

# Approximation of Discrete and Orbital Koopman Operators over Subsets and Manifolds

**Andrew J. Kurdila**

Virginia Tech: Virginia Polytechnic Institute and State University

**Sai Tej Paruchuri** (✉ [saitejp@lehigh.edu](mailto:saitejp@lehigh.edu))

Lehigh University <https://orcid.org/0000-0003-2372-3888>

**Nathan Powell**

Virginia Tech: Virginia Polytechnic Institute and State University

**Jia Guo**

Georgia Tech: Georgia Institute of Technology

**Parag Bobade**

General Electric Company

**Boone Estes**

Virginia Tech: Virginia Polytechnic Institute and State University

**Haoran Wang**

Virginia Tech: Virginia Polytechnic Institute and State University

---

## Research Article

**Keywords:** Koopman Operator, Orbital Koopman Operator, Koopman Operator Convergence Rates, Reproducing Kernel Hilbert Spaces, Projection-based Koopman Approximations

**Posted Date:** August 1st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1727747/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

---

# Approximation of Discrete and Orbital Koopman Operators over Subsets and Manifolds

Parag Bobade · Boone Estes ·  
Jia Guo · Andrew J. Kurdila ·  
Sai Tej Paruchuri · Nathan Powell · Haoran Wang

Received: date / Accepted: date

**Abstract** This paper describes a general approach for the construction of approximations of the Koopman operator associated with deterministic continuous semiflows over a complete metric space  $X$ . A primary contribution of the paper is the derivation of rates of convergence for approximations of Koopman operators for large classes of evolution equations. The rates of convergence in the paper are derived in two substantially different scenarios. If the state space  $X$  is compact, we consider when the samples  $\Xi$  are dense in the entire state space  $X$ . We also study cases when samples are dense in a limiting subset  $\Omega \subset X$  that is a proper subset of  $X$ . Two general classes of methods are described in the paper, referred to as intrinsic and extrinsic methods. Intrinsic methods define bases of approximation from kernels that are defined in terms of, or having knowledge of, the limiting set  $\Omega$ . Extrinsic methods use kernels that do not depend on the knowledge of the limiting set  $\Omega$ . In both types of approximations, the regularity of the underlying set and the smoothness of the space of functions on which the Koopman operator acts determine the rate of approximation. In the strongest error bounds derived in the paper, it is shown that the error in approximation of the Koopman operator decays like  $O(h_{\Omega_n, \Omega})^p$  where  $h_{\Omega_n, \Omega}$  is the fill rate of the samples  $\Omega_n$  in the limiting set  $\Omega$  and  $p$  is an exponent related to the choice of the kernel and the smoothness of functions on which the Koopman operator acts. Such error bounds are obtained when either the limiting subset  $\Omega = X$ , when it is a proper subset  $\Omega \subset X$  that is sufficiently regular, or when it is a type of smooth manifold  $\Omega = M$ .

**Keywords** Koopman Operator · Orbital Koopman Operator · Koopman Operator Convergence Rates · Reproducing Kernel Hilbert Spaces · Projection-based Koopman Approximations

**Mathematics Subject Classification (2020)** 37 · 93

## 1 Introduction

In this paper we derive a family of error estimates for the approximation of certain kinds of Koopman operators that are defined for deterministic semidynamical systems. Koopman operators for semiflows in discrete time  $U_f$ , as well as orbital Koopman operators for continuous time semiflows  $\mathcal{U}_\phi$ , are studied. Here  $f$  and  $\phi$  are functions that govern the evolution of a discrete dynamical system and the state trajectory generated by a continuous dynamical system, respectively. For each of the two Koopman operators, two different approximations are considered: (i) projection-based approximation, (ii) data-dependent approximation. The principal theoretical results of this paper consist of estimates that make precise in what sense these approximations of the Koopman operators converge to their infinite-dimensional counterparts.

Error bounds are derived for cases when the Koopman operators act on a reproducing kernel Hilbert (RKH) space  $H_X$  that is embedded in a Sobolev space. The RKH space  $H_X$  of functions over  $X = \mathbb{R}^d$  is defined in terms of a reproducing kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$ . The Koopman operator approximations are built using

---

Parag Bobade  
GE Research, Bengaluru, India

Boone Estes, Andrew Kurdila, Nathan Powell, Haoran Wang  
Department of Mechanical Engineering, Virginia Tech, Blacksburg, USA

Jia Guo  
School of Electrical and Computer Engineering, Georgia Tech, Atlanta, USA

Sai Tej Paruchuri  
Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, USA

Corresponding Authors:  
Andrew Kurdila (kurdila@vt.edu), Sai Tej Paruchuri (saitejp@lehigh.edu)

finite-dimensional spaces of approximants  $H_{\Omega_n} = \text{span}\{\mathfrak{K}_\xi | \xi \in \Omega_n\}$  with  $\mathfrak{K}_\xi := \mathfrak{K}(\xi, \cdot)$ . Here, the term  $\Omega_n$  represents a set of samples collected along the trajectory of the dynamical system. The assumption is that the collection of samples becomes dense in a limiting set  $\Omega$  as  $n$  increases. That is, the set  $\Xi = \cup_{n \in \mathbb{N}} \Omega_n$  is dense in the limiting set  $\Omega \subseteq X$ . The theorems presented in this work illustrate that the convergence rate depends on (i) the smoothness properties of the native space  $H_X$ , (ii) the structure of  $\Omega$ . In particular, we prove that the convergence rate is directly proportional to the smoothness of  $H_X$ . In addition, convergence results are described such that we get sharper results as the knowledge about the limiting set  $\Omega$  improves. For example, error bounds are derived that are applicable to samplings collected over portions of a trajectory for the system in Example 1, Figure 1 (a), that has little discernible asymptotic structure. Stronger bounds are derived when the uncertain semiflow exhibits additional long term structure. In Examples 2 and 3 shown in Figures 1 (b) and 1 (c), the semiflow is either supported on or attracted to a smooth manifold. In this latter case, we show that the error in approximations of the Koopman operator decays at a rate that depends on the fill rate of samples in the manifold. The best results are obtained when it so happens that the limiting set  $\Omega$  is a smooth manifold, and in this case it is shown that the finite-dimensional error in approximating the Koopman operator decreases with  $O(h_{\Omega_n, \Omega}^p)$ . Here  $h_{\Omega_n, \Omega}$  is the fill distance of the samples in the limiting set  $\Omega$ , and  $p$  depends on the smoothness properties of the native space  $H_X$ .

We introduce the motivation behind Koopman theory in Section 1.1. We review the goals and philosophy of the Koopman operator in Section 1.2. Related work and open questions are reviewed in Section 1.3. The strategy of the overall paper is summarized in Section 1.4. We give an overview of the new results derived in the paper in Section 1.5. The theory developed in the paper begins with a discussion of background material on RKH spaces, Sobolev spaces, and interpolation or projection operators in Section 2. Essential material on dynamical systems theory is contained in Section 3. Sections 4 and 5 give detailed accounts of the development of the newly introduced theory and the main technical results. Numerical studies and conclusions are presented in Sections 6 and 7, respectively.

## 1.1 Motivation

Koopman theory is a body of research that uses operator theory to study dynamical systems. Owing to the amenability to data-driven methodologies that emerged from Koopman theory, it has gathered attention from the researchers across disciplines in the past decade [1–7]. While the original paper by Bernard O. Koopman [8] appeared early in the 1930s, a number of recent texts are evidence of the renewed interest in Koopman operator theoretic approaches. General studies of the Koopman operator can be found in texts on ergodic theory such as [9], and in the popular text [10] on deterministic or stochastic dynamical systems. Just over the past five years careful studies such as [11–33] have explored refinements of the theory of approximation of Koopman operators as well as associated data-driven modeling methods for uncertain dynamical systems. These references contain many individual contributions to the theory that underlies the approximation of Koopman operators. In view of the breadth of these studies on Koopman theory, a full and detailed account of the finer points studied in these papers would far exceed the limit of a single paper. Here we only review some of the broader issues and emphasize what we consider to be important open questions that remain, as of yet, largely unresolved. For the sake of researchers across diverse fields, we briefly review Koopman theory and its philosophy next so that this paper is self-contained. Such a review has been carried out in greater length and detail in a number of places, and we recommend discussions such as in [34–37] for more extensive treatments.

## 1.2 The Philosophy of Koopman Theory

One of the principal reasons for the popularity of methods derived from Koopman theory is that they are amenable to data-driven approaches. In general, these studies build approximations of quantities or mathematical objects associated with an unknown flow from samples or observations. The approximations can take the form of estimates or predictors of the state, estimates of an observable function, or approximations of the propagation law of the dynamical system itself, among other examples. Moreover, it is not only the efficacy of data-driven algorithms that has resulted in such focus on Koopman theory. The theory is generally applicable to, indeed in a sense expressly designed for, the study of nonlinear systems. Koopman theory provides an elegant framework in which to carry out analysis of uncertain nonlinear dynamics as well as to develop data-driven algorithms for modeling and identification of such systems.

Because the general theory is so broadly applicable, a diverse set of definitions of a Koopman operator have appeared in the literature. These definitions are tailored to a class of uncertain dynamical systems under study. Variants are defined for deterministic and stochastic systems, and for both discrete time and continuous time evolutions. A good account of the variety of definitions can be found in [10], and a detailed discussion of the duality structures that accompany some of these representations can be found in [38]. Intuition about the nature

of Koopman theory, its goals and strategies, is perhaps best achieved by considering a simple model of a discrete time, deterministic, nonlinear system. Such a typical system is governed by the equations

$$\begin{aligned}\phi_{n+1} &= f(\phi_n) \\ y_{n+1} &= h(\phi_{n+1})\end{aligned}\tag{1}$$

where  $\phi_n$  is the state at time  $t_n \in \mathbb{R}^+ := [0, \infty)$  in the state space  $X := \mathbb{R}^d$ ,  $f : X \rightarrow X$  is an unknown, generally nonlinear function that determines the dynamics, the function  $h : X \rightarrow \mathbb{R}^m$  is an observable or output function, and  $y = h(x)$  is a measurement or observation evaluated at the state  $x \in X$ . Because it is normally the case that the function  $f$  is unknown in applications of Koopman theory, we say that this is an example of an uncertain system. In Koopman theory, it is ordinarily assumed that the history of input-output observations  $\{(\phi_i, y_i)\}_{i \in \mathbb{N}_0}$  has been collected, and it is desired to use this information to estimate properties of, or some mathematical object associated with, the flow.

For this type of dynamical system, the Koopman operator  $U_f$  is defined as a mapping that takes functions  $g : X \rightarrow \mathbb{R}$  defined on (perhaps some subset of) the state space  $X$  to another function  $U_f g$  via the rule

$$(U_f g)(x) := (g \circ f)(x)\tag{2}$$

for all  $x \in X$ . In this definition note that the operator  $U_f$  is characterized by the unknown function  $f$  that appears in the evolution law: it is fixed in the definition above. It is now possible to define another discrete evolution law, one that defines the evolution of functions that satisfy

$$g_{n+1} = U_f g_n,\tag{3}$$

which can be used to investigate various properties of the original system in Equation 1. This evolution law determines a discrete, linear, generally infinite dimensional dynamical system, in contrast to the original system that is discrete, nonlinear, and finite dimensional. In the language of Koopman theory, Equation 3 is said to induce discrete linear dynamics in the ‘‘space of observables,’’ or observable functions.

We can fill in a bit more details, at least for one iconic problem, to explain how the study of Equation 3 might inform the study of the original system in Equation 1. Suppose we set the number of outputs  $m = d$ , define the output function  $h := (h_1, \dots, h_d)^T$ , and introduce the vector-valued Koopman operator  $\mathbb{U}_f h := \{U_f h_1, \dots, U_f h_d\}^T$ . If we select the observable function to correspond to the full-state observation function  $h(x) = x = \{x_1, \dots, x_d\}^T$ , then we have

$$\phi_{n+1} = \underbrace{(f \circ f \circ \dots \circ f)}_{n \text{ times}}(\phi_0) = \underbrace{(\mathbb{U}_f \circ \mathbb{U}_f \circ \dots \circ \mathbb{U}_f h)}_{n \text{ times}}(\phi_0).$$

We see from this identity that if we are able to build approximations of the Koopman operator  $U_f$ , it in principle enables the development of state estimators or predictors for the original nonlinear system. A similar strategy can be followed to study forecasting of the output  $y_{n+1}$ . Since we have  $y_{n+1} = h(x_{n+1}) = h(f(x_n)) = (U_f h)(x_n)$ , we see that approximations of the Koopman operator can be used to forecast the future output  $y_{n+1}$  from the present state  $x_n$ .

This simple model problem illustrates clearly a tradeoff that arises in Koopman theory that largely motivates its use. The dynamics of nonlinear systems can be much more difficult a topic of study than that for linear systems. Koopman theory replaces the study of the original nonlinear system with that of a linear system. Still, as attractive as it is to study a linear system instead of a nonlinear system, there is at least one significant drawback here. The original system above is finite dimensional while the latter one induced by the Koopman operator  $U_f$  is generally infinite dimensional. In fact, the Koopman operator  $U_f$  that defines the dynamics in the space of observables depends on the unknown function  $f$ , and it therefore is also unknown when the dynamics are uncertain. Considerations of how to build approximations of the Koopman operator in finite dimensional spaces necessarily plays an important role in Koopman theory, and these associated technical questions can be delicate and nuanced. It is the nature of the approximations of the Koopman operator that is the concern of the recent papers summarized above.

Any approximation of the Koopman operator necessarily entails a choice of the finite dimensional spaces of approximants, and the types of approximants that have been studied for this problem is vast. Common choices of bases include Fourier modes, globally supported polynomials, piecewise polynomials in the form of finite element functions or splines, wavelets, bases defined in terms of reproducing kernels of native spaces, as well as eigenfunctions of Koopman operators or approximations of eigenfunctions of Koopman operators. A discussion of the approximation spaces that can be defined in terms of some of these choices above, and how they influence convergence rates, is given by the authors in [38].

However, even in view of the diverse choices summarized above, perhaps it is most common to construct approximations of Koopman operators in terms of eigenfunctions of the Koopman operator, or in terms of approximations of eigenfunctions. The eigenfunctions, which are simply known as Koopman modes, are solutions

of the (functional) eigenvalue problem for the Koopman operator. Just as the eigenvectors of the coefficient matrix of a linear, time-invariant system can be used to decouple, diagonalize, and simplify the study of that system, *for some* types of nonlinear systems a similar strategy in terms of the Koopman modes is possible. In essence, the Koopman modes sometimes can be used to transform the nonlinear system rigorously into a linear system. With this ideal scenario in mind, a common strategy for a nonlinear problem at hand is to generate a reduced order, finite dimensional model by the superposition of a finite number of Koopman modes. When the calculation of the Koopman modes can be problematic either analytically or computationally, alternatives rely on computing approximations of the eigenfunctions. The study to establish for what systems this transformation can be accomplished exactly or approximately with sufficient accuracy, and the assessment of the magnitude of any resultant errors is a very active area of research. The studies [11–13, 16, 22–24, 26, 27, 31, 38, 39] all investigate aspects of how to go about generating approximations of Koopman modes, which has become a topic of interest in and of itself.

### 1.3 Broader Issues and Open Questions

As suggested in the above discussion, the fine details of the theory advanced in [11–20, 22–33, 40] can vary a lot from reference to reference. Here we identify a few of these recent publications that, in the authors' view, appear most relevant to this paper.

#### 1.3.1 Most Relevant Recent Work

The early work by [31] studies the relationship of Koopman approximations to matrix representations that appear in the dynamic mode decomposition (DMD) and Extended DMD (EDMD) algorithms, and uses the algorithms to estimate the Koopman modes. Reference [39] analyzes approximations of the Koopman and Perron-Frobenius operators, their relationship to the EDMD algorithm, and interprets them as different types of Galerkin approximations. A central conclusion of [39] is that as the number of samples  $m$  increases, a finite dimensional approximation of the Koopman operator converges over a fixed finite dimensional subspace of approximants. That is, if  $n$  is the dimension of the space of approximants or of the basis, and  $m$  is the number of samples, this paper gives sufficient conditions to ensure that when  $n$  is fixed, convergence is guaranteed as  $m \rightarrow \infty$ . Examples are studied in which the bases of the approximant spaces are selected to be globally defined monomials. Reference [21] builds upon the study in [39] and further explores that approximation of transfer operators, and in particular the Koopman and Perron-Frobenius operators. The relationships among various approximations, the DMD method, the EDMD methods, the time-lagged independent component analysis (TICA) method, and the variational approach to conformation dynamics (VAC) are discussed.

