# Kernel Methods for Learning Dynamical Systems

Boya Hou[1], Subhonmesh Bose[2], Nathan Dahlin[3] and Umesh Vaidya[4]

[1] *University of Illinois Urbana-Champaign; boyahou2@illinois.edu*
[2] *University of Illinois Urbana-Champaign;boses@illinois.edu*
[3] *University at Albany, SUNY; ndahlin@albany.edu*
[4] *Clemson University;uvaidya@clemson.edu*

ABSTRACT

A dynamical system is a mathematical framework that describes how the system evolves through time. One can learn the mechanism of this evolution from data. A popular approach is to consider a parametric representation of that evolution mechanism and learn said parameters from data. Such an approach can be limiting for complex, high-dimensional, and stochastic systems. In this monograph, we describe a non-parametric operator-theoretic approach to that representation and consider the question of learning that description from data. These "transfer operators" define how a dynamical system moves functions or probability measures over states through time. While the operators themselves are infinite-dimensional, they are also linear, and this linearity allows investigation of the properties of the dynamical systems through their spectra.

We study the interaction of the Koopman and the Perron-Frobenius transfer operators with a reproducing kernel Hilbert space (RKHS). Under suitable assumptions, these operators can be represented as elements of a vector-valued

Boya Hou, Subhonmesh Bose, Nathan Dahlin and Umesh Vaidya (2025), "Kernel Methods for Learning Dynamical Systems", : Vol. XX, No. XX, pp 1–194. DOI: XXX.

RKHS, opening doors to learning their kernel approximations. We describe two approaches to learning these kernel approximations from data, one via sample average approximation and another via stochastic approximation, and provide sample complexity guarantees for the approximation quality.

As is well-known for function learning via kernel methods, learned non-parametric descriptions of functions, and in our case operators, grow with the size of the data. To counter this difficulty, we review various techniques to reduce redundancy in datasets and find a sparse non-parametric description of a dynamical system with data. Again, we provide sample complexity guarantees for the approximation quality of learning that sparse representation. A key element of our analysis in this monograph is to address the important mathematical consequences of model mis-specification, by which we mean that the action of the operator on the RKHS may not be closed. In a way, this monograph attempts to transport techniques from statistical learning in RKHS and employ them to study properties of non-linear dynamical systems from data.

Our development in this manuscript is largely self-contained. It provides the necessary details of integral operators and RKHS theory that our operator-theoretic description is built upon. The theoretical developments are elementary, targeting an audience in systems and control theory, signal processing, and statistical machine learning. We also provide illustrative numerical examples of the algorithms from specific application domains. We conclude this manuscript with a discussion of open research questions within the broad scope of learning dynamical systems from data.

# Contents

# 1

# Introduction



## 1.1 Dynamical Systems and Model Representation

Many real-world phenomena in biological processes, engineered systems, and social behaviors exhibit temporal patterns. These patterns can be viewed as a dynamical system–a mathematical framework that describes how "states" of a system evolve over time. Conceptually, the state encapsulates system memory, and the state-space model provides an internal description of a system. The study of dynamical systems has its origins in the works of Poincaré and Birkhoff (Poincaré, 1899; Birkhoff, 1927). Rooted in celestial mechanics, these works described dynamical systems (in continuous time) through differential equations–a tradition that has continued till the present day. In this approach, these equations are derived from physical, biological, economic, or other scientific laws. For example, descriptions of electromechanical systems can be obtained using Maxwell's equations and Newton's laws. These descriptions allow bottom-up modeling of complex interacting systems, where component descriptions obtained using scientific laws such as Maxwell's equations can be juxtaposed to obtain a description of the whole system. These descriptions are inherently parametric. For example, the motion of a block of mass $m$ under the action of a force $F$ is described by the

dynamical system $\ddot{x} = F/m$, where $x$ suitably captures the position of the block and hence, $\ddot{x}$ is its acceleration. The mass of the block $m$ is a parameter one can obtain through experimentation of different forces on the block and measuring its acceleration. Thus, arriving at a differential equation model requires the knowledge of foundational scientific laws that govern the system as a whole or its components, followed by measurements to ascertain the model parameters. Naturally, this modeling approach suffers when one is faced with studying systems, where either prior knowledge of the underlying structure is not available or the model description becomes way too complex when it is built up component-by-component. With the presumption of a fixed parametric form for the system description, inaccuracies can result when the true system falls outside the chosen model class. In this monograph, we ask the question if there is a direct way to learn a system model directly from abundant measurements collected from the trajectories of the evolution of the system without assuming a fixed parametric form, thereby circumventing the challenges of a parametric system description derived from scientific modeling laws.

