Fadzil. "The practical identification of systems with nonlinearities". In: *IFAC Proceedings Volumes* 18.5 (1985), pp. 155–160.

[4] SA Billings and IJ Leontaritis. "Parameter estimation techniques for nonlinear systems". In: *IFAC Proceedings Volumes* 15.4 (1982), pp. 505–510.

[5] Stephen A Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[6] Paul S Bourdon. "Similarity of parts to the whole for certain multiplication operators". In: *Proceedings of the American Mathematical Society* 99.3 (1987), pp. 563–567.

[7] Bingni W Brunton et al. "Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition". In: *Journal of neuroscience methods* 258 (2016), pp. 1–15.

[8] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the national academy of sciences* 113.15 (2016), pp. 3932–3937.

[9] Sheng Chen and Steve A Billings. "Representations of non-linear systems: the NARMAX model". In: *International journal of control* 49.3 (1989), pp. 1013–1032.

[10] D Coca and SA Billings. "Non-linear system identification using wavelet multiresolution models". In: *International Journal of Control* 74.18 (2001), pp. 1718–1736.

[11] Warren E Dixon et al. *Nonlinear control of engineering systems: a Lyapunov-based approach*. Springer Science & Business Media, 2003.

[12] Timothy G Feeman. "The mathematics of medical imaging". In: *Sringer,* (2010).

[13] Herbert I Freedman. *Deterministic mathematical models in population ecology*. Vol. 57. Marcel Dekker Incorporated, 1980.

[14] Stephan Ramon Garcia, Javad Mashreghi, William T Ross, et al. *Finite Blaschke products and their connections*. Springer, 2018.

[15] G-OA Glentis, Panos Koukoulas, and Nicholas Kalouptsidis. "Efficient algorithms for Volterra system identification". In: *IEEE Transactions on Signal Processing* 47.11 (1999), pp. 3042–3057.

[16] Israel Gohberg, Seymour Goldberg, and Marius A Kaashoek. *Classes of linear operators*. Vol. 63. Springer, 2013.

[17] Brian C Hall. *Quantum theory for mathematicians*. Vol. 267. Springer, 2013.

[18] John K Hunter and Bruno Nachtergaele. *Applied analysis*. World Scientific Publishing Company, 2001.

[19] Alexandre Janot, Maxime Gautier, and Mathieu Brunot. "Data set and reference models of EMPS". In: *Nonlinear System Identification Benchmarks*. 2019.

[20] Rushikesh Kamalapurkar et al. *Reinforcement learning for optimal feedback control*. Springer, 2018.

[21] Yitzhak Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, 2004.

[22] I Kevrekidis, Clarence W Rowley, and M Williams. "A kernel-based method for data-driven Koopman spectral analysis". In: *Journal of Computational Dynamics* 2.2 (2016), pp. 247–265.

[23] HK Khalil. *Nonlinear Systems Third Edition (2001)*.

[24] Farhan Mohammad Khan and Rajiv Gupta. "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India". In: *Journal of Safety Science and Resilience* 1.1 (2020), pp. 12–18.

[25] Dhruv Khandelwal, Maarten Schoukens, and Roland Toth. "On the Simulation of Polynomial NARMAX Models". In: *2018 IEEE Conference on Decision and Control (CDC)*. 2018, pp. 1445–1450. DOI: 10.1109/CDC.2018.8619372.

[26] J Nathan Kutz et al. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[27] EG Ladopoulos. *Singular integral equations: linear and non-linear theory and its applications in science and engineering*. Springer Science & Business Media, 2013.

[28] Stéphane Mallat. "A wavelet tour of signal processing". In: Elsevier, 1999, pp. 45–70.

[29] Alexandre Mauroy and Jorge Goncalves. "Koopman-based lifting techniques for nonlinear systems identification". In: *IEEE Transactions on Automatic Control* 65.6 (2019), pp. 2550–2565.

[30] Igor Mezić. "Analysis of fluid flows via spectral properties of the Koopman operator". In: *Annual review of fluid mechanics* 45 (2013), pp. 357–378.

[31] Igor Mezić. "Spectral properties of dynamical systems, model reduction and decompositions". In: *Nonlinear Dynamics* 41.1 (2005), pp. 309–325.

[32] Igor Mezić and Andrzej Banaszuk. "Comparison of systems with complex behavior". In: *Physica D: Nonlinear Phenomena* 197.1-2 (2004), pp. 101–133.

[33] Charles A Micchelli and Massimiliano Pontil. "On learning vector-valued functions". In: *Neural computation* 17.1 (2005), pp. 177–204.

[34] Jeff Moehlis. *Dynamical Systems in Neuroscience: The geometry of excitability and bursting*. 2008.

[35] Oliver Nelles. *Nonlinear system identification: from classical approaches to neural networks, fuzzy models, and gaussian processes*. Springer Nature, 2020.

[36] Rolf Nevanlinna. "Uber beschrankte Funktionen, die in gegeben Punkten vorgeschrieben Werte annehmen". In: *Ann. Acad. Sci. Fenn. Ser. A 1 Mat. Dissertationes* (1919).