The recent reference [41] further studies the convergence of approximations of Koopman operators in terms of projections on the  $L^2_\mu(X)$  space with  $X$  the state space of the semiflow. Several conclusions regarding convergence as the dimension  $n$  of the space of approximants or the number  $m$  of samples approaches infinity are made, but again, no discussion about rates of approximation in terms of fixed numbers  $m$  of samples or fixed dimensions  $n$  of the approximant space is given.

Other references that are particularly relevant to this study include the quite recent efforts in [12, 42–44]. These papers explore methods for framing the approximation of Koopman operators in terms of bases built from kernels of RKH spaces. The paper [42] studies the approximation of the Koopman and Perron-Frobenius operators and concentrates on the realization of approximations in terms of kernels of RKH spaces. In contrast to many of the articles on the approximation of Koopman operators, reference [42] discusses probabilistic rates of convergence of various approximations introduced for the regularized, stochastic Koopman operators. Specifically, it is shown in Theorem 3.14 that, for systems in which the samples are independent and identically distributed (IID), approximations converge in probabilistic order  $\mathcal{O}_P(m^{-1/2}\epsilon^{-1})$  with  $m$  the number of samples and  $\epsilon$  the regularization parameter used to define the approximation to the Koopman operator. (It should be noted that this estimate is derived by choosing  $m = n$ , much as we do later in our data-driven approximations of the Koopman operator.) This bound is derived for IID processes, which limits its applicability to cases when the samples are generated along the sample path of many nonlinear stochastic recursions. This result should be contrasted with Section 8 of [38] that is based on concentration of measure inequalities for certain types of dependent or independent processes; the latter yields error estimates for approximations of stochastic Koopman operators that are explicit in the number of samples  $m$  and the dimension of the space of approximants  $n$ .

The approach in [12] uses diffusion eigenfunctions that are intrinsic to the manifold  $M$  over which evolutions take place. In particular, the eigenfunctions of the Laplace-Beltrami operator over  $M$  are employed as the basis for theoretical considerations. In this sense the method can be understood as an intrinsic method, which we talk about in further detail later in this paper. Since the construction of the intrinsic basis over the manifold is impossible when the manifold is unknown, data-driven approximations of the unknown intrinsic kernel are computed using a variable bandwidth kernel approximation defined over  $\mathbb{R}^d$ . In some sense, the results described

in the current paper can be viewed as a method to obtain explicit bounds on the error of approximation in a similar setup. Specifically, in this paper we make precise when we can restrict a kernel defined over the state space  $\mathbb{R}^d$  to a particular limiting subset  $\Omega$  (which may be a manifold) and derive associated rates of convergence. By employing trace theorems for Sobolev spaces that are regular enough to be RKH spaces, it is possible to make explicit predictions of rates of convergence in terms of the fill distance, a topic not pursued in [12].

The authors of [43] study approximations of Koopman operators for some unitary groups, again constructing approximations in terms of bases for RKH spaces over a smooth manifold  $M$ . An important contribution of this paper is the choice of the RKH framework that induces dynamics described by the unitary group  $e^{tW_\tau}$  for a compact, self-adjoint operator  $W_\tau$  that is amenable to approximation. Overall, the analysis in the paper investigates the convergence of spectral approximations of Koopman operators, and it explores the convergence of their associated kernel integral operators over  $L_\mu^2(X)$ . Typical results (see for example, Theorem 21) ensure the convergence of approximate eigenfrequencies, approximate eigenfunctions, and observables as the dimension  $n$  of the subspace of approximants increases to infinity. Again, to emphasize the contrast to results in this paper, reference [43] does not study the rate of convergence, nor error in say  $L_\mu^2(M)$ , of approximations of the Koopman operator that is induced by a particular choice of a kernel of the RKH space. Different choices of kernels will induce different rates of convergence, and it would be a great advantage to understand how to control or influence the error by virtue of selecting a different basis.

References [45] and Sections 8.3 through 8.5 of [38] are noteworthy to the current paper because they emphasize the relevance of a large body of research on statistical learning theory, approximation theory, and nonlinear regression that is pertinent to Koopman theory. As shown in [38], the conventional result whereby the EDMD algorithm is used to generate a data-driven approximation of the Koopman operator can be identified, in some situations, with the estimates that result from the method of empirical risk minimization in statistical learning theory. In fact, this equivalence can be used to derive rates of convergence for EDMD-based approximations of the Koopman operator that improve some of the convergence guarantees that have been derived in recent literature. A similar strategy is explored in [45], although the convergence results stated are not as sharp. Reference [45] casts the problem of approximating the Koopman operator as a classical empirical risk minimization problem, chooses a basis in terms of a reproducing kernel, and subsequently shows in Theorem 1 that approximations converge in the Lebesgue space  $L_\mu^2(X)$  with  $X$  the state space of the system. As outlined in Section 3 of [45], rates of convergence follow in principle by making assumptions about the rate of decay of the eigenvalues that describe the reproducing kernel. See also [38] how this strategy is equivalent to the definition of linear, spectral approximation spaces that characterize the rates of convergence.

The authors in [38], as noted above, discuss rates of convergence for approximations of Koopman operators in terms of spectral approximation spaces. These spaces are defined via the eigenfunctions of integral operators that are defined in terms of the kernel of an RKH space. As suggested above, this is a well-known technique to study the approximation rates of functions. It is widely used *when the eigenfunction basis* is known, or at least can be readily approximated. However, it cannot be emphasized enough that computing a basis that is intrinsic to some complex underlying set  $\Omega$ , or even a smooth manifold  $M$ , can be extraordinarily difficult. The authors in [38] discuss rates of convergence for some common choices of bases including normalized B-splines, piecewise polynomials, orthonormal wavelets, and orthonormal multiwavelets. But the analysis there is carried out assuming that the bases used for approximations are “well-aligned” with the limiting set  $\Omega = X$ . The definition of such nice bases may be possible in some highly structured cases, such as flows on a circle, a sphere, or on the torus. But it may be far from trivial to come up with such well-designed, domain-adapted bases in typical applications of Koopman theory since the domain over which samples are collected is unknown, has a complicated structure, or both.

### 1.3.2 Open Questions

Considered as an entire class, there are some important subproblems in the approximation of Koopman operators that are not covered in a systematic way in the literature so far. We can summarize some of these open questions by being a bit more precise about the structure of the problem that arises when trying to approximate a Koopman operator. The general setup in this paper is described in terms of the following key features that are used to frame the approximation problem. We are given a continuous semiflow in either discrete or continuous time over the state space  $X$ . In some situations, the state space may be compact but in many cases it is not. A set of samples  $\Xi := \bigcup_{n \in \mathbb{N}} \Omega_n$  are collected that are dense in a “limiting subset”  $\Omega \subseteq X$ . Here  $\Omega_n$  is a finite set that contains  $n$  samples that is used to build a particular finite dimensional approximation of the Koopman operator. The limiting set  $\Omega$  may, or may not be, the entire state space  $X$ . A finite dimensional space  $H_{\Omega_n}$  of approximants is defined in terms of a basis  $\{\psi_i\}_{i=1, \dots, n}$ , and this basis is used to fashion an approximation of the Koopman operator. The basis may be supported on  $\Omega \subset X$ , or it may be supported on all of  $X$ . This means that we can discuss approximation of Koopman operators on functions supported on the entire state space, or we can talk about the approximation of Koopman operators over functions supported on a proper subset  $\Omega \subset X$ .

We can gain an appreciation of some of the open problems that motivate this paper by considering a few of the most common scenarios. Perhaps most frequently, it is simply assumed that the samples  $\Xi$  are dense ( $\Xi \subset X$  and  $X \subset \overline{\Xi}$ ) in the compact state space  $X$  over which the evolution is defined. In this event, we have the special case in which the limiting set  $\Omega = X$  is known *a priori*. By definition, the basis functions could be any of a large number of possible choices for functions over  $\Omega = X$ , including splines, polynomials, Fourier modes, wavelets, or kernel functions. It is frequently the case in the papers above that the basis  $\{\psi_i\}_{i=1}^n$  is a collection of eigenfunctions of the Koopman operator, which are then necessarily functions over the set  $X$ . It is also commonly the case that the basis functions are in fact other types of eigenfunctions, associated with different differential operators or semigroups, like the heat kernel over  $X$ . For any of these choices of bases, in the highly idealized scenarios when samples fill the state space  $X$ , it is still seldom the case that the rates of convergence of the approximation of Koopman operators are related to the specific choice of the bases  $\{\psi_i\}_{i=1, \dots, n}$  defined over the set  $X$ . It is much more commonly the case that the approximations are shown to converge in terms of  $L^2_\mu(X)$ , for instance, but the rates of convergence are not derived based on the choice of bases  $\{\psi_i\}_{i=1, \dots, n}$  that are contained in  $L^2_\mu(X)$ . This situation for the approximation of Koopman operators should be contrasted to studies of rates of convergence for approximations of solutions of partial differential equations (PDEs). A few decades of study have developed a rich theory that relates approximation theory, the definition of (finite element or finite volume or spline or wavelet or kernel or eigenfunction) bases, and the rates of convergence of approximations of solutions of PDEs. To be sure, some of the standard machinery for building approximations of PDEs by refinements or enrichments over fixed, bounded domains are either unsuitable or would be difficult to apply to approximations in Koopman theory. Even if the state space  $\Omega = X$  is known *a priori*, the distribution of samples over  $X$  is generally not known, and it is the distribution  $\mu$  in  $L^2_\mu(X)$  that intuitively would guide enrichment, refinement, or other forms of basis selection over  $X$ . The difficulty of course is that the portions of the domain over which samples are collected or concentrated is inextricably connected to the dynamical system under study. Moreover, sometimes approximations are sought to induce a dynamics that is somehow consistent with the original semiflow, and this topic does not seem to have a precedent in most methods for approximating solutions of PDEs. All of these factors combine to make it more challenging to derive rates of convergence of Koopman approximations. Still, there is a notable lack of results on rates of convergence of Koopman operator approximations, even in this simplest scenario when samples are dense in  $X$ . While these results on convergence may be harder to come by, or to prove, there definitely is a need for further study along these lines.

Perhaps more importantly, when the limiting set  $\Omega \neq X$ , but rather it is a proper subset of the state space  $X$ , the problem of proving convergence of approximations is still more subtle. From a practical point of view, this is an important issue. The driving motivation behind the use of Koopman theory in many real applications is that it is known, or it is believed, that the high dimensional dynamics on the state space  $X$  actually corresponds to a dynamics on a simpler or smaller subset in a particular motion regime or case under study. Perhaps the trajectory at hand actually evolves over a “smaller” invariant set, or perhaps the trajectories over time accumulate near such a set. The goal of Koopman theory in such a case is to provide a principled, data-driven method to build a model having reduced complexity. It would certainly be valuable to characterize the rates of convergence in this situation, but part of the problem is choosing bases and associated spaces of approximants that provide a tractable setting for analysis over the unknown set  $\Omega$ . This is complicated by the fact that the limiting set  $\Omega$  may have no apparent or particularly notable structure, or it can have a high degree of regularity when measured in a certain sense. As noted on page 341 of [46], the case when the limiting subset is a proper subset of the state space is a common and difficult case that can pose its own set of difficulties. The limiting set  $\Omega$  may have zero measure as a subset of  $X$ , and choosing function spaces over  $\Omega$  that are well-defined can be a difficult task in its own right. In any event, when the limiting set  $\Omega$  is an unknown proper subset of  $X$ , the approximation problem is more complex. We can define the basis used for approximations over the limiting subset  $\Omega$ , or we can define the basis for approximation over  $X$ . In both cases we want to understand how the choice of bases affects the convergence rate, what are appropriate function spaces in which to estimate errors, and how the approximation results over  $\Omega$  or over  $X$  might be related.

Before proceeding to a summary of the new results of this paper, we would like to emphasize one final source of complication when putting together a general approach to the approximation of Koopman operators. A data-driven approximation of the Koopman operator will always be a function of the number of samples  $m$  of the observations and the dimension  $n$  of the space of approximants. For Koopman operators associated with *stochastic flows*, approximations of the Koopman operator will typically consist of a bias and variance term: this is a fundamental decomposition of the error when samples are generated according to a stochastic process. The bias term can be traced to the error that is induced by the specific choice of the space of approximants, and it is also known as the approximation space error. The variance term is associated with the stochastic nature of the system. A simple overview of the breakdown of the error in approximating the Koopman operator in terms of the bias and variance is given in [38]. The approximation space or bias error decreases as the dimension  $n$  of the space of approximants is increased. The stochastic error is typically a complicated function of the dimension  $n$  and the number of samples  $m$ . However, for a fixed number of samples  $m$ , the variance error term increases as  $n \rightarrow \infty$ . The important issue to be remembered regarding this decomposition is that the bias-variance tradeoff

implies that the dimension of the approximant space  $n$  must be adapted as a function of the number of samples  $m$  to achieve an optimal error reduction.

This paper considers only deterministic systems, which means that some of these finer issues regarding the balance of probabilistic and deterministic errors do not arise.

#### 1.4 Strategy of this Paper

In short, the purpose of this paper is to provide a general and systematic framework for the construction of approximations of Koopman operators for deterministic systems, one that can be applied either when the samples are dense in the entire state space  $\Omega = X$ , or when samples are dense in a limiting subset  $\Omega \subset X$  that is properly contained in the state space. The goal of the theory is to generate approximations for which the rates of convergence can be studied for a wide class of Koopman operators and bases  $\{\psi_i\}_{i=1,\dots,n}$ . We want to explore how the rates depend on the choice of function spaces, the limiting set  $\Omega$ , and the choice bases that define the space of approximants.

The analysis in this paper differs in two essential ways from many approaches to the approximation of Koopman operators that have appeared over the past few years. First, the study makes a careful use of *priors* to establish *rates of convergence* that depend on the smoothness and approximation properties of the observables on which the Koopman operator acts. Priors in this paper refer to membership in certain native spaces of reproducing kernels, or in some cases to Sobolev spaces that measure a scale of smoothness  $s$ . The parameter  $s$  can likewise be interpreted as describing the approximation properties of the functions in the Sobolev space. This is a standard framework in which other types of general (nonparametric) estimation problems are cast, such as in the study of (distribution-free) learning theory as it is applied to nonlinear regression [47–51]. In this paper we demonstrate the strength of such an approach when it is applied to Koopman theory also. A pragmatic implication of this framework is that we are able to derive strong error bounds in a number of different, but common, scenarios that arise in the approximation of Koopman operators. Overall, we can view the current paper as a continuation of the study and approach in [38]. That reference studies rates of convergence for some common choices of basis functions like splines, orthonormal wavelets, and orthogonal eigenfunctions of integral operators that are often associated with the kernels of RKH spaces. The results in [38] apply when the bases are “well-aligned” with the domain  $\Omega = X$ . Here, in contrast, the emphasis is on bases built from the translates of the kernel functions of a RKH space, and much of the theoretical background is explained by methods of scattered data approximation. We show that the choice of bases in terms of scattered data samples is a powerful tool for the derivations of rates of approximation of Koopman operators in that the conditions that bases be adapted to the domain are not required.

The second distinguishing feature of the approach in this paper is the investigation of the way in which the *structure of the underlying set* on which the dynamics is supported or accumulates plays a critical role in the analysis of Koopman operator approximations. The authors feel that in fact it is this dependence on the set that supports the underlying dynamics that fundamentally differentiates approximation in Koopman theory from other fields of study in estimation and approximation theory like nonlinear regression. We will see that ultimately the interplay between the dynamics and rate of convergence of approximations of the Koopman operator is described by the rate at which the samples fill the limiting set. To the authors’ knowledge, this is a qualitatively new result in the analysis of approximations of Koopman operators.