A dynamical system defines how quantities pertaining to a system at a point in time are mapped to the same at a future point in time. For a system that evolves over discrete time steps, if all such quantities at time $t$ are captured in the state $x(t)$, then the dynamical system is the law that describes how $x(t+1)$ can be obtained from $x(t)$. A difference equation, the discrete-time counterpart of a differential equation, precisely defines such a law. Instead of a fixed parameterized functional form for such a law through a difference equation, one can view the dynamical system to be a mechanism that maps functions of or measures over states at time $t$ to that at time $t+1$. This mechanism then assumes the form of an infinite-dimensional operator that acts on spaces of functions of the state or measures defined over the state space. These "transfer" operators, formally known under the names of the Koopman and the Perron-Frobenius operators, then define a dynamical system model. As will soon become clear from their formal descriptions, these operators are linear. Consequently, they can be studied through their spectra. In this monograph, we focus on learning these operators directly from data. Operators are infinite-dimensional objects that act

on suitably-defined function spaces and space of probability measures or embeddings thereof. For our study, we focus on learning the actions of these operators on reproducing kernel Hilbert spaces (RKHSs), the rich theory of which has come to define the foundation of "kernel methods" in statistical learning. By design, the approach is data-centric and non-parametric. Any tractable learned model of the target transfer operator must ultimately have a finite description, and as a result, can only be an approximation. In this monograph, we review algorithms to learn these approximations from data, but then provide an estimate of the quality of that approximation by leveraging tools from functional analysis, measure theory, and statistical learning theory. We attempt to provide a comprehensive overview of the state of the art in operator-theoretic and kernel-based approach to learning dynamical systems and discuss promising research directions in the same.

## 1.2 A Gentle Introduction to Transfer Operators

Consider a discrete-time deterministic dynamical system, $x_{t+1} = T(x_t)$, $t \in \mathbb{N}$, where $T : \mathbb{X} \to \mathbb{X}$, $t \in \mathbb{N}$, and $\mathbb{X} \subseteq \mathbb{R}^n$ is the state space. The action of the Koopman operator on a function $g : \mathbb{X} \to \mathbb{C}$ is defined as

$$(Kg)(x_t) = (g \circ T)(x_t) = g(T(x_t)) = g(x_{t+1}), \quad t \in \mathbb{N}. \tag{1.1}$$

This operator was introduced by Bernard Koopman in (Koopman and Neumann, 1932) to study Hamiltonian dynamical systems. Indeed, it captures what the dynamics does to functions of the state, often referred to as *observables*. While technically, the operator can be studied for complex-valued functions, in this monograph, we will largely focus on its action over real-valued function spaces.

The (stochastic) Koopman operator proposed in (Mezić, 2005) generalizes the above to discrete-time Markov processes. Formally, let $\{X_t\}_{t\in\mathbb{N}}$ be an $\mathbb{X} \subseteq \mathbb{R}^n$-valued time-homogeneous Markov process defined on a filtered probability space $(\Omega, \mathbb{X}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \rho)$, where $X_t$ is the $\mathcal{F}_t$-adapted state of the system at time $t$. Such a stochastic dynamical system can be characterized by the transition kernel $P(\cdot|x)$. Suppose the transition kernel $P(\cdot|x)$ is absolutely continuous with respect to the Lebesgue measure and it admits a transition density function $p$ for almost all

$x \in \mathbb{X}$ such that

$$P\left(X_{t+1} \in \mathbb{A} | X_t = x\right) = \int_{\mathbb{A}} p(y|x) \mathrm{d}y, \tag{1.2}$$

for all measurable $\mathbb{A} \subseteq \mathbb{X}$. Let $\mathcal{G}$ be the space of scalar-valued measurable functions and $\rho_X$ a measure on $\mathbb{X}$. The Koopman operator $K : \mathcal{G} \to \mathcal{G}$ with respect to measure $\rho_X$ is defined as

$$\left(Kf\right)(x) = \mathbb{E}\left[f\left(X_{t+1}\right) | X_t = x\right], \quad f \in \mathcal{G}. \tag{1.3}$$