[37] Anup Parikh, Rushikesh Kamalapurkar, and Warren E Dixon. "Integral concurrent learning: Adaptive control with parameter convergence using finite excitation". In: *International Journal of Adaptive Control and Signal Processing* 33.12 (2019), pp. 1775–1787.

[38] Vern I Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*. Vol. 152. Cambridge university press, 2016.

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

[40] Georg Pick. "Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden". In: *Mathematische Annalen* 77.1 (1915), pp. 7–23.

[41] Joshua L Proctor and Philip A Eckhoff. "Discovering dynamic patterns from infectious disease data using dynamic mode decomposition". In: *International health* 7.2 (2015), pp. 139–145.

[42] R. Rado. In: *The Mathematical Gazette* 20.239 (1936), pp. 215–217. ISSN: 00255572. URL: http://www.jstor.org/stable/3608084 (visited on 02/06/2023).

[43] Mrinal Raghupathi. *Constrained Nevanlinna-Pick interpolation*. University of Houston, 2008.

[44] Alireza Rahrooh and Scott Shepard. "Identification of nonlinear systems using NARMAX model". In: *Nonlinear Analysis: Theory, Methods & Applications* 71.12 (2009), e1198–e1202.

[45] Michael Renardy and Robert C Rogers. *An introduction to partial differential equations*. Vol. 13. Springer Science & Business Media, 2006.

[46] Joel A Rosenfeld and Rushikesh Kamalapurkar. "Dynamic mode decomposition with control liouville operators". In: *IFAC-PapersOnLine* 54.9 (2021), pp. 707–712.

[47] Joel A Rosenfeld et al. "Dynamic mode decomposition for continuous time systems with the Liouville operator". In: *Journal of Nonlinear Science* 32 (2022), pp. 1–30.

[48] Joel A Rosenfeld et al. "Occupation kernels and densely defined liouville operators for system identification". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 6455–6460.

[49] Joel A Rosenfeld et al. "On occupation kernels, Liouville operators, and dynamic mode decomposition". In: *2021 American Control Conference (ACC)*. IEEE. 2021, pp. 3957–3962.

[50] Joel A Rosenfeld et al. "The occupation kernel method for nonlinear system identification". In: *arXiv preprint arXiv:1909.11792* (2019).

[51] Joel A. Rosenfeld et al. "Occupation Kernels and Densely Defined Liouville Operators for System Identification". In: *2019 IEEE 58th Conference on Decision and Control (CDC)* (2019), pp. 6455–6460.

[52] Joel A. Rosenfeld et al. "On Occupation Kernels, Liouville Operators, and Dynamic Mode Decomposition". In: *2021 American Control Conference (ACC)*. 2021, pp. 3957–3962. DOI: 10.23919/ACC50511.2021.9483121.

[53] Walter Rudin. "Real and complex analysis. 1987". In: *Cited on* 156 (1987), p. 16.

[54] Saburou Saitoh. "Theory of reproducing kernels". In: *Analysis and Applications—ISAAC 2001*. Springer, 2003, pp. 135–150.

[55] Donald Sarason. "Generalized Interpolation in H". In: *Transactions of the American Mathematical Society* 127.2 (1967), pp. 179–203. ISSN: 00029947. URL: http://www.jstor.org/stable/1994641.

[56] Donald Sarason. "Operator-Theoretic Aspects of the Nevanlinna-Pick Interpolation Problem". In: *Operators and Function Theory*. Ed. by S. C. Power. Dordrecht: Springer Netherlands, 1985, pp. 279–314. ISBN: 978-94-009-5374-1. DOI: 10.1007/978-94-009-5374-1_9. URL: https://doi.org/10.1007/978-94-009-5374-1_9.

[57] Donald Sarason. "Unbounded Toeplitz operators". In: *Integral Equations and Operator Theory* 61.2 (2008), pp. 281–298.

[58] Peter J Schmid. "Dynamic mode decomposition of numerical and experimental data". In: *Journal of fluid mechanics* 656 (2010), pp. 5–28.

[59] JG Skellam, MS Bartlett, and RW Hiorns. "The mathematical theory of the dynamics of biological populations". In: *by MS Bartlett and RW Hiorns, Academic Press, New York* (1973), p. 63.

[60] Elias M Stein and Rami Shakarchi. *Complex analysis*. Vol. 2. Princeton University Press, 2010.

[61] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[62] DAVID E Stewart, ROBERT L Dewar, et al. "Non-linear dynamics". In: *Complex systems* (2000), p. 167.

[63] Dragan Vukotić. "Analytic Toeplitz operators on the Hardy space H^p: a survey". In: *Bulletin of the Belgian Mathematical Society-Simon Stevin* 10.1 (2003), pp. 101–113.

[64] H-L Wei et al. "Forecasting relativistic electron flux using dynamic multiple regression models". In: *Annales geophysicae*. Vol. 29. 2. Copernicus GmbH. 2011, pp. 415–420.

[65] Kehe Zhu. "Operator theory in function spaces". In: 138. American Mathematical Soc., 2007, pp. 133–160.

[66] Kehe Zhu. *Operator theory in function spaces*. 138. American Mathematical Soc., 2007.