Consideration of a few examples can illustrate the breadth of behavior that the authors feel that approximation of the Koopman operator should take into account. The examples below are meant to emphasize an important feature about data-driven models of dynamical systems: they are often used to capture the “underlying physics” of some high-dimensional flow. In rough terms this means that although the state space may have many degrees of freedom, there actually is a low dimensional mathematical structure lurking in the larger space. Although, the examples are only two dimensional, they each illustrate how the limiting behavior of the flow can exhibit quite a bit of structure. They are also suggestive of how error bounds for approximation of Koopman operators reflect or depend intrinsically on that underlying structure. While discrete semiflows with the form of Equation 1 are more common in the existing studies about Koopman operators, we will also treat orbital Koopman operators, defined in Section 3.1, that are associated with semiflows in continuous time such as the following two examples. Of course, discretization in time of the continuous time dynamical systems in the following examples could also define discrete time flows of the type in Equations 1.

*Example 1 (Example 2.1, [52])* The following system of equations represent the model of a tunnel-diode circuit. Figure 1, (a) depicts trajectories for several initial conditions of the flow,

$$\begin{aligned} \dot{\phi}(t) &= \begin{Bmatrix} \frac{1}{C}(-h(\phi_1(t)) + \phi_2(t)) \\ \frac{1}{L}(-\phi_1(t) - R\phi_2(t) + u) \end{Bmatrix}, \\ \phi(0) &= \phi_0 \in \mathbb{R}^2, \end{aligned}$$



with  $u = 1.2$ ,  $R = 1.5 \times 10^3$ ,  $C = 2 \times 10^{-12}$ ,  $L = 5 \times 10^{-6}$ , and  $h(\phi_1) := 17.76\phi_1 - 103.79\phi_1^2 - 229.62\phi_1^3 - 226.31\phi_1^4 + 83.2\phi_1^5$ . This example has been selected since the phase portrait is highly complex and does not exhibit any recognizable “simple underlying structure” that will feature prominently in the next few examples. Samples  $\Xi(\phi_0)$  and  $\Xi(\psi_0)$  that are collected along two distinct trajectories are shown in the figure. The positive limit sets consist of a finite collection of isolated points in this example, and the qualitative asymptotic behavior of a trajectory can vary drastically depending on the location of the initial condition. We are interested in obtaining precise descriptions of convergence of Koopman operator approximations built from samples such as these two sets, which we intuitively expect to be quite different finite dimensional operators. In addition, we are interested in understanding over what spaces of functions the Koopman approximations converge, how these spaces compare, what convergence results can be stated if the two sets are merged, and so forth. In the current example, the emphasis would be on identifying useful function spaces that are well-defined for the problem of estimating error in this “unstructured” scenario.

*Example 2* Figure 1, (b) depicts trajectories for several initial conditions of the flow that is governed by the Van Der Pol oscillator equations

$$\begin{aligned} \dot{\phi}(t) &= \left\{ \begin{array}{l} \phi_2(t) \\ -\phi_1(t) + \epsilon(1 - \phi_1^2(t))\phi_2(t) \end{array} \right\}, \\ \phi(0) &= \phi_0 \in \mathbb{R}^2. \end{aligned}$$

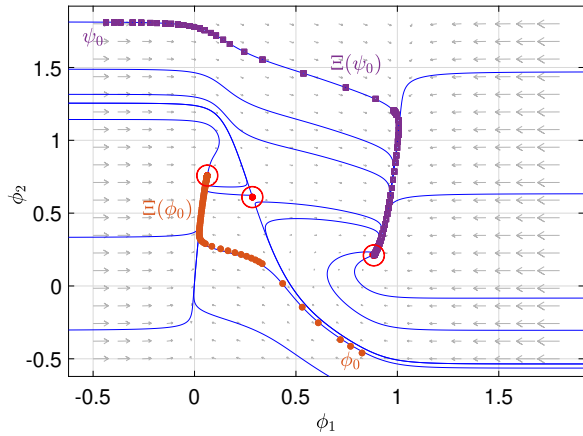
This semiflow exhibits behavior that is qualitatively quite different from that in Example 1 above. There is a unique limit cycle for this flow, and for any  $\phi_0$  on the limit cycle, the orbit  $\Gamma^+(\phi_0)$  is a smooth, compact manifold  $M \subset \mathbb{R}^2$ . However, for other initial conditions  $\phi_0$ , the orbit  $\Gamma^+(\phi_0)$  is not a smooth compact manifold. It is easy to see from the phase portrait that the limit cycle  $M$  attracts the trajectory generated by any initial condition as  $t \rightarrow \infty$ . If we have two different initial conditions  $\phi_0$  and  $\psi_0$  and generate two different samplings  $\Xi(\phi_0)$  and  $\Xi(\psi_0)$ , then the limiting sets  $\overline{\Xi(\phi_0)}$  and  $\overline{\Xi(\psi_0)}$  are different sets. It is not hard to imagine that finite dimensional approximations, especially those constructed from samples during the “transient regime,” could be substantially different. However, both sets  $\overline{\Xi(\phi_0)}$  and  $\overline{\Xi(\psi_0)}$  contain the same positive limit set  $M$ . This paper makes precise in what sense the approximations of the Koopman operator generated from  $\Xi(\phi_0)$  and  $\Xi(\psi_0)$  are comparable: they converge in some sense to the same operator restricted to the RKH space  $H_M$  that is indexed by the manifold  $M$ . Furthermore, if the RKH space  $H_M$  happens to be a Sobolev space, we derive sufficient conditions to guarantee rates of convergence of approximations of the Koopman operator. In this paper we will describe methods in which approximations of the Koopman operator are derived in terms of *intrinsic* kernels that depend on knowledge of the manifold  $M$  and *extrinsic* kernels that do not require knowledge of the underlying manifold  $M$ .

*Example 3* In Figure 1, (c), the trajectories of the Lotka-Volterra predator-prey model given by the system

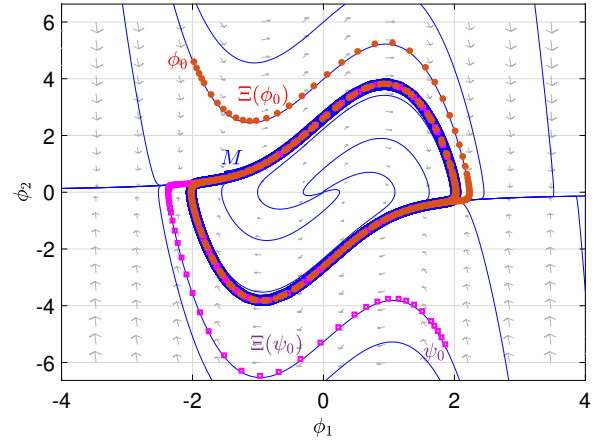
$$\begin{aligned} \dot{\phi}(t) &= \left\{ \begin{array}{l} \phi_1(t)(\phi_2(t) - 2) \\ \phi_2(t)(1 - \phi_1(t)) \end{array} \right\}, \\ \phi(0) &= \phi_0 \in \mathbb{R}^2 \end{aligned}$$

are depicted. The phase portrait illustrates the fact that this system exhibits a great deal of structure, in some sense perhaps more than the other examples considered thus far. For infinite number of choices of initial conditions  $\phi_0$ , the orbit  $\Gamma^+(\phi_0) := \cup_{t \geq 0} \phi(t)$  is a smooth, compact, 1-dimensional Riemannian submanifold  $M \subseteq \mathbb{R}^2$ . Several samples are shown along one trajectory, where the orange points denoted with a ‘.’ are collected during the first transit around the limit cycle, then the samples denoted by purple squares are added during the second transit around the limit cycle, and so forth. The collection of samples  $\Xi := \Xi(\phi_0) := \{\phi_i\}_{i \in \mathbb{N}_0} := \{\phi(t_i)\}_{i \in \mathbb{N}_0}$  can be built progressively as  $t_i \rightarrow \infty$  so that  $M := \overline{\Xi}$ . Basic questions arise that can be traced to the nature of the set  $\Omega$ . For instance, if we build an approximation  $\mathcal{U}_\phi^n$  of the Koopman operator  $\mathcal{U}_\phi$  in terms of the sampling  $\Xi(\phi_0)$ , for what functions  $g$  is  $\mathcal{U}_\phi^n g$  a good approximation of  $\mathcal{U}_\phi g$ ? We can state this another way: over what subspace of functions is  $\mathcal{U}_\phi^n$  a good approximation to  $\mathcal{U}_\phi$ ? This paper will show that since  $M$  is a smooth manifold, we can get precise answers to this question for some approximations of the Koopman operator.

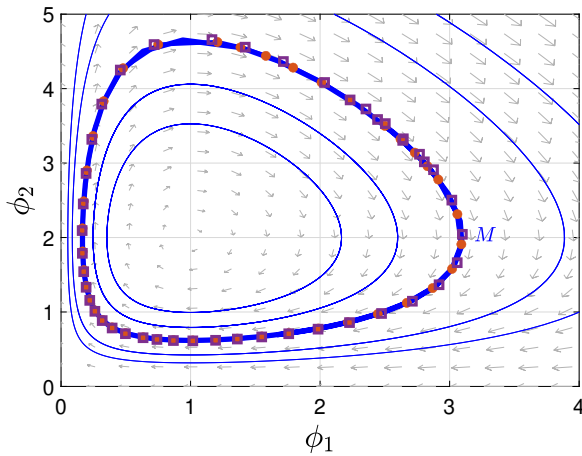
*Example 4* Here we briefly overview the geometric method of numerical integration [53] that is often used to generate simulations that preserve certain qualitative behaviors of systems. We assume the discrete evolution used for numerical integration is an approximation of the underlying continuous one. Given state variables  $\phi \in M$ , where  $M$  is a  $d$ -dimensional differentiable manifold, the tangent space to  $M$  at a point  $\phi \in M$  is denoted by  $T_\phi M$ . For a curve  $t \mapsto \phi(t)$  on the manifold  $M$ , the velocity  $\dot{\phi}(t) \in T_{\phi(t)} M$  at a time  $t$  is tangent to the manifold  $M$ . The tangent bundle  $TM$  is the union of all of the tangent spaces to  $M$ , which serves as



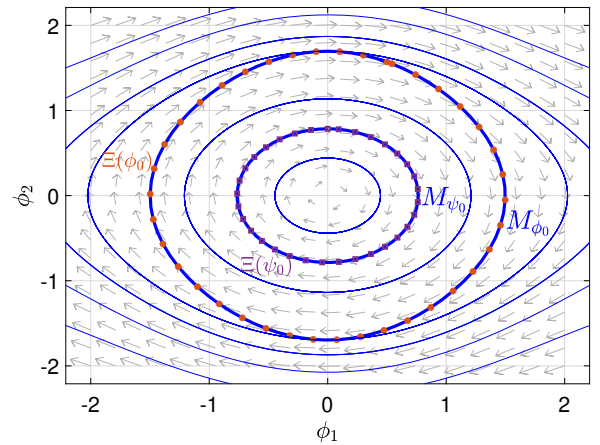
(a) A flow having little apparent underlying structure. The positive limit set is a collection of attractive equilibria circled in red. Here samples are taken along a few trajectories during transient regimes.



(b) A flow having “asymptotic structure,” where trajectories accumulate on the positive limit set, which is an invariant set of the flow.



(c) A flow exhibiting considerably more structure. Each orbit generated by an initial condition in the upper right quadrant is an embedded submanifold of  $\mathbb{R}^2$ .



(d) Samples taken from a geometric numerical integrator scheme. The integration scheme determines discrete flows of orbits defined by 1-dimensional manifolds. The underlying manifolds are completely determined by the initial conditions. For certain orbits that do not lie on the same manifold, the estimates will never convergence to one another as  $t \rightarrow \infty$ .

Fig. 1: A collection of flows exhibiting various kinds of underlying structure. (a) Phase portrait of Example 1, from [52] page 46, Example 2.1, (b) Phase portrait of Example 2, the Van Der Pol oscillator, (c) Phase portrait of Example 3, the Lotka-Volterra predator-prey equations, [53] page 4, and (d) Phase portrait of simple pendulum dynamics and samples that are taken from a discrete numerical integrator with an invariant flow along a manifold defined by the dynamics.

the state space of  $(\phi(t), \dot{\phi}(t))$ . In analogy to the ODEs in Examples 1, 2 and 3, continuous evolution law on the manifold  $M$  is governed by the equation

$$\dot{\phi}(t) = F(\phi(t)),$$

where the map  $F : M \rightarrow TM$ . It defines a semigroup  $\{S(t)\}_{t \geq 0}$  on the manifold  $M$  so that the solutions are given by  $\phi(t) = S(t)\phi_0$ . In this case, we can discretize the semiflow and define the familiar discrete nonlinear recursion,  $\phi_{i+1} = f(\phi_i)$  where  $f(\phi_i) = S(h)\phi_i$  with a fixed timestep  $h$ . The theoretical results of this paper, for instance, can be applied when the system eventually approaches a limit set  $\Omega$  which is itself a manifold.

In practice, a closed form expression for the semigroup  $S(t)$  is not available for many systems of interest. Geometric integration methods [53] are popular methods for approximating the system on the manifold  $M$  by a discrete evolution law. For example, we consider the evolution corresponding to the dynamics of a pendulum with a mass  $m = 1$  hanging from a massless rod of length  $l = 1$  that is approximated using the Störmer-Verlet discrete integration scheme as given in Example 1.4 in [53]. In this example,  $\phi_1$  and  $\phi_2$  correspond to the

pendulum's momentum and position, respectively. The equations of motion are given as

$$\begin{aligned}\dot{\phi}_1 &= -\sin(\phi_2) = f_1(\phi_2), \\ \dot{\phi}_2 &= \phi_1.\end{aligned}$$

The discrete evolution at the  $i^{\text{th}}$  step is determined from the pendulum's position  $\phi_{2,i}$ , as well as approximations  $\phi_{1,i}$  and  $\phi_{1,i+1/2}$  of the derivative  $\dot{\phi}_1 = \dot{\phi}_{(2)}$  at the  $i^{\text{th}}$  and  $i+1/2$  step. The approximations are calculated by introducing the difference equations

$$\phi_{1,i} = \frac{\phi_{2,i+1} - \phi_{2,i-1}}{2h}, \quad (4)$$

$$\phi_{1,i+1/2} = \frac{\phi_{2,i+1} - \phi_{2,i}}{h}. \quad (5)$$

Note that in this formulation  $\ddot{\phi}_2 = f_1(\phi_2)$ , which can be approximated by a second-order difference quotient

$$\phi_{2,i+1} - 2\phi_{2,i} + \phi_{2,i-1} = h^2 f_1(\phi_{1,i}). \quad (6)$$

Using the positions  $\phi_{2,i}$  and the momentum  $\phi_{1,i}$  along with the second derivative  $f_1(\phi_{1,i})$  and Equations 4, 5 and 6, we can determine a one-step evolution  $f : X \rightarrow X$  to get  $\{\phi_{1,i+1}, \phi_{2,i+1}\}$  defined by the following recursion:

$$\phi_{1,i+1/2} = \phi_{1,i} + \frac{h}{2} f_1(\phi_{2,i}), \quad (7)$$

$$\phi_{2,i+1} = \phi_{2,i} + h\phi_{1,i+1/2}, \quad (8)$$

$$\phi_{1,i+1} = \phi_{1,i+1/2} + \frac{h}{2} f_1(\phi_{2,i+1}). \quad (9)$$

Like the previous example, for any  $\phi_0$ , the orbit  $\Gamma^+(\phi_0) := \cup_{t \geq 0} \phi(t)$  through  $\phi_0$  is a smooth, compact, 1-dimensional Riemannian submanifold  $M_{\phi_0} \subseteq \mathbb{R}^2$ . In Figure 1, (d), several samples  $\Xi(\phi_0)$  denoted by the orange circles are collected along the discrete flow determined by the geometric numerical integration scheme of Equations 7 through 9 starting from initial condition  $\phi_0$ . These samples actually come from a limit cycle defined by a 1-dimensional manifold  $M_{\phi_0}$  which contains  $\phi_0$ . Similarly, samples  $\Xi(\psi_0)$  denoted by purple squares are collected around a separate limit cycle defined by the manifold  $M_{\psi_0}$  where the integration starts at different initial conditions  $\psi_0$ . Samples  $\Xi(\phi_0)$  and  $\Xi(\psi_0)$  will never intersect as they are confined to the manifolds  $M_{\phi_0}$  and  $M_{\psi_0}$  respectively as  $t \rightarrow \infty$ . Again, we are interested in understanding in what spaces approximations of the Koopman operator converge for such systems.