This operator acts on functions of states and not on states themselves; as a result, $K$ is infinite-dimensional. Consider two functions $g_1, g_2 \in \mathcal{G}$ and two scalars $a_1, a_2 \in \mathbb{R}$. Using (3.70), we have

$$\begin{aligned} K\left(a_1 g_1 + a_2 g_2\right)(x) =& \mathbb{E}\left[\left(a_1 g_1 + a_2 g_2\right)(y) | x_t = x\right] \\ =& a_1 \mathbb{E}\left[g_1(y) | x_t = x\right] + a_2 \mathbb{E}\left[g_2(y) | x_t = x\right] \\ =& a_1 K g_1(x) + a_2 K g_2(x). \end{aligned} \tag{1.4}$$

A similar calculation holds for deterministic dynamical systems. As a result, the infinite-dimensional Koopman operator is *linear*, even when the underlying dynamical system can be nonlinear in how states map from one time step to the next. This linearity permits the use of linear operator theory to study dynamical systems through their spectra. In particular, several works by Igor Mezic have established precise connections between spectral properties of the Koopman operator and properties of the dynamical system, e.g., see (Mezić and Banaszuk, 2004; Mezić, 2005; Mezić, 2020; Mezić, 2021) and the references therein.

The Perron-Frobenius (PF) operator is closely related to the Koopman operator. Let $\rho_X$ be a probability measure over $\mathbb{X}$. The PF operator $\mathcal{P}$ is defined as

$$\left(\mathcal{P}\rho_X\right)(y) = \int P(y|x) \rho_X(\mathrm{d}x). \tag{1.5}$$

The PF operator is also infinite-dimensional and linear. We refer interested readers to (Huang and Vaidya, 2022; Moyalan *et al.*, 2023) for a more thorough study of the PF operator.

The Koopman operator and the PF operator form an adjoint pair in appropriate function spaces (Lasota and Mackey, 2013). To illustrate through an example, consider that the Koopman operator acts on the

space of bounded measurable functions $L_\infty(\mathbb{X})$ and assume $L_\infty(\mathbb{X})$ is closed under the action of $K$, i.e., $K : L_\infty(\mathbb{X}) \to L_\infty(\mathbb{X})$. Similarly, let the PF operator $\mathcal{P}$ act on the Banach space of finite signed measures $\mathcal{M}(\mathbb{X})$, and assume $\mathcal{M}(\mathbb{X})$ is closed under the action of the PF operator, i.e., $\mathcal{P} : \mathcal{M}(\mathbb{X}) \to \mathcal{M}(\mathbb{X})$. Then, $K$ and $\mathcal{P}$ form an adjoint pair since for $h \in L_\infty(\mathbb{X})$, $g \in \mathcal{M}(\mathbb{X})$, we have

$$\begin{aligned}\int (Kh)(x)\rho_X(\mathrm{d}x) &= \int \int h(y)P(\mathrm{d}y|x)\rho_X(\mathrm{d}x) \\ &= \int h(y) \left( \int P(\mathrm{d}y|x)\rho_X(\mathrm{d}x) \right) \\ &= \int h(y)(\mathcal{P}\rho_X)(\mathrm{d}y).\end{aligned} \tag{1.6}$$

In the derivation, we use Fubini's theorem to exchange the integrals. Said plainly, the Koopman operator governs the evolution of functions, while the Perron-Frobenius operator evolves probability measures over the state space forward. Since the Koopman and the Perron-Frobenius operators provide equivalent descriptions of a dynamical system, in the following, we primarily focus on the Koopman operator as the representation of choice.

Why is an infinite-dimensional linear operator-theoretic description of a dynamical system useful? The power of this framework becomes apparent from the definition of the Koopman operator in (3.70). Roughly, one can estimate this expectation in the definition of the operator using data of the states separated by one time-step. Therefore, the Koopman operator allows estimation of the system model from data, whose spectrum then allows inference of qualitative information about the nature of the dynamics. A variety of algorithms exist that approximate this operator from data, such as dynamic mode decomposition (DMD) (Schmid, 2010; Rowley *et al.*, 2009), and its extended version (Williams *et al.*, 2015). These procedures do not rely on explicit knowledge of the equations that describe the state evolution from one time-step to the next. As a result, in contrast to approaches such as (Sattar and Oymak, 2022; Mania *et al.*, 2022; Foster *et al.*, 2020; Kowshik *et al.*, 2021), which assume specific near-linear behavior, the Koopman framework supports learning and analysis of general nonlinear systems.