## 1.5 A Review of the New Results

The systems for which the theory in this paper applies include some types of discrete or continuous semidynamical systems that evolve on a complete metric space  $(X, d_X)$ . Most of the examples choose  $X$  as Euclidean space  $X := \mathbb{R}^d$  or as a compact Riemannian manifold  $M$ . Some examples are also described when the system evolves on or is attracted to certain types of  $k$ -dimensional smooth manifolds  $M$  that are regularly embedded in Euclidean space,  $M \subset X := \mathbb{R}^d$ . These latter cases are important in applications, such as illustrated in the examples, since it is not uncommon that samples  $\Xi := \Xi(\phi_0)$  are generated that fill sets that have zero measure as subsets of the state space  $X := \mathbb{R}^d$ .

Similar error bounds are derived in this paper for the two types of deterministic dynamical systems, both of which are very general. An evolution law like Equation 1 is the exemplar for case with discrete dynamics determined by a mapping  $f$ . For such systems it is well-known that the Koopman operator  $U_f$  is given as shown in Equation 2 with  $(U_f g)(x) = (g \circ f)(x)$  for each  $x \in X$ . We also define a type of Koopman operator for continuous semidynamical systems in continuous time. If  $t \mapsto \phi(t) \in X$  is a trajectory of the system through the initial condition  $\phi_0 \in X$ , we define the *orbital* Koopman operator  $\mathcal{U}$  from the identity  $(\mathcal{U}_\phi g)(t) = (g \circ \phi)(t)$  for all  $t \in [0, \infty) := T$ .<sup>1</sup>

The set of all samples collected along the flow is denoted by  $\Xi := \bigcup_{n \in \mathbb{N}} \Omega_n$  with  $\Omega_n := \{\xi_i \mid 1 \leq i \leq n\}$  the finite set of samples used to construct a particular approximation of the Koopman operator. The analysis of approximations of the Koopman operators is carried out by identifying a limiting subset  $\Omega \subseteq X$  in which the samples  $\Xi$  are dense. The study of approximations in the paper progresses from simple, rather unstructured cases to those that are more structured and can be considered as direct implications of well-known results on

<sup>1</sup> As discussed more fully in Section 3.1, the orbital Koopman operator  $\mathcal{U}_\phi$  associated with a trajectory  $t \mapsto \phi(t)$  differs from the conventional definition of the Koopman operator associated with a semiflow in continuous time. We are interested in the approximation of the orbital Koopman operator precisely because its approximation arises in the study of approximations of the RKH embedding method for adaptive estimation, a topic of much interest to the authors. [54–56]. Also see Example 7.

scattered data interpolation [57]. The study begins with the definition of a reproducing kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  that induces the RKH space  $H_X := \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in X\}$  of functions over the state space, which is the largest set of functions used to formalize the approximation problem. In this definition  $\mathfrak{K}_{\xi_i} := \mathfrak{K}(\xi_i, \cdot)$  is the kernel function centered at  $\xi_i \in X$ . See Section 2.2 for the properties of such spaces and associated definitions. Finite dimensional spaces of approximants  $H_{\Omega_n}$  are defined as finite spans of collections of such bases centered over scattered data as in  $H_{\Omega_n} := \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Omega_n\}$ , and the convergence of the approximate Koopman operators is cast in terms of the closed subspace  $H_\Omega := \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Omega\}$  that is *generated or indexed* in terms of the limiting set  $\Omega \subseteq X$ .

### 1.5.1 Projection Approximation $U_f^n$ Without Regularity Assumptions on $\Omega$

The first and simplest approximation  $U_f^n$  of the Koopman operator  $U_f$  in this paper is defined in terms of the  $H_X$ -orthogonal projection  $P_{\Omega_n} : H_X \rightarrow H_{\Omega_n}$ , and we set  $U_f^n g := (P_{\Omega_n} g) \circ f$ . Theorems 1 and 2 together imply that we have the error bound

$$|(U_f g)(x) - (U_f^n g)(x)| \leq \|(I - P_{\Omega_n})g\|_{H_X} \rightarrow 0$$

for each  $x \in X$  and  $g \in H_\Omega$  as  $n \rightarrow \infty$ . Roughly speaking, this convergence result is similar in nature to that in Theorem 3 of [41], which shows that as the dimension of the space of approximants goes to infinity, the approximate Koopman operator  $U_f^n$  approaches the true Koopman operator  $U_f$  strongly in the strong operator topology over  $L_\mu^2(X)$ . The equation above is entirely analogous and implies that convergence is guaranteed in the strong operator topology on maps from  $H_\Omega \rightarrow C(X)$ . Neither the result in [41], nor the analysis in Theorems 1 and 2 of this paper, describe the rates of convergence for different choices of spaces of approximants, however. They only imply convergence as the dimension of the space of approximants approaches infinity. A quick inspection of this pair of theorems shows that Theorem 1 implies the leftmost inequality above, while Theorem 2 implies that  $\|(I - P_{\Omega_n})g\|_{H_X} \rightarrow 0$  for  $g \in H_\Omega$ . This latter limit is shown to hold if the kernel  $\mathfrak{K}$  that induces the native space  $H_X$  has strong separation properties. These properties are known to hold for a large collection of kernels such as the Matern-Sobolev kernels, the Abel kernel, or the  $\ell^1$ -exponential kernel. However, the Gaussian kernels are known to lack the separation properties assumed in Theorem 2. [58] Thus, while very popular, the Gaussian kernels are not guaranteed by our analysis to generate such convergent approximations: in their case, more or other types of analysis are needed.

### 1.5.2 Projection Approximations $U_f^n$ , $\Omega$ a Regular set or Manifold

In this paper, the rather weak results in Theorems 1 and 2 for “unstructured” limit sets  $\Omega \subseteq X$  are improved when either the limiting set  $\Omega \subset X$  is a compact set having a Lipschitz boundary, or when  $\Omega := X := M$  is in fact a smooth Riemannian manifold. Theorem 11 gives sufficient conditions that the projection-based Koopman operator  $U_f^n(\cdot) := (P_{\Omega_n}(\cdot)) \circ f$  satisfies a bound such as

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(M))} \lesssim h_{\Omega_n, M}^{t-s} \|g\|_{W^{t,2}(M)} \quad (10)$$

for all  $g \in W^{t,2}(M)$  when the limiting set  $\Omega$  is in fact a smooth, connected, compact, Riemannian manifold  $\Omega := M$ . In this equation the error is measured in the pullback space  $f^*(W^{s,2}(M))$  that is defined in Section 2.2, the parameter  $h_{\Omega_n, M}$  is the fill distance of the finite samples in  $\Omega_n$  in the manifold  $M$ , and the ranges for the indices  $t, s$  are dictated by the Sobolev embedding theorem and the many zeros theorem on manifolds (see Appendix). When the samples  $\Xi$  are dense in a limiting set  $\Omega \subset X$  that is sufficiently regular, in the sense that it is compact and has a Lipschitz boundary, sufficient conditions for the similar bound

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(\Omega))} \lesssim h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)} \quad (11)$$

for all  $g \in W^{t,2}(\Omega)$ . Pointwise error bounds that are qualitatively similar to the last two Equations 10 and 11, ones that also bound the pointwise difference  $|(U_f g)(x) - (U_f^n g)(x)|$  in terms of a power of the fill distance, are given in Theorems 4, 5, and 6.

### 1.5.3 Data-Driven Approximations $\mathfrak{U}_f^n$ over $M$ a Smooth Manifold

The approximation  $U_f^n(\cdot) := (P_{\Omega_n}(\cdot)) \circ f$  studied above in Equations 10 and 11 is for the projection-based operator  $P_{\Omega_n} : H_M \rightarrow H_{\Omega_n}$ , but this expression cannot be evaluated unless the function  $f$  is known. Bounds on the error induced by the projection-based approximation  $U_f^n$  are certainly valuable to understand the “worst-case” performance of approximations built from a given finite dimensional space of approximants  $H_{\Omega_n}$ , and they are also important in their role in studying *data-dependent* approximations  $\mathfrak{U}_f^n g := P_{\Omega_n}((P_{\Omega_n} g) \circ f)$ . We discuss these data-dependent approximations next. As shown in Section 4, the operators  $\mathfrak{U}_f^n$  can be constructed

from the input-output samples  $\{(\phi_i, y_i)\}_{1 \leq i \leq n} = \{(\phi_i, f(\phi_i))\}_{1 \leq i \leq n}$  along the discrete trajectory of the system in Equation 1. It is also worth noting that the realization of the coordinate representation of  $\mathfrak{U}_f^n$  is closely related to the approximation of the Koopman operator that is defined in terms of the EDMD algorithm, provided that the number of samples is equal to the dimension of the space of approximants. This fact is explored in detail in Example 6. The definition of  $\mathfrak{U}_f^n$  makes sense only so long as  $(P_{\Omega_n}g) \circ f \subseteq H_M$ . Thus, a standing assumption in this case is that the pullback space  $f^*(H_M) \subseteq H_M$ . Since  $(P_{\Omega_n}g) \circ f \in f^*(H_M)$ , this structural assumption is enough to ensure that the data-driven operator  $\mathfrak{U}_f^n$  is well-defined. Theorem 11 is representative of the type of bound that can be derived in this case. We have a pointwise error bound

$$|(U_f g)(x) - (\mathfrak{U}_f^n g)(x)| \leq C_M h_{\Omega_n, M}^{t-s} (\|g\|_{W^{t,s}(M)} + \|U_f g\|_{W^{t,2}(M)})$$

in terms of the fill distance of the samples  $\Omega_n$  in the manifold  $M$ .

#### 1.5.4 Data-Driven Approximations $\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n$ Over a Regular Subset $\Omega \subseteq X$

In the last section it is noted that the data-driven operator  $\mathfrak{U}_f^n$  is well-defined if we assume that  $f^*(H_M) \subseteq H_M$ . But it is not immediately clear that this assumption is warranted, nor does it appear easy to verify in many cases. Alternatively, we can make a direct structural assumption on the function  $f$  that appears in the discrete evolution Equation 1. If we assume that  $f \in (W^{t,2}(\Omega))^d$ , we show following Theorem 8 that we have the pointwise bound

$$|(U_f g)(x) - (\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n g)(x)| \leq (\tilde{C}_\Omega + C_\Omega \|f\|_{(W^{t,2}(\Omega))^d}) h_{\Omega_n, \Omega}^{t-s} (\|g\|_{W^{t,s}(\Omega)} + \|U_f g\|_{W^{t,2}(\Omega)}),$$

which again depends on the fill distance of the samples  $\Omega_n$  in the limiting set  $\Omega$ . In contrast to the analysis of the data-dependent operator  $\mathfrak{U}_f^n$ , however, the data-dependent operator  $\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n$  requires a higher order embedding of the underlying RKH space in space of Lipschitz continuous functions. This is achieved in practice by appealing to the embedding of the Sobolev spaces in the Lipschitz spaces, a well-studied topic in the classical analysis of these spaces.

#### 1.5.5 Koopman Approximations over a Smooth Manifold $M$ Embedded in $\mathbb{R}^n$

The motivation behind considering the Koopman approximation over a manifold is that the samples generated by the dynamical system cover a smooth manifold in certain practical cases. However, it is essential to note that the manifold's explicit formulation (equations) is unknown in most cases. This, in turn, implies that determining the Riemannian metric on the manifold  $M$  can be challenging. Theoretically, we can derive such a metric by defining a kernel  $\mathfrak{K}$  on the manifold  $M$ . Then we can apply the results discussed in the previous sections to the problem at hand. However, this approach can be complex, even in the most straightforward practical cases. Fortunately, we can evoke the structure of the Euclidean space in which the manifold  $M$  is embedded to make the computation easier. Defining reproducing kernels over Euclidean space is a well-studied topic. We can now restrict the kernel to the manifold to obtain the convergence rates over the manifold  $M$ . The trace theorem, discussed in Section A.3 and [59], explicitly defines the Sobolev space in which the native RKH space induced by the restricted kernel is contained. It is now possible to understand the convergence rates in the restricted Sobolev space. The kernel restriction to the manifold  $M$  results in a loss of smoothness of the Sobolev space, leading to weaker convergence rates. However, this extrinsic approach to approximating the Koopman operator does not require an explicit definition of the kernel over the manifold  $M$ , thus making it easier to implement than the intrinsic approach detailed above.

In the remainder of this paper, we give the details of the analysis that establishes and supports the new results summarized above. The presentation of the details includes many auxiliary results that may be of independent interest to the reader, and the authors emphasize these points whenever they appear.

## 2 Definitions and Background

### 2.1 Notation and Symbols

In this section we begin a description of the underlying theory, and make clear common notation and symbols. In the paper, the symbols  $\mathbb{R}, \mathbb{R}^+, \mathbb{N}, \mathbb{N}_0 := \mathbb{N} \cup \{0\}$  denote the real numbers, nonnegative real numbers, positive integers, and nonnegative integers, respectively. The functions  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$  are the ceiling and floor functions that return the smallest integer greater than, or the largest integer less than, a given real number. When we write  $a \lesssim b$ , it means that there is a positive constant  $c$  that does not depend on  $a, b$  such that  $a \leq c \cdot b$ . A similar definition holds for the relation ' $\gtrsim$ ', and we write  $a \approx b$  when  $a \gtrsim b$  and  $a \lesssim b$ . We use  $X$  to denote the state space of a system, which is always assumed to be a complete metric space  $(X, d_X)$ . In the paper

the most general system is taken to be a semidynamical system defined in terms of a continuous semigroup  $\{S(t)\}_{t \in T}$  with the time-indexing set  $T = \mathbb{R}^+$  or  $T = \mathbb{N}_0$ . By  $L^p(\Omega) := L^p_\mu(\Omega)$  we refer to the usual Banach space of  $p$ -integrable functions over  $\Omega \subseteq X$  with respect to the measure  $\mu$  on  $X$  for  $1 \leq p < \infty$ , with the usual modification for  $p = \infty$ . The space  $C(\Omega)$  is the collection of bounded continuous functions  $f : \Omega \rightarrow \mathbb{R}$  on  $\Omega \subseteq X$  equipped with the usual supremum norm, and  $C(\Omega, \mathbb{R}^d)$  is the associated space of vector-valued, continuous, and bounded functions  $f : \Omega \rightarrow \mathbb{R}^d$ . We denote by  $C^{0,1}(\overline{\Omega})$  the space of Lipschitz function over a closed subset  $\overline{\Omega}$ , where the norm  $\|f\|_{C^{0,1}(\overline{\Omega})} = \|f\|_{C(\Omega)} + |f|_{C^{0,1}(\overline{\Omega})}$  is defined in Equation 25 in Section A.3. As discussed more carefully in Section 2.2 below,  $H_X$  will always denote an RKH space of real-valued functions defined over the state space  $X$  that is induced by an admissible kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$ . When  $(U, \|\cdot\|_U)$  and  $(V, \|\cdot\|_V)$  are normed vector spaces, we say that  $U$  is continuously embedded in  $V$  whenever  $U \subseteq V$ , and the canonical injection  $i : u \in U \mapsto i(u) = u \in V$  is a bounded operator. That is,  $\|u\|_V := \|i(u)\|_V \lesssim \|u\|_U$  for all  $u \in U$ . We abbreviate the property by writing  $U \hookrightarrow V$ , or  $U \xrightarrow{i} V$  if we want to be precise about which canonical injection  $i$  is involved in the embedding.

## 2.2 Structures in RKH Spaces

As noted in the introduction, in this paper it is frequently the case that  $H_X$  is an RKH space of real-valued functions over the state space  $X := \mathbb{R}^d$ . However, we also study some examples where the entire state space  $X$  is in fact a smooth manifold  $X := M$ , and we also consider the RKH spaces  $H_M$  when  $M$  is a submanifold that is regularly embedded in the state space  $X := \mathbb{R}^d$ . The theory of RKH spaces in this section is stated with  $X$  an arbitrary set, keeping in mind these intended applications. This section reviews the basic definitions and properties of the RKH space, as well as several auxiliary spaces that are commonly associated with such an RKH space  $H_X$ . These spaces include the pullback space  $\phi^*(H_X)$  of time-dependent functions defined in terms of a mapping  $\phi : T \rightarrow X$  with respect to an arbitrary time set  $T$ , the pullback space  $f^*(H_X)$  of spatial functions defined in terms of a mapping  $f : X \rightarrow X$ , the space of restrictions  $R_\Omega(H_X)$  of functions in  $H_X$  to a domain  $\Omega \subseteq X$ , and the subspace  $H_\Omega \subseteq H_X$  defined in terms of a generating set  $\Omega \subseteq X$ . We also discuss and relate interpolation and projection operators defined on these spaces. The following discussion of the supporting theory of RKH spaces is necessarily concise, and the reader is referred to comprehensive treatments in [60–62] for a detailed account.

### 2.2.1 RKH Space $H_X$ and $H_\Omega \subseteq H_X$ of Functions over $X$

The set  $X$  is the state space of the uncertain evolution law in this paper, and all the function spaces introduced in the paper are derived from a real RKH space  $H_X$  of functions over  $X$ . A real-valued function  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  is an admissible kernel function if it is symmetric, continuous, and of positive type. The kernel is of positive type if for any finite collection of points  $\{\xi_i\}_{1 \leq i \leq n} \subseteq X$ , the Grammian or collocation matrix  $\mathbb{K}_n := [\mathfrak{K}(\xi_i, \xi_j)]$  is positive semi-definite. The kernel is of strictly positive type if the Grammian matrix is positive definite for any finite collection of distinct points. The real Hilbert space  $H_X$  is defined in terms an admissible kernel [60] function  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  as

$$H_X := \overline{\text{span}\{\mathfrak{K}_x \mid x \in X\}} \quad (12)$$

with  $\mathfrak{K}_x(\cdot) := \mathfrak{K}(x, \cdot)$  referred to as the kernel basis function located at  $x \in X$ . The closure is taken with respect to the candidate inner product  $(\cdot, \cdot)_{H_X}$  that is defined for any two functions  $\mathfrak{K}_x, \mathfrak{K}_y$  by  $(\mathfrak{K}_x, \mathfrak{K}_y)_{H_X} := \mathfrak{K}(x, y)$  for all  $x, y \in X$ . The RKH space  $H_X$  is said to be the native space induced by the kernel  $\mathfrak{K}$ . An RKH space gets its name from the fact that  $(f, \mathfrak{K}_x)_{H_X} = f(x)$  for all  $f \in H_X$  and  $x \in X$ , which is known as the reproducing property of the RKH space. Not every Hilbert space  $H$  is an RKH space: it is known for instance that the Hilbert space of Lebesgue square integrable functions  $L^2(X)$  is not an RKH space. However, it is also known that a Hilbert space  $H$  is an RKH space if and only if all the evaluation functionals  $\mathcal{E}_x : f \mapsto f(x)$  are bounded operators from  $H$  to  $\mathbb{R}$ . A sufficient condition for this is that there is a constant  $\bar{k}$  such that  $\sup_{x \in X} \mathfrak{K}(x, x) \leq \bar{k}^2$ . Then we have the uniform norm bound  $\|\mathfrak{K}_x\| \leq \bar{k}$  and the uniform operator norm bound  $|\mathcal{E}_x| \leq \bar{k}$ . Since  $\mathfrak{K}$  that induces  $H_X$  is assumed to be continuous, the existence of such an upper bound  $\bar{k}$  is sufficient to ensure that  $H_X \hookrightarrow C_b(X)$ .

It is evident from the above definitions that it is always possible to define a closed subspace  $H_\Omega \subseteq H_X$  that is generated by an arbitrary set  $\Omega \subseteq X$  as

$$H_\Omega := \overline{\text{span}\{\mathfrak{K}_x \mid x \in \Omega \subseteq X\}}. \quad (13)$$

We refer to the set  $\Omega$  in the definition of the above space  $H_\Omega$  as the generating set of the space  $H_\Omega$ . Note carefully that  $H_\Omega$  are functions defined over all of  $X$ , not functions only defined just on  $\Omega$ . We consider the restriction of functions below in discussions of the RKH spaces  $R_\Omega(H_X)$ . In the definition of  $H_\Omega$ , the closure is taken with respect to the norm  $\|\cdot\|_{H_X}$ , and  $H_\Omega$  is endowed with the usual topology it inherits as a closed

subspace of  $H_X$ . It is a standard result in the theory of RKH spaces that the space  $H_\Omega$  is an RKH space itself, and its kernel  $\mathfrak{K}_\Omega$  can be written in terms of the  $H_X$ -orthogonal projection  $P_\Omega$  of  $H_X$  onto  $H_\Omega$  and the kernel  $\mathfrak{K}$  as

$$\mathfrak{K}_\Omega(x, y) := (P_\Omega \mathfrak{K}_x, P_\Omega \mathfrak{K}_y)_{H_X} \quad (14)$$

for all  $x, y \in X$ . If we set

$$Z_\Omega := \{f \in H_X \mid f|_\Omega = 0\},$$

then it can be shown directly from the definition of  $H_\Omega$  that we have the  $H_X$ -orthogonal decomposition

$$H_X = H_\Omega \oplus Z_\Omega.$$

The orthogonal complement of  $H_\Omega$  is thereby the set of functions in  $H_X$  that vanish identically on  $\Omega \subseteq X$ . In the above discussion, the definition of  $P_\Omega$  is given for scalar-valued functions in  $H_X$ . We overload this notation for vector-valued functions and define  $P_\Omega f := \{P_\Omega f_1, \dots, P_\Omega f_d\}^T$  for all  $f \in H_X^d$ .

### 2.2.2 The RKH Space $R_\Omega(H_X)$ of Restrictions to $\Omega \subseteq X$

When we have defined a kernel  $\mathfrak{K}$  on  $X$  that induces a native space  $H_X$ , there is a standard way to define an RKH space  $R_\Omega(H_X)$  of restrictions of functions in  $H_X$  to a subset  $\Omega \subseteq X$ . This set is given by

$$R_\Omega(H_X) := \{f \in H_X \mid g = R_\Omega f, \quad f \in H_X\}$$

with  $R_\Omega$  the restriction operator  $R_\Omega f = f|_\Omega$ .<sup>2</sup> As discussed in detail in [60, 61], the kernel  $r : \Omega \times \Omega \rightarrow \mathbb{R}$  of the RKH space  $R_\Omega(H_X)$  is

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

for all  $x, y \in \Omega$ . For any  $h \in R_\Omega(H_X)$ , this definition of the kernel induces a norm that is equivalent to

$$\|h\|_{R_\Omega(H_X)} \approx \min \{\|f\|_{H_X} \mid h = R_\Omega f, \quad f \in H_X\}.$$

Note that, by virtue of the above definition, it is immediate that  $\|R_\Omega g\|_{R_\Omega(H_X)} \leq \|g\|_{H_X}$  for all  $g \in H_X$ , and the operator  $R_\Omega : H_X \rightarrow R_\Omega(H_X)$  is norm bounded with  $\|R_\Omega\| \leq 1$ .

The spaces  $H_\Omega$  and  $R_\Omega(H_X)$  are closely related. We can identify the space  $Z_\Omega$  that appears in in the orthogonal decomposition  $H_X := H_\Omega \oplus Z_\Omega$  as the kernel of the restriction operator  $R_\Omega : H_X \rightarrow R_\Omega(H_X)$ , that is,  $Z_\Omega = \ker(R_\Omega)$ . The restriction map  $R_\Omega$  is linear and bounded on  $H_X$ . Moreover,  $R_\Omega|_{Z_\Omega^\perp} = R_\Omega|_{H_\Omega}$  is injective. This means that  $(R_\Omega|_{H_\Omega})^{-1}$  exists and is a bounded map from  $R_\Omega(H_X) \rightarrow H_\Omega$ . We define the extension operator  $\mathbb{E}_X : R_\Omega(H_X) \rightarrow H_X$  by the equation

$$\mathbb{E}_X := (R_\Omega|_{H_\Omega})^{-1}.$$

The extension operator enables an explicit equivalent expression for the inner product on  $R_\Omega(H_X)$  as

$$(g, h)_{R_\Omega(H_X)} = (\mathbb{E}_X g, \mathbb{E}_X h)_{H_X},$$

and in fact it follows from this definition that the operator norm  $\|\mathbb{E}_X\| \leq 1$  too. With this definition of the inner product on  $R_\Omega(H_X)$ , it is possible to show that the orthogonal projection  $P_\Omega$  satisfies

$$P_\Omega = \mathbb{E}_X R_\Omega.$$

It is also pointed out in some of the proofs that follow that the identity above implies some useful norm equivalences. For any  $g \in H_\Omega \subseteq H_X$  we have

$$\|g\|_{H_\Omega} = \|P_\Omega g\|_{H_X} = \|\mathbb{E}_X R_\Omega g\|_{H_X} = \|R_\Omega g\|_{R_\Omega(H_X)}.$$

A function  $g \in H_\Omega$ , which is defined on all of  $X$ , is the zero function in  $H_\Omega$  if and only if the restriction  $R_\Omega g$  is zero over  $\Omega$ . As for the projection operator  $P_\Omega$ , we define the action of  $\mathbb{E}_X$  and  $R_\Omega$  on vector-valued functions as  $\mathbb{E}_X g := \{\mathbb{E}_X g_1, \dots, \mathbb{E}_X g_d\}^T$  and  $R_\Omega f := \{R_\Omega f_1, \dots, R_\Omega f_d\}^T$  for  $g \in (R_\Omega(H_X))^d$  and  $f \in H_X^d$ , respectively.

<sup>2</sup> It is well-known [60–62] that the construction of the space of restrictions  $R_\Omega(H_X)$  can be understood a special case of a pullback space generated by the canonical injection  $\mathcal{I} : \Omega \rightarrow X$  that associates  $\mathcal{I} : x \in \Omega \mapsto x = \mathcal{I}x \in X$ . However, we will only use the notion of a pullback space for  $\phi^*(H_X)$  and  $f^*(H_X)$  in this paper. One reason for this choice is that when we introduce Sobolev spaces over the spatial domain, and relate them to RKH spaces, trace operator has a natural interpretation and relationship to the restriction operator  $R_\Omega$ .

### 2.2.3 The Pullback RKH Spaces $\gamma^*(H_X)$ for $\gamma : S \rightarrow X$

When  $H_X$  is an RKH space, there is a standard construction of pullback spaces that are defined in terms of  $H_X$  and a mapping into  $X$ , which we describe next. For the RKH space  $H_X$ , any set  $S$ , and any mapping  $\gamma : S \rightarrow X$ , the pullback RKH space  $\gamma^*(H_X)$  is defined to be the set of functions

$$\gamma^*(H_X) := \{g : S \rightarrow \mathbb{R} \mid g = f \circ \gamma, \quad f \in H_X\},$$

which is the space of compositions of functions in  $H_X$  with the fixed map  $\gamma$ . The norm on the pullback space  $\gamma^*(H_X)$  is defined to be

$$\|g\|_{\gamma^*(H_X)} := \min \left\{ \|f\|_{H_X} \mid g = f \circ \gamma, \quad f \in H_X \right\}.$$

As discussed in [62], the pullback space  $\gamma^*(H_X)$  is itself an RKH space with kernel  $\mathfrak{K}_\gamma$  defined as

$$\mathfrak{K}_\gamma(\tau, s) := \mathfrak{K}(\gamma(\tau), \gamma(s))$$

for all  $\tau, s \in S$ . This implies that  $\gamma^*(H_X) := \overline{\text{span}\{\mathfrak{K}_{\gamma, s} \mid s \in S\}}$  with  $\mathfrak{K}_{\gamma, s} := \mathfrak{K}(\gamma(s), \gamma(\cdot))$  for each  $s \in S$ .

In this paper we interpret the trajectory  $t \mapsto \phi(t)$  of the uncertain dynamical system as defining the mapping  $\phi : T \rightarrow X$  with  $T := [0, \infty)$ , and the specific pullback space  $\phi^*(H_X)$  in this paper is subsequently a space of functions of time. The pullback space  $\phi^*(H_X)$  is therefore used to study semiflows in continuous time. We also have occasion to use the pullback space  $f^*(H_X)$  that is the pullback space generated by the mapping  $f : X \rightarrow X$  that appears in the discrete evolution law in Equation 1. The definition of either of the spaces  $f^*(H_X)$  or  $\phi^*(H_X)$ , their kernels, their norms, etc, are defined exactly as in the case of  $\gamma^*(H_X)$ , but replacing the mapping  $\gamma$  with either  $f$  or  $\phi$ , respectively. Finally, we note that since  $R_\Omega(H_X)$  and  $H_\Omega$  are RKH spaces in their own right, it also makes sense to define the pullback spaces such as  $\phi^*(R_\Omega(H_X))$  or  $f^*(H_\Omega)$ . These spaces are particularly useful when the samples  $\Xi$  of the underlying flow are not dense in the state space  $X$ , but are dense in a limiting set that is a proper subset  $\Omega \subset X$ .

*Example 5* The last few sections have reviewed some of the general properties of RKH spaces, and in this section we summarize a few of the common kernels that define RKH spaces in some of the examples. There are a large variety of admissible kernels  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  that induce native spaces  $H_X$ . See [57] for a good overview that focuses for the most part on kernels defined over subsets  $\Omega \times \Omega \subseteq \mathbb{R}^d \times \mathbb{R}^d$ . A very brief discussion of certain kernels defined on smooth manifolds  $M$  can be found in Chapter 17 of the same reference [57], but most of the guarantees of convergence in the paper for are cast in terms of kernels described in the family of recent papers [63–65]. We should also point out that the kernels described below and used in this paper are all examples of *positive definite* kernels. As carefully discussed in [57], there is a larger family of *conditionally positive definite* kernels, but we will not consider this level of generality when building the native spaces in this paper.

*Exponential Kernels:* Of all of the kernels that appear in applications to dynamical systems, it seems as if the most popular are the exponential kernels. Many kernels can be defined in terms of the exponential function, and we refer in this paper to a few that have the form

$$\mathfrak{K}_{\alpha, p, q}(x, y) := e^{-\alpha \|x - y\|_{\ell^q(\mathbb{R}^d)}^p}$$

with  $\alpha > 0$  a real constant and  $1 \leq p, q \leq \infty$ . The most well-known of these kernels are the Gaussian kernels  $\mathfrak{K}_{\alpha, 2, 2}$ , but we use the Abel kernel  $\mathfrak{K}_{\alpha, 1, 2}$  and the  $\ell^1$ -exponential kernel  $\mathfrak{K}_{\alpha, 1, 1}$  too because of their excellent separation properties. See [58], Section 3 for an excellent commentary on the separation properties of some of these exponential kernels. The exponential kernel is popular since it has a simple closed form representation, it is smooth, and it is a positive definite kernel for any  $\alpha > 0$ . This kernel also appears in many papers that study statistical or machine learning theory, or Gaussian process methods, in terms of RKH spaces. See [66] for a clear and concise discussion of the interrelationships among these fields. Those familiar with Bayesian estimation will recognize that native spaces defined in terms of exponential functions are used to characterize covariance operators.

*Inverse Multiquadrics* The kernels  $\mathfrak{K}_\beta(x, y) := 1/(c^2 + \|x - y\|_{\mathbb{R}^d}^2)^\beta$  are known as the family of inverse multiquadrics, and they have much in common with the exponential kernels. For any  $\beta > 0$  these functions are also smooth, have a simple closed form representation, and are positive definite on  $\mathbb{R}^d \times \mathbb{R}^d$  for any  $\beta > 0$ . In principle, the use of the exponential or inverse multiquadric kernels can be used to define interpolation operators that have convergence rates of arbitrarily high polynomial order, even exponential rates of convergence, as discussed in Section 11.4 of [57]. Unfortunately, the condition number of the Grammian matrices for exponential kernels can be prohibitively high, which induces numerical stability issues that can make implementations problematic.