## 1.3 Learning With Kernels

The Koopman operator is studied through its interaction with a function space, and the choice of that space dictates how well the system dynamics encoded in the operator can be analyzed. Of the existing *parametric* techniques that learn the Koopman operator, extended dynamic mode decomposition (EDMD) (Williams *et al.*, 2015) is perhaps the most widely used, where the function space is the span of a fixed pre-selected basis of functions. One then aims to learn how the Koopman operator moves functions in this space to other functions in the same space. It is plausible that functions from this space map to a function outside of this space, i.e., the function space is *not closed* under the dynamics and is referred to as the *mis-specified* setting. Then, the target operator becomes a solution of a regression problem that minimizes the error between the one-step propagated maps and its distance to the subspace spanned by the basis.

The EDMD method subscribes to a pre-defined basis of functions. As one might expect, such a selection can be difficult to approach systematically. To circumvent that limitation, we adopt the non-parametric route in this manuscript and attempt to learn the Koopman operator through its interaction with a reproducing kernel Hilbert space (RKHS) as in (Williams *et al.*, 2014; Kawahara, 2016; Klus *et al.*, 2020c; Hou *et al.*, 2023b; Kostic *et al.*, 2022; Bevanda *et al.*, 2023; Köhne *et al.*, 2025). An RKHS is a Hilbert space of functions in which the evaluation functional is continuous, and equivalently admits a reproducing kernel. See Chapter 2 for a detailed introduction. RKHSs offer a rich set of functions and have a myriad of properties that make analysis of the Koopman operator easy. To motivate our operator learning using RKHS, consider the classical problem of learning an unknown function $f$ in RKHS $\mathcal{H}$ from samples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$. One can case this problem as a regularized least squares problem of the form,

$$\underset{f \in \mathcal{H}}{\text{minimize}} \sum_{i=1} \frac{1}{N} \|f(x_i) - y_i\|^2 + \lambda \|f\|_{\mathcal{H}}^2 \tag{1.7}$$

for a regularization parameter $\lambda > 0$. The celebrated Representer Theorem (Schölkopf *et al.*, 2001) then guarantees that the solution to

the above optimization problem is a linear combination of the *feature map* of this RKHS $\mathcal{H}$ centered at the data points $x_1, \ldots, x_N$, where the feature map is derived from the kernel associated with $\mathcal{H}$. In a sense, this framework generates a set of basis functions from data, circumventing the need for an upfront selection of a basis as EDMD. In addition, RKHS offers a mathematically rigorous framework for statistical learning and enjoys both strong theoretical guarantees, such as consistency, rate of convergence, and practical flexibility in learning high-dimensional, nonlinear functions. See (Steinwart and Christmann, 2008; Berlinet and Thomas-Agnan, 2011; Muandet *et al.*, 2016) . This manuscript generalizes that theory to the question of Koopman operator learning.

When using the learned non-parametric Koopman operator as a representation of the dynamical system, the model complexity is characterized by the size of the dataset. As a result, the non-parametric representation becomes burdensome with growth in the size of the input dataset (Hou *et al.*, 2023b), and poses computational and data storage challenges. To address the scalability challenge, in Chapter 4 and 5, we review sparsification schemes that control the growth in complexity of the learned representation by selectively retaining elements in the dataset, with particular attention to coherence-based sparsification (Richard *et al.*, 2008) and kernel matching pursuit (Vincent and Bengio, 2002; Koppel *et al.*, 2019).

As mentioned before, it is natural to consider the setting in which the function space is closed under the action of the system dynamics, but this closedness assumption is restrictive and challenging to verify, e.g., see (Mezić, 2020; Colbrook *et al.*, 2024; Köhne *et al.*, 2025). In Chapter 3, we provide a simple example where a function from a given space under the action of the dynamics may not belong to that space. We then review techniques to accommodate "mis-specification" in the operator learning setting to relax the usual closedness assumption, thereby allowing the Koopman operator to map a function from an RKHS to another one outside of it.