*Compactly Supported Polynomial RBF of Minimal Degree:* We will see that interpolation operators generated in terms of kernel functions of a RKH space are expressed in terms of the inverse of the Grammian matrix



$\mathbb{K}(\Omega_n) := [\mathfrak{K}(\xi_i, \xi_j)]_{\xi_i, \xi_j \in \Omega_n}$  associated with a finite dimensional number of samples  $\Omega_n := \{\xi_1, \dots, \xi_n\}$ . For kernels that are globally supported like the exponential or inverse multiquadric functions, the Grammian matrix is fully populated. It can be much more efficient to use kernels that are only locally supported, which can imply that many entries of the Grammian (and its inverse) are zero. This leads to much more efficient representations of approximations of functions. In a fashion that is analogous or reminiscent of techniques from finite element methods [67], there are standard constructions of kernel functions  $\mathfrak{K}_{d,k}$  on  $\mathbb{R}^d \times \mathbb{R}^d$  that are piecewise polynomials, have compact support, and are positive definite for any  $d \geq 1$  and even integers  $k > 0$ . The definition of these kernels varies as a function of the dimension  $d$  and even integer index  $k > 0$ , so they are not as simple in form as either exponential or inverse multiquadric functions. See [57], Table 9.1 for a summary of some of the more common, low order versions of these functions.

*Intrinsic Matern-Sobolev Kernels:* So far, the positive definite kernels have been defined over  $\mathbb{R}^d \times \mathbb{R}^d$ , which means that their restrictions to any subset  $\Omega \times \Omega$  defines a positive definite kernel on  $\Omega \times \Omega$ . In this paper we also make use of Matern-Sobolev kernels that are defined on certain types of smooth Riemannian manifolds  $M$ . The definitions of these kernels require a bit more information about smooth manifolds. When  $M$  is a  $d$ -dimensional, closed, compact, connected, smooth Riemannian manifold, we denote by  $W^{s,2}(M)$  the Sobolev space having real smoothness index  $s > 0$ . The careful definition and basic properties of these spaces are described in Section A that follows. If  $s > d/2$ , these are examples of RKH spaces. In the case that the smoothness index is an integer  $s := m$ , the reproducing kernel  $\mathfrak{K}_m$  of the space  $W^{m,2}(M)$  is the fundamental solution of the operator equation  $\mathcal{L}u := \sum_{j=0}^m (\nabla^j)^* \nabla^j u = \delta$  of order  $2m$  with  $\delta$  the Dirac distribution,  $\mathcal{L}$  the Laplace-Beltrami operator on  $M$ , and  $\nabla$  the covariant derivative operator on  $M$ . In some classical choices of the compact manifold  $M$ , say, if  $M$  is the circle  $S^1$ , the sphere  $S^2$ , or the torus  $T^d$ , closed form expressions for the Matern-Sobolev kernel are known, see [63–65]. Spherical harmonics on the sphere are eigenfunctions of the Laplace-Beltrami operator  $\mathcal{L}$  on  $S^{d-1}$ , [57], for instance. The Sobolev-Matern kernels are strictly positive definite on  $M$ .

*Extrinsic Matern-Sobolev Kernels:* In applications to the approximation of Koopman operator, there is an underlying assumption that the dynamics is unknown, or at least uncertain. While the Sobolev-Matern kernels are attractive in that they are an example of a positive definite kernel over a compact manifold, if the exact form of the manifold  $M$  is not known, the closed form expression for these kernels will not be readily available. However, the Sobolev-Matern kernels can also be defined for the Sobolev spaces  $W^{m,2}(\mathbb{R}^d)$  when  $m > d/2$ , and closed form expressions for these these RKH spaces are known. In the case that  $M$  is a  $k$ -dimensional, closed, compact, connected, smooth, regularly embedded submanifold  $M \subset X := \mathbb{R}^d$ , we can define kernels for  $W^{r,2}(M)$  by the restriction of the kernels defined on  $X := \mathbb{R}^d$  to the set  $M \subset X$ , with certain restrictions on the index  $r$ . Even though intrinsic kernels on  $M$  are unavailable or intractable to compute, this strategy can be used to define a kernel on  $M$  that is easy to implement in calculations even if a closed form expression for  $M$  is unavailable.

#### 2.2.4 Interpolation and Projection

Approximations of the Koopman operators are constructed in terms of finite dimensional subspaces of the spaces  $H_X, H_\Omega$ , or  $R_\Omega(H_X)$ , depending on the particular application. Associated with the finite set of points

$$\Omega_n := \{\xi_1, \dots, \xi_n\} \subseteq X,$$

we let  $H_{\Omega_n}$  be the  $n$ -dimensional space of approximants  $H_{\Omega_n} := \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Omega_n\}$ . We define the  $H_X$ -orthogonal projection  $P_{\Omega_n} : H_X \rightarrow H_{\Omega_n}$  in the usual way, in terms of the identity

$$((I - P_{\Omega_n})f, g)_{H_X} = 0$$

for all  $g \in H_{\Omega_n}$ . We define the interpolation operator  $I_{\Omega_n} : H_X \rightarrow H_{\Omega_n}$  to be the unique operator that satisfies the interpolation conditions over  $\Omega_n$  for each  $f \in H_X$ ,

$$(I_{\Omega_n} f)(\xi_i) = f(\xi_i)$$

for all  $\xi_i \in \Omega_n$ . While it is not true for functions in a general Hilbert space, for the RKH space  $H_X$  the interpolation and projection operators coincide,  $I_{\Omega_n} = P_{\Omega_n}$ .

In the same fashion, we define the  $R_\Omega(H_X)$ -orthogonal projections  $\mathbb{P}_{\Omega_n}$  and  $\mathbb{I}_{\Omega_n}$  with respect to the  $n$ -dimensional approximant spaces  $R_\Omega(H_{\Omega_n})$  that are given by

$$R_\Omega(H_{\Omega_n}) := \text{span}\{r_{\xi_i} \mid \xi_i \in \Omega_n\}$$

where as discussed above  $r_\xi := R_\Omega \mathfrak{K}_\xi$  for  $\xi \in \Omega$ . The projections  $P_{\Omega_n}$  and  $\mathbb{P}_{\Omega_n}$  are related by the identities

$$\begin{aligned} R_\Omega P_{\Omega_n} &= \mathbb{P}_{\Omega_n} R_\Omega, \\ \mathbb{E}_X \mathbb{P}_{\Omega_n} &= P_{\Omega_n} \mathbb{E}_X. \end{aligned}$$

These identities can be inferred by manipulation of coordinate expressions, as is carried out in [68]. Alternatively, we know that for each  $g \in R_\Omega(H_X)$  the interpolation operator  $\mathbb{I}_{\Omega_n}$  satisfies  $(\mathbb{I}_{\Omega_n}g)(\xi) = g(\xi_i)$  for all  $\xi_i \in \Omega_n$ . Because  $\Omega_n \subseteq \Omega$ , we have

$$(\mathbb{E}_X \mathbb{I}_{\Omega_n} g)(\xi_i) = (\mathbb{E}_X g)(\xi_i) = (I_n \mathbb{E}_X g)(\xi_i) = (P_{\Omega_n} \mathbb{E}_X g)(\xi_i)$$

for each  $\xi_i \in \Omega_n$ . Since the range of  $\mathbb{E}_X \mathbb{I}_{\Omega_n}$  and  $P_{\Omega_n} \mathbb{E}_X$  are  $n$ -dimensional, these relations and the positive definiteness of the kernel suffice to prove that  $\mathbb{E}_X \mathbb{I}_{\Omega_n} = \mathbb{E}_X \mathbb{P}_{\Omega_n} = P_{\Omega_n} \mathbb{E}_X = I_{\Omega_n} \mathbb{E}_X$ . We note that identities such as the above have been used elsewhere in the study of approximations in RKH spaces, see [69].

The operators  $P_{\Omega_n}, \mathbb{P}_{\Omega_n}$  above are defined for scalar-valued functions. With an abuse of notation we will overload these definitions so that they apply component-wise on vector-valued functions. We set

$$P_{\Omega_n} f := \begin{Bmatrix} P_{\Omega_n} f_1 \\ \vdots \\ P_{\Omega_n} f_d \end{Bmatrix} \quad \text{and} \quad \mathbb{P}_{\Omega_n} g := \begin{Bmatrix} \mathbb{P}_{\Omega_n} g_1 \\ \vdots \\ \mathbb{P}_{\Omega_n} g_d \end{Bmatrix}$$

for functions  $f \in H_X^d$  and  $g \in (R_\Omega(H_X))^d$ .

### 2.3 Sobolev Spaces

The theory of RKH spaces outlined in the last section is general and allows for functions defined on an arbitrary set  $X$ . The strongest convergence results derived in this paper are derived when the RKH space  $H_X$ , or one of its associated spaces such as  $H_\Omega$  or  $R_\Omega(H_X)$ , can be interpreted as a Sobolev space. The Sobolev spaces are some of the most common spaces used to describe smoothness and approximation properties of functions. [70] In this paper we will use Sobolev spaces  $W^{r,2}(\Omega)$  for real  $r > 0$  that contain functions defined over a subset of  $\Omega \subseteq X := \mathbb{R}^d$ , as well as the spaces  $W^{r,2}(M)$  that contain functions defined over certain types of manifolds  $M \subseteq X$ . The definitions and theorems for these spaces are very similar in form, but differ in the fine details of certain hypotheses. A more detailed discussion on Sobolev spaces is given in Appendix A.

## 3 Dynamical Systems and Koopman Theory

As noted in the introduction, Koopman theory has been developed for deterministic or stochastic systems, and for both of these classes in discrete or continuous time. Variants of the theory have most often been studied when the unknown underlying system takes values in the Euclidean space  $\mathbb{R}^d$ , but some work has also been carried out for evolutions that take values in certain types of manifolds  $M$ . While we certainly cannot treat all the diverse and nuanced differences in these cases, it is a primary goal of this paper to derive error bounds in a framework which, hopefully, can be extended to many other scenarios. We limit the study here to two types of *deterministic* dynamic systems, one class in continuous time, and a second in discrete time. We find in the study that the derivation of the rate of convergence of the approximations of the Koopman operators for the two classes have much in common.

Also, as discussed more fully in [38], the error analysis of stochastic Koopman operators can be carried out by decomposing the total error into an approximation space error that is known as the bias term, and a stochastic error that is referred to as the variance term. This is a classical decomposition of the error for function estimates that are built from stochastically generated samples. This paper, then, can be viewed as a study of the approximation space or bias error term: it can inform the analysis of part of the total error for stochastic Koopman operators. The full study of the stochastic case would far exceed the scope of this paper, and we leave the topic for future study.

### 3.1 Deterministic Evolutions in Continuous Time

We first discuss deterministic evolutions in continuous time, and we denote the time indexing set  $T := \mathbb{R}^+$ . We assume that the underlying unknown dynamics constitute a continuous semidynamical system that we take to be defined in terms of continuous semigroup of operators  $\{S(t)\}_{t \in T}$  that act on the complete metric space  $(X, d_X)$ . In most of the examples we choose  $X := \mathbb{R}^d$ , although we do consider cases in which the evolution takes place in an embedded submanifold  $M \subseteq X$ . As per the usual definitions [71, 72], the positive or forward orbit  $\Gamma^+(\phi_0)$  through the initial condition  $\phi_0 \in X$  is

$$\Gamma^+(\phi_0) := \bigcup_{t \in [0, \infty)} S(t)\phi_0 = \bigcup_{t \in [0, \infty)} \phi(t)$$

with  $t \mapsto \phi(t) := S(t)\phi_0$  the motion or trajectory through  $\phi_0$  at  $t = 0$ . It will be useful to distinguish the positive limit set  $\omega^+(\phi_0)$  associated with an orbit  $\Gamma^+(\phi_0)$ :

$$\omega^+(\phi_0) = \left\{ y \in X \mid \exists t_i \rightarrow \infty \text{ such that } S(t_i)y \rightarrow y \right\}.$$

The positive limit set will be important to our considerations in that it will make precise the largest set for which some of the error bounds for the data-dependent approximations make sense. Note that if the forward orbit  $\Gamma^+(\phi_0)$  is precompact, then the positive limit set  $\Gamma^+(\phi_0)$  is nonempty, connected, compact, and attracts the flow. The positive limit set in this case provides a well-defined starting point from which to extract sequences to generate convergent approximations of Koopman operators in this paper.

Associated with the state space  $X$ , a real-valued RKH space  $H_X$  of functions over  $X$  is defined in terms of an admissible kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$ . We define the *orbital* Koopman operator  $\mathcal{U}_\phi$  for a trajectory  $t \mapsto \phi(t) := S(t)\phi_0$  of the system via the familiar relation

$$(\mathcal{U}_\phi g)(t) = (g \circ \phi)(t), \quad (15)$$

and in this way the orbital Koopman operator is a map  $\mathcal{U}_\phi : H_X \rightarrow \phi^*(H_X)$ , with  $\phi^*(H_X)$  the pullback space generated by the map  $\phi : T \rightarrow X$  and  $T := [0, \infty)$ . In fact  $\mathcal{U}_\phi$  is a bounded and linear mapping between  $H_X$  and  $\phi^*(H_X)$ . Boundedness follows since

$$\|\mathcal{U}_\phi g\|_{\phi^*(H_X)} = \min\{\|h\|_{H_X} \mid g \circ f = h \circ f, h \in H_X\} \leq \|g\|_{H_X}, \quad (16)$$

which shows that  $\|\mathcal{U}_\phi\| \leq 1$ .

The definition of the orbital Koopman operator  $\mathcal{U}_\phi$  in Equation 15 is given for  $g \in H_X$ , but we also study these operators for  $g \in R_\Omega(H_X)$ . Since in this case  $g : \Omega \rightarrow \mathbb{R}$ , here the Koopman operator  $\mathcal{U}_\phi g := g \circ \phi$  makes sense when the orbit  $\Gamma^+(\phi_0) \subset \Omega$ . This case will be studied used when the samples  $\Xi$  do not fill the state space  $X$ , but are dense in the limiting set  $\Omega \subset X$ . Carefully note that the identity  $\mathcal{U}_\phi g := g \circ \phi$  actually defines two different operators for the cases when  $g \in H_X$  or  $g \in R_\Omega(H_X)$  since they have different domains. We will abuse notation and refer to both as  $\mathcal{U}_\phi$ . Since, following the same logic as in Equation 16 above, we have  $\|\mathcal{U}_\phi g\|_{\phi^*(R_\Omega(H_X))} \leq \|g\|_{R_\Omega(H_X)}$ , it follows that the operator norm  $\|\mathcal{U}_\phi\| \leq 1$  in either definition of  $\mathcal{U}_\phi$ .

Those familiar with constructions in RKH spaces will recognize the above definition of the Koopman operator  $\mathcal{U}_\phi$  as that of the composition operator encountered in RKH theory. [62] On the other hand, those familiar with Koopman theory will note that this definition differs somewhat from that given for continuous time systems. For continuous evolutions, it is more common to define a different Koopman operator, one that is time-dependent. The time-dependent Koopman operator is defined via the identity  $(U_t f)(x) := f(S(t)x)$ , and then we have  $U_t : H_X \rightarrow H_X$ . Carefully observe that the domain and range of the operator  $U_t$  is the same space, while they are different spaces entirely for the orbital Koopman operator. It is this difference that motivates referring to  $\mathcal{U}_\phi$  as the orbital Koopman operator since it depends on the entire orbit, and is not an operator defined for fixed  $t$ . We do not study the approximation of operator  $U_t$  in this paper. Historically, the authors' interest in bounds for approximations of the orbital Koopman operator  $\mathcal{U}_\phi$  has arisen in the study the RKH embedding method for uncertain systems in [56, 68].