Transfer operators for discrete-time Markovian dynamical systems are closely related to the conditional mean embedding (CME) operators in an RKHS that have been widely studied in the kernel learning literature. First presented in (Song *et al.*, 2009), the CME roughly captures

the stochastic dependency of one random variable with another correlated random variable. In a way, it encodes the conditional distribution of one random variable with another. The parallel between CME and the transfer operators becomes apparent by allowing the aforementioned random variables to correspond to states of a discrete-time Markovian dynamical system across successive time-steps. CME can be viewed as the solution to a vector-valued regression problem in a vector-valued RKHS (details in chapter 3), using which one can analyze learning rates to estimate such operators with data as in (Li *et al.*, 2022; Kostic *et al.*, 2023). We remark that the regression viewpoint for CME is new-ish to the literature; a prior definition in Muandet *et al.*, 2016 requires a closedness assumption similar to that we described for the Koopman operator. We formalize the connection between CME and the transfer operators for the mis-specified case in chapter 3, and present learning rates in both batch-data and streaming data settings in chapters 4 and 5, respectively.

Notice that Koopman operators are technically defined for an uncontrolled Markovian dynamical system, or for one that is controlled, but the control is a pre-specified function of the state. Being the gateway to embed conditional distributions (or stochastic kernels), CME paves the way to analyze Markov decision processes (MDPs) as expounded in (Grunewalder *et al.*, 2012; Hou *et al.*, 2023a). The CME framework holds promise in integrating control design for MDPs, thus generalizing the Koopman framework. In this vein, Chapter 6.3 explores the application of the Koopman–CME learning framework to model-based reinforcement learning.

## 1.4 Monograph Overview and Contributions

This monograph lays the groundwork for learning and analyzing nonlinear dynamical systems from data. Building upon statistical learning theory and the operator-theoretic framework, we aim to construct nonparametric estimates of the Koopman operator as a representation of system dynamics through its interactions with an RKHS.

In Chapter 2, we provide the reader with a brief introduction to the theory of RKHS. We start with an overview of classical concepts,

including the space of square-integrable functions, kernel functions, reproducing property, Mercer's theorem, and integral operator. We further review several theoretical results related to the powers of kernels and the construction of intermediate spaces between an RKHS and the space of square-integrable functions.

In Chapter 3, we first summarize recent developments regarding Hilbert space embedding of probability distributions, in particular, the embedding of conditional distributions. We then review approaches to study discrete-time Markovian processes through a linear-operator-theoretic lens and leverage the Koopman operator (embedded in a suitably defined RKHS) as a representation of a Markovian dynamical system. As the Koopman operator describes the dynamics of functions, choosing the right function space is critical to accurately represent system dynamics. We argue why accounting for the "mis-specified" scenario is important to learn an operator model for a dynamical system from data. Our treatment connects the Koopman operator and the CMEs, allowing us to unleash a well-developed statistical learning theory toolbox on the study of dynamical systems.

In Chapter 4, we review kernel-based approaches for learning the Koopman operator from batch data. As we have mentioned before, kernel learning becomes cumbersome with large data. To enable scaling, we discuss sparsification techniques that limit the growth of the representation. The goal of this chapter is the development of sample complexity guarantees for the Koopman operator from data under a variety of sampling paradigms. The results highlight the trade-off between model complexity (controlled by sparsification) and approximation accuracy.

In Chapter 5, building on the regression-based interpretation of CME learning, we explore the online learning or streaming data scenario where one updates the model *incrementally* as new data becomes available. We again resort to data sparsification in the interest of controlling model complexity. Unlike the batch learning setting, compounding bias might arise due to sparsification at each iteration. To this end, we discuss recent work that addresses this difficulty via adaptive step-size control. This chapter provides guarantees of asymptotic convergence and last-iterate guarantees with Markovian sampling for the Koopman operator.

In Chapter 6, we explore applications of the sparse Koopman operator learning framework. We start by illustrating the application of kernel-based operator learning to problems that arise in the domain of dynamics and stability of power systems. Given the drive to integrate renewable energy resources, distributed generation, and demand response, traditional physics-based models of system components are becoming less accurate. Our data-driven framework provides an alternate route. We present applications of sparse kernel learning in what is known as the transient stability assessment problem and the propagation of uncertainty from initial conditions. Then, we explain how our framework can be extended to handle MDPs via the bridge between Koopman operators and the CME.

Chapter 7 discusses suggestions for future research and concludes the monograph.