### 3.2 Deterministic Evolutions in Discrete Time

Evolutions in discrete time in this paper are governed by the nonlinear, deterministic recursion

$$\phi_{n+1} = f(\phi_n)$$

where  $\phi_n \in X$ ,  $(X, d)$  is a complete metric space, and  $f : X \rightarrow X$  is a generally nonlinear, but unknown, function. To keep the exposition and notation simple, we use an overloaded notation to describe this system in a way that is similar to the continuous case. For the discrete system we set the time index set to be the nonnegative integers  $T = \mathbb{N}_0^+$  and let the discrete semigroup be defined as  $\{S(t)\}_{t \in T} := \{S(t)\}_{t \in \mathbb{N}_0} := \{f^t\}_{t \in \mathbb{N}_0}$ . The positive orbit and positive limit set are correspondingly defined as

$$\begin{aligned} \Gamma^+(\phi_0) &:= \bigcup_{t \in \mathbb{N}_0} S(t)\phi_0 = \bigcup_{t \in \mathbb{N}_0} \phi_t, \\ \omega^+(\phi_0) &:= \{y \in X \mid \exists t_k \rightarrow \infty \text{ such that } S(t_k)\phi_0 \rightarrow y \text{ as } k \rightarrow \infty\}. \end{aligned}$$

This overloaded notation will enable the discussion of the error bounds in the discrete and continuous time cases have a common structure.

The Koopman operator for the discrete semiflow is defined by the operator  $U_f$  given by

$$U_f g := g \circ f$$

for each  $g \in H_X$ . Again, the Koopman operator  $U_f$  is a mapping from  $H_X$  to the pullback space  $f^*(H_X)$ . It is linear and bounded, and boundedness follows from exactly the same argument by which the orbital Koopman operator  $\mathcal{U}_\phi$  is shown to be bounded.

As in our discussion of the orbital Koopman operator  $\mathcal{U}_\phi$ , we use the same notation  $U_f$  to denote the operator defined by the identity  $U_f g := g \circ f$  for  $g \in R_\Omega(H_X)$ . Again, this definition makes sense as long as the orbit  $\Gamma^+(x_0)$  takes values in the subset  $\Omega \subset X$ . This definition is used in cases in which the samples  $\Xi$  do not fill the state space  $X$ , but are dense in a proper subset  $\Omega \subset X$ . By the same reasoning as outlined above, the operator norm  $\|U_f\| \leq 1$  as a mapping  $U_f : R_\Omega(H_X) \rightarrow f^*(R_\Omega(H_X))$ .

### 3.3 Samplings and Fill Distances

To begin our discussion of sampling, we treat the case that samples come from a single orbit  $\Gamma^+(\phi_0)$ . For either of the systems above characterized by the continuous semigroups  $\{S(t)\}_{t \in T}$  with either  $T = [0, \infty)$  or  $T = \mathbb{N}_0$ , we build approximations from a single orbit  $\Gamma^+(\phi_0)$  based on some set of *distinct* samples  $\Xi \subset X$  with

$$\Xi = \Xi(\phi_0) := \{\xi_i := \phi(t_i), t_i \in T\} \quad \text{or} \quad \{\xi_i := \phi_{k(i)} \mid i \in \mathbb{N}_0\}.$$

Note that this definition does not require that  $t_i \rightarrow \infty$  for the continuous time case, nor that the sampling is uniform in either case, although these will be the most common cases in applications. Approximations of functions and operators in the paper is achieved using the finite dimensional spaces  $H_{\Omega_n}$  defined as

$$H_{\Omega_n} := \text{span}\{\mathfrak{R}_{\xi_i} \mid \xi_i \in \Omega_n\} \subset H_X,$$

with

$$\Omega_n := \{\xi_i \in \Xi \mid 0 \leq i \leq n\}.$$

The sets  $\Omega_n$  are nested with  $\Omega_n \subset \Omega_{n+1}$ , and the approximant spaces are likewise nested,  $H_{\Omega_n} \subset H_{\Omega_{n+1}}$ . Correspondingly, for the continuous time case we also define the finite sets of discrete sample times  $T_n := \{t_i \in T \mid \phi(t_i) \in \Omega_n\}$ .

We denote by  $\Omega$  the limiting set that is approximated by the samples in the sense that  $\Xi$  is dense in  $\Omega$ , that is,  $\Xi \subseteq \Omega$  and  $\Omega \subseteq \overline{\Xi}$ . Clearly, one candidate for the limiting set is given by

$$\Omega := \overline{\bigcup_{n \in \mathbb{N}} \Omega_n} := \overline{\Xi},$$

but we also allow for open sets  $\Omega$ . The rate of convergence will ultimately be expressed in terms of the fill distance that is defined for any finite set  $\Omega_n \subset \Omega$  as

$$h_{\Omega_n, \Omega} := \sup_{x \in \Omega} \min_{\xi \in \Omega_n} d_X(x, \xi).$$

It should be noted that in order that the fill distance is finite, it must be the case that  $\Omega$  is bounded. In this paper when we say that the samples  $\Xi : \bigcup_{n \in \mathbb{N}} \Omega_n$  fill up the limiting set  $\Omega$ , we mean the fill distance approaches zero,  $d_{\Omega_n, \Omega} \rightarrow 0$  as  $n \rightarrow \infty$ .

In general an orbit  $\Gamma^+(x_0)$  of a semidynamical system need not have any easily or particularly notable structure that is readily characterized, and this fact will be reflected in the samples  $\Xi := \Xi(\phi_0)$  extracted from  $\Gamma^+(\phi_0)$ . Different error bounds for the Koopman operators in this paper will be derived depending on how much structure the set  $\Omega$  exhibits. The following list of examples illustrate some of the situations that may be encountered in practice.

*Samples  $\Xi$  are dense in the entire state space,  $\Omega := X$ :* In some sense this is one of the most basic situations that must be studied. Note that if we want  $h_{\Omega_n, \Omega} \rightarrow 0$  as  $n \rightarrow \infty$ , we must have  $\Omega$  bounded in this case, which puts limits on the choice of the state space  $X$ . The state space  $X$  cannot be a vector space, for instance, if we want the fill distance of finite sets to converge to zero. It makes sense then to focus on cases where the samples  $\Xi$  are dense in a compact state space and  $\Omega := X$ . One canonical system of this type is Hamiltonian flow on the torus  $\mathbb{T}^2$ , since it is well-known that the trajectory  $t \mapsto \phi(t)$  starting from any initial condition is dense in the torus. Since the torus  $\mathbb{T}^2$  is a compact submanifold of  $\mathbb{R}^2$ , it follows that, at least in principle, a collection of samples  $\Xi$  can be generated to fill the entire state space  $X = \Omega = \mathbb{T}^2$  in the sense that  $h_{\Omega_n, \Omega} \rightarrow 0$  as  $n \rightarrow \infty$ . Here we might be interested in the approximation of an observable function  $h : \mathbb{T}^2 \rightarrow \mathbb{R}$  defined over the torus. More generally, this case includes flows that are dense in a compact manifold  $M = X$ , and a typical task would be the approximation of an observable function  $h : M \rightarrow \mathbb{R}$  defined over the manifold.

*Samples are dense in an arbitrary subset  $\Omega \subset X$ :* Here, when we say arbitrary, we mean that  $\Omega$  does not exhibit any readily characterized structure. An example of this case can be when  $\Omega$  is an arc of the motion, that is, the range of the trajectory over some fixed time interval such as  $[a, b) \subset \mathbb{R}^+$  so that  $\Omega = \bigcup_{t \in [a, b)} \phi(t) \subset \Gamma^+(\phi_0)$ . In

Example 2 this could correspond to the approximation of the Koopman operator over some “transient regime” of the motion. The function spaces for characterizing rates of convergence of approximations in this case must make sense for such subsets. For this class of problems, approximation error is measured in the norm of a native space  $H_\Omega \subseteq H_X$  since the native space  $H_\Omega$  can be defined for any subset  $\Omega \subset X$ , even those that have zero measure as subsets of  $X := \mathbb{R}^d$ . Error bounds in this case are expressed in terms of the norm of the complementary projection operators  $\|(I - P_{\Omega_n})g\|_{H_\Omega}$ . If the rate at which  $\|I - P_{\Omega_n}\| \rightarrow 0$  is known, then we obtain rates of convergence of Koopman operator approximations.

*Samples dense in regular subsets  $\Omega \subseteq X$  or manifolds  $M \subseteq X$ :* The last category did not require assumptions regarding the properties of the limiting subset  $\Omega$ . However, when  $\Omega$  exhibits more structure, stronger conclusions regarding convergence can be derived. When  $\Omega$  is either a sufficiently regular subset  $\Omega \subset X$ , or even a submanifold  $M \subseteq X$ , we choose to express errors in terms of Sobolev norms that are equivalent to  $R_\Omega(H_X)$  in some cases. Here regularity refers to the regularity of the domain, and the precise notion of what is needed is described later in the paper. The error bounds derived in this case are in terms of the fill distance  $h_{\Omega_n, \Omega}$  of the sets  $\Omega_n$  in the limiting set  $\Omega$ .

#### 4 Approximations of the Koopman Operator $U_f$

In this section we carry out the details of the analysis of the errors in approximating the Koopman operators  $U_f : H_X \rightarrow H_X$  associated with the discrete dynamical system, and the orbital Koopman operator  $U_\phi$  for the continuous time dynamical system. We first study  $U_f$  in this section since this case seems most relevant to a host of recent papers that study approximation of such Koopman operators. We subsequently derive error bounds for the orbital Koopman operator in Section 5, but only give brief outlines of how the proofs in Section 4 must be modified for the continuous time case. In both cases, we build up and refine error estimates based on how much we know about the structure of the limiting set  $\Omega \subset \bar{X}$ .

Two general types of approximation are defined for the Koopman operator  $U_f$  in this section. First, we define  $U_f^n := U_f P_{\Omega_n}$  with

$$U_f^n g = (P_{\Omega_n} g) \circ f, \quad (17)$$

where it is assumed that  $f : X \rightarrow X$ . An explicit representation of this operator is given by

$$(U_f^n g)(x) = \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) g(\xi_i) \mathfrak{K}_{\xi_j}(f(x))$$

for each  $x \in X$  and  $g \in H_X$ , with  $\mathbb{K}^{-1}(\Omega_n)$  the inverse of the Grammian matrix  $\mathbb{K}(\Omega_n) := [\mathfrak{K}(\xi_m, \xi_n)]$  associated with the samples in  $\Omega_n$ . The above expression makes sense when the limiting set  $\Omega = X$ . The above definition also makes sense if  $g \in R_\Omega(H_X)$ ,  $f : \Omega \rightarrow \Omega$ , and  $t \mapsto \phi(t) \in \Omega \subseteq X$ , since in this case it is the coordinate expression for the data-driven approximation  $U_f^n := (P_{\Omega_n} g) \circ f$ . This fact follows since the kernel that generates  $R_\Omega(H_X)$  is  $r := \mathfrak{K}|_{\Omega \times \Omega}$ , and the Grammian matrix  $\mathbb{K}(\Omega_n) := [r(\xi_m, \xi_n)] = \mathbb{K}(\Omega_n)$ . This latter definition of  $U_f^n := P_{\Omega_n} g \circ f$  is used to study cases when the sample  $\Xi$  fill  $\Omega$ , but do not fill all of  $X$ . As discussed in Section 2.2, in this expression, the projections  $P_{\Omega_n}$  are the orthogonal projections of  $R_\Omega(H_X)$  onto  $R_\Omega(H_{\Omega_n}) := \text{span}\{r_{\xi_i} \mid \xi_i \in \Omega_n\}$ .

Note that  $U_f^n$  can be applied to a function  $g$  only if  $f$  is known in closed form in either of the above definitions of  $U_f^n$ , so these approximations are not realizable in applications that seek data-driven approximations. Still, bounds on the error in approximating  $U_f$  by  $U_f^n$  are of theoretical interest in their own right. They tell us upper bounds on the approximation rates that can be achieved from a given finite dimensional subspace  $H_{\Omega_n}$ , if the function  $f$  is known. In this sense they describe limits on what is feasible or achievable in principle when using the subspaces  $H_{\Omega_n}$ . Pragmatically speaking, these bounds are also directly applicable in two specific, relevant problems. These bounds can be used to derive rates of convergence of estimates for the reproducing kernel Hilbert (RKH) space embedding method, see Section III of [68]. They are also important in that they are used to construct some error bounds for the data-driven approximations that follow.

For constructing data-dependent approximations of the Koopman operator, we discuss two different types of approximations, each of which makes a different type of structural assumption about the nature of the Koopman operator. In the first data-driven method, we assume (1) that the samples  $\Xi$  are dense in a limiting set  $M := X$  that is a type of smooth manifold, and we also require that (2) the Koopman operator  $U_f$  is such that the pullback space  $f^*(H_X) \subset H_X$ . In this case we can define the approximation  $\mathfrak{U}_f^n := P_{\Omega_n}((P_{\Omega_n} g) \circ f)$ . Since by definition  $P_{\Omega_n} : H_X \rightarrow H_X$ , this definition makes sense as long as  $f^*(H_X) \subset H_X$  because  $(P_{\Omega_n} g) \circ f \in f^*(H_X)$  is then an element of  $H_X$ . As we explain in Example 6, if  $g := \sum_{1 \leq j \leq n} c_j \mathfrak{K}_{\xi_j}$ , the operator  $\mathfrak{U}_f^n$  has the coordinate representation

$$\mathfrak{U}_f^n g := \sum_{i,j,m} c_j \mathfrak{K}_{\xi_j}(y_i) \mathbb{K}_{m,i}^{-1}(\Omega_n) k_{\xi_m}. \quad (18)$$

In fact, Example 6 shows that the operator above, which acts on functions, is a special case of that induced in the usual derivation of the EDMD algorithm in references such as [21, 31, 41, 42]

The definition of the approximation  $\mathfrak{U}_f^n$  makes sense provided that the pullback space  $f^*(H_X) \subseteq H_X$ . As useful as this assumption is, it may not be a simple condition to establish or verify. This fact motivates our second method, which is intended to be applicable in some cases when it is not clear that  $f^*(H_X) \subseteq H_X$ . In the second method, we assume that (1) the samples  $\Xi$  are dense in a limiting set  $\Omega$  that is a sufficiently regular, proper subset of  $X := \mathbb{R}^d$ , and (2) the unknown function  $f \in H_X^d$ . In contrast to the approximations  $\mathfrak{U}_f^n$  defined above that make structural assumptions on the composite function  $U_f g := g \circ f$ , in this case we make hypotheses directly about the function  $f$ . Data-driven approximations  $\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n$  of the Koopman operator are generated from expressions like

$$\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n := U_{\mathbb{P}_{\Omega_n} f} \mathbb{P}_{\Omega_n} g := (\mathbb{P}_{\Omega_n} g) \circ (\mathbb{P}_{\Omega_n} f).$$

For completeness, we record the explicit representations

$$\mathbb{P}_{\Omega_n} f := \{\mathbb{P}_{\Omega_n} f_1, \dots, \mathbb{P}_{\Omega_n} f_d\}^T := \begin{Bmatrix} \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) f_1(\xi_i) \mathfrak{K}_{\xi_j}(\cdot) \\ \vdots \\ \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) f_d(\xi_i) \mathfrak{K}_{\xi_j}(\cdot) \end{Bmatrix}$$

where the samples  $\Omega_n := \{\xi_i\}_{1 \leq i \leq n} = \{\phi_i\}_{1 \leq i \leq n}$ . Note also that since by assumption the discrete evolution is given by  $y_i := \phi_{i+1} = f(\phi_i)$ , the data-dependent approximation  $\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n := U_{\mathbb{P}_{\Omega_n} f} \mathbb{P}_{\Omega_n} g$  can be formed from the pairs of samples  $\{(\xi_i, y_i)\}_{1 \leq i \leq n} = \{\xi_i, f(\xi_i)\}_{1 \leq i \leq n}$  via

$$(\mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n g)(x) = \sum_{1 \leq \ell, m \leq n} \mathbb{K}_{m,\ell}^{-1}(\Omega_n) g(\xi_\ell) \mathfrak{K}_{\xi_m} \left( \begin{Bmatrix} \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) f_1(\xi_i) \mathfrak{K}_{\xi_j}(\cdot) \\ \vdots \\ \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) f_d(\xi_i) \mathfrak{K}_{\xi_j}(\cdot) \end{Bmatrix} \right).$$

At the end of this section we combine the error bounds for  $U_f^n$  and  $U_{\mathbb{P}_{\Omega_n} f}^n$  to generate overall error estimates for the data-driven approximations of the Koopman operators. We build bounds on rates of convergence using the triangle inequalities that have the form

$$\|U_f g - \mathfrak{U}_f^n g\|_B \leq \|U_f g - U_f^n g\|_B + \|U_f^n g - \mathfrak{U}_f^n g\|_B$$

or

$$\|U_f g - \mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n g\|_B \leq \|U_f g - U_f^n g\|_B + \|U_f^n g - \mathfrak{U}_{\mathbb{P}_{\Omega_n} f}^n g\|_B$$

with  $B$  a Banach space of functions.

#### 4.1 The Approximations $U_f^n$

We start with some very straightforward results, ones that are not too surprising and hold without any particular assumptions about the structure of the limiting set  $\Omega$ .

##### 4.1.1 Approximations $U_f^n$ , Arbitrary $\Omega \subseteq X$

The following theorem bounds the error in approximations  $U_f^n$  of  $U_f$  in terms of the norms of the complementary projections  $I - P_{\Omega_n}$ . In Theorem 2 that follows we give sufficient conditions to conclude the pointwise convergence  $(U_f^n g)(x) \rightarrow (U_f g)(x)$  for  $x \in X$  for  $g \in H_\Omega \subseteq H_X$ .

**Theorem 1** *Suppose that the samples  $\Xi$  are dense in a bounded set  $\Omega \subseteq X$ , there is a constant  $\bar{k} > 0$  such that the kernel  $\mathfrak{K}$  that defines the native space  $H_X$  satisfies  $\sup_{x \in X} \mathfrak{K}(x, x) \leq \bar{k}^2$ , and let  $f : X \rightarrow X$ . Then we have the error bound*

$$\|U_f g - U_f^n g\|_{f^*(H_X)} \leq \|(I - P_{\Omega_n})g\|_{H_X} \quad (19)$$

for all  $g \in H_X$ . If furthermore the mapping  $f \in C(X, X)$  and  $g \in H_\Omega$ , we have the pointwise estimate

$$|U_f g(x) - U_f^n g(x)| \leq \|(I - P_{\Omega_n})g\|_{C(X)} \leq \|(I - P_{\Omega_n})g\|_{H_\Omega} \quad (20)$$

for each  $x \in X$ .

*Proof* We want to derive a bound for the norm of the difference

$$\|U_f g - U_f^n g\|_{f^*(H_X)} = \|U_f(I - P_{\Omega_n})g\|_{f^*(H_X)}.$$

But by definition we have

$$\|U_f h\|_{f^*(H_X)} = \|h \circ f\|_{f^*(H_X)} = \inf\{\|q\|_{H_X} \mid h \circ f = q \circ f\} \leq \|h\|_{H_X},$$

and we see that the operator norm  $\|U_f\| \leq 1$ . From this line Equation 19 follows. By definition, the kernel on  $f^*(H_X)$  is given by  $\mathfrak{K}_f(x, y) := \mathfrak{K}(f(x), f(y))$ . Since  $\sup_x \mathfrak{K}_f(x, x) \leq \bar{k}^2$  and  $\mathfrak{K}_f$  is continuous, we know that  $f^*(H_X) \hookrightarrow C(X)$ . We can then write

$$\begin{aligned} |U_f g(x) - U_{f,n} g(x)| &= |(g - P_{\Omega_n})(f(x))| \\ &\leq \|(I - P_{\Omega_n})g\|_{C(X)} \leq \|(I - P_{\Omega_n})g\|_{H_X} \end{aligned}$$

for each  $x \in X$ . Additionally, if  $g \in H_\Omega$  we have

$$\begin{aligned} \|(I - P_{\Omega_n})g\|_{C(X)} &\leq \|(I - P_{\Omega_n})g\|_{H_X} = \|P_\Omega(I - P_{\Omega_n})g\|_{H_X} \\ &= \|\mathbb{E}_X R_\Omega(I - P_{\Omega_n})g\|_{H_X} = \|(I - \mathbb{P}_{\Omega_n})R_\Omega g\|_{R_\Omega(H_X)} \end{aligned}$$

The norm  $\|(I - \mathbb{P}_{\Omega_n})R_\Omega g\|_{R_\Omega(H_X)} = \|(I - P_{\Omega_n})g\|_{H_\Omega} \rightarrow 0$  as  $n \rightarrow \infty$  provided that we can show  $\{P_{\Omega_n}\}_{n \in \mathbb{N}}$  converge to the identity operator on  $H_\Omega$ , or equivalently,  $\{\mathbb{P}_{\Omega_n}\}_{n \in \mathbb{N}}$  converge to the identity operator on  $R_\Omega(H_X)$ . This is the topic of the next proposition and theorem.

So far, we have established that the error in the approximation of the Koopman operator can be bounded by  $\|(I - P_{\Omega_n})g\|_{H_\Omega} = \|(I - \mathbb{P}_{\Omega_n})R_\Omega g\|_{R_\Omega(H_X)}$  for  $g \in H_\Omega \subseteq H_X$ . In many of the theorems of the next few sections, precise bounds for  $\|(I - P_{\Omega_n})g\|_{H_\Omega}$  are derived depending on the smoothness properties of the kernel  $\mathfrak{K}$  that defines  $H_X$  and the regularity of the limiting set  $\Omega$ .

In the situation at hand, it is assumed that the samples  $\Xi = \bigcup_{n \in \mathbb{N}} \Omega_n$  are dense in  $\Omega$ . But with so little known about the set  $\Omega$ , the kernel  $\mathfrak{K}$ , and the RKH space  $H_X$ , as of yet we do not even know that the set of functions  $\{\mathfrak{K}_x \mid x \in \Xi\}$  is dense in  $H_\Omega := \overline{\text{span}\{\mathfrak{K}_x \mid x \in \Omega\}}$ . Proving this fact is equivalent to proving that the operator norm  $\|I - P_{\Omega_n}\| \rightarrow 0$  when here we interpret  $P_{\Omega_n} := P_{\Omega_n}|_{H_\Omega}$  as an operator on  $H_\Omega$ . Establishing this fact is carried out using the following proposition, which is implied by Propositions 1 and 2 of [58].

**Proposition 1** *Suppose that the admissible kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  induces the RKH space  $H_X$  and that the map  $x \mapsto \mathfrak{K}_x \in H_X$  is injective. Then  $d_{\mathfrak{K}} : X \times X \rightarrow \mathbb{R}$  given by  $d_{\mathfrak{K}}(x, y) := \|\mathfrak{K}_x - \mathfrak{K}_y\|_{H_X}$  is a metric on  $X$ , and the map  $x \mapsto \mathfrak{K}_x$  is an isometry from  $(X, d_X)$  into  $H_X$ .*

*We say that the native space  $H_X$  separates a subset  $S \subset X$  if for each  $x \notin S$  there is a function  $g \in H_X$  such that  $g(x) \neq 0$  and  $g(y) = 0$  for all  $y \in S$ . If  $H_X$  separates all the  $d_X$ -closed subsets of  $X$ , then  $(X, d_{\mathfrak{K}})$  and  $(X, d_X)$  are equivalent metric spaces.*

This theorem suffices to establish that  $\{P_{\Omega_n}\}_{n \in \mathbb{N}}$  converges to the identity operator on  $H_\Omega$ , as summarized in the following theorem.

**Theorem 2** *Suppose that the kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  satisfies the hypotheses of Proposition 1. Then  $\{P_{\Omega_n}\}_{n \in \mathbb{N}}$  converges to the identity operator on  $H_\Omega$ .*

*Proof* The theorem follows if we can establish that  $H_\Omega := \overline{\text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Xi\}}$ . Since the metric spaces  $(X, d_X)$  and  $(X, d_{\mathfrak{K}})$  are assumed to be equivalent, there are two positive constants  $c_1, c_2$  such that

$$c_1 d_X(x, y) \leq d_{\mathfrak{K}}(x, y) \leq c_2 d_X(x, y)$$

for all  $x, y \in X$ . Suppose that  $f \in H_\Omega$  and fix an arbitrary  $\epsilon > 0$ . By the definition of  $H_\Omega$ , there is an integer  $n = n(\epsilon)$ , real numbers  $\alpha_{n,i} \in \mathbb{R}$ , and locations  $x_{n,i} \in \Omega$  such that

$$\left\| f - \sum_{1 \leq i \leq n} \alpha_{n,i} \mathfrak{K}_{x_{n,i}} \right\|_{H_X} < \epsilon/2.$$

But since  $\Xi$  is dense in  $\Omega$ , for each  $x_{n,i}$  we can find an  $\xi_{n,\ell_i} \in \Xi$  such that

$$d_X(\xi_{n,\ell_i}, x_{n,i}) \leq \frac{1}{2c_2 n \|\{\alpha_{n,i}\}\|_\infty} \epsilon.$$

We then have

$$\begin{aligned}
\left\| f - \sum_{1 \leq i \leq n} \alpha_{n,i} \mathfrak{K}_{n,\ell_i} \right\|_{H_X} &\leq \left\| f - \sum_{1 \leq i \leq n} \alpha_{n,i} \mathfrak{K}_{x_{n,i}} \right\|_{H_X} + \left\| \sum_{1 \leq i \leq n} \alpha_{n,i} (\mathfrak{K}_{x_{n,i}} - \mathfrak{K}_{\xi_{n,\ell_i}}) \right\|_{H_X} \\
&\leq \epsilon/2 + \|\{\alpha_{n,i}\}\|_\infty \sum_{1 \leq i \leq n} \|\mathfrak{K}_{x_{n,i}} - \mathfrak{K}_{\xi_{n,\ell_i}}\|_{H_X} \\
&= \epsilon/2 + \|\{\alpha_{n,i}\}\|_\infty \sum_{1 \leq i \leq n} d_{\mathfrak{K}}(x_{n,i}, \xi_{n,\ell_i}) \\
&\leq \epsilon/2 + c_2 \|\{\alpha_{n,i}\}\|_\infty \sum_{1 \leq i \leq n} d_X(x_{n,i}, \xi_{n,\ell_i}) \leq \epsilon.
\end{aligned}$$

This means that  $\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Xi\}$  is dense in  $H_\Omega$ , and the proof is complete.

#### 4.1.2 Approximations $U_f^n$ , $\Omega := M$ a Manifold

As noted above, the RKH space  $f^*(H_X)$  is defined for *any* function  $f : X \rightarrow X$ , and Theorem 2 holds even when and the limiting set  $\Omega$  does not exhibit any particular structure. We next discuss a cases when the limiting set  $\Omega$  exhibits some additional, but specific, kinds of structure. We begin with a discussion of the case in which the limiting set  $\Omega$  is a smooth, connected, compact Riemannian manifold  $M := \Omega$ , and the samples are dense in  $M$ . Three types of theorems are derived in this section below: the first uses the many zeros Theorem 11 on manifolds, while the latter two theorems are examples of error bounds that generate pointwise estimates from Theorems 12 and 13. While there are differences in the fine details of the hypotheses for these two theorems, it should be noted that all three theorems are quite similar qualitatively speaking. In all cases the errors in approximations of the Koopman operator are bounded by powers of the fill distance  $h_{\Omega_n, M}^r$ , and the exponent  $r$  depends on the smoothness of the functions in certain Sobolev spaces. Note also that all three of these theorems can be applied when the samples fill the entire state space  $X$ , that is, when  $M := \Omega := X$ . We begin with a result that depends on the many zeros Theorem 11 in a manifold  $M$ .

**Theorem 3** *Suppose that  $M$  is a  $d$ -dimensional, connected, compact, Riemannian manifold without boundary, let  $\mathfrak{K} : M \times M \rightarrow \mathbb{R}$  be a positive definite kernel that induces a native space  $H_M$ , and suppose that  $H_M$  is equivalent to the Sobolev space  $W^{t,2}(M)$  for some  $t \in \mathbb{R}$  that satisfies  $d/2 < s \leq \lceil t \rceil - 1$  for a given  $s \in \mathbb{N}$ . Then there are constants  $C_M, h_M > 0$  such that for all  $\Omega_n \subset \Omega$  that satisfy  $h_{\Omega_n, \Omega} \leq h_M$  we have*

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(M))} \leq C_M h_{\Omega_n, M}^{t-s} \|g\|_{W^{t,2}(M)}$$

for  $g \in W^{t,2}(M)$ .

*Proof* Since  $s > d/2$ , the Sobolev embedding theorem implies that  $W^{s,2}(M)$  is a RKH space, and therefore the pullback space  $f^*(W^{s,2}(M))$  is well-defined. By the definition of the pullback space we have

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(M))} \leq \|U_f\| \|(I - P_{\Omega_n})g\|_{W^{s,2}(M)}.$$

As discussed in Section 2.2, the operator norm of  $U_f$ , as a mapping from the RKH space  $W^{s,2}(M)$  into  $f^*(W^{s,2}(M))$ , is less than or equal to one. But on  $\Omega_n$ , we know that  $((I - P_{\Omega_n})g)|_{\Omega_n} = 0$  since the projection is identical to the interpolant over  $\Omega_n$ . By the many zeros Theorem 11, we conclude that

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(M))} \leq C_M h_{\Omega_n, M}^{t-s} \|g\|_{W^{t,2}(M)}.$$

Now we turn to a pair of theorems that ensure pointwise convergence rates over manifolds.

**Theorem 4** *Suppose that the collection of samples  $\Xi$  is dense in a compact, connected,  $C^\ell$  Riemannian manifold  $M := \Omega$  without boundary, the kernel  $\mathfrak{K}$  that induces  $H_M$  is positive definite, the kernel  $\mathfrak{K} \in C^{2k}(M \times M)$ , and  $\ell \geq 2k$ . Then there exist constants  $h_M, C_M > 0$  such that for all  $g \in H_M$  and  $\Omega_n \subset M$  with  $h_{\Omega_n, M} < h_M$  we have*

$$|(U_f g)(x) - (U_f^n g)(x)| \leq C_M h_{\Omega_n, M}^k \|g\|_{H_M}$$

for all  $x \in M$ .

*Proof* By definition we have

$$\begin{aligned}
|(U_f g)(x) - (U_f^n g)(x)| &:= |g(f(x)) - (P_{\Omega_n} g)(f(x))| \\
&\leq \sup_{\eta \in M} |g(\eta) - (P_{\Omega_n} g)(\eta)| \\
&\leq C_M h_{\Omega_n, M}^k \|g\|_{H_M}
\end{aligned}$$

for each  $x \in M$ . The last line in the above sequence of inequalities follows from the approximation Theorem 13.



**Theorem 5** *Suppose that the collection of samples  $\Xi$  is dense in a  $d$ -dimensional, compact, connected, smooth Riemannian manifold  $M := \Omega$  without boundary, the kernel  $\mathfrak{K}$  that induces  $H_M$  is positive definite, and  $H_M$  is equivalent to a Sobolev space  $W^{t,2}(M)$  for some  $t > d/2$ . Then there exist constants  $h_M, C_M > 0$  such that for all  $g \in H_M$  and  $\Omega_n \subset M$  with  $h_{\Omega_n, M} < h_M$  we have*

$$|(U_f g)(x) - (U_f^n g)(x)| \leq C_M h_{\Omega_n, M}^{t-d/2} \|g\|_{H_M}$$

for all  $x \in M$ .

*Proof* The proof of this theorem proceeds exactly as in the last Theorem 4, but by replacing the last line of the proof to use Theorem 12 instead of Theorem 13.

#### 4.1.3 Approximations $U_f^n$ , $\Omega \subset X := \mathbb{R}^d$

The last two theorems are designed to apply to cases when the samples  $\Xi$  fill up the entire manifold  $M := \Omega$  in the sense that  $d_{\Omega_n, M} \rightarrow 0$  as  $n \rightarrow \infty$ . We now consider cases when the samples  $\Xi$  are dense in a limiting set  $\Omega$  that is a proper subset of the state space  $X := \mathbb{R}^d$ .

**Theorem 6** *Suppose that the collection of samples  $\Xi$  is dense in the compact subset  $\Omega \subset X := \mathbb{R}^d$  that has a Lipschitz boundary, the kernel  $\mathfrak{K}$  induces the RKH space  $H_X$  of functions over  $X$ , there is a constant  $\bar{k}$  such that  $\sup_{x \in X} \mathfrak{K}(x, x) \leq \bar{k}^2$ , the integer  $s \in \mathbb{N}_0$  with  $d/2 \leq s \leq \lceil t \rceil - 1$ , and  $R_\Omega(H_X) \approx W^{t,2}(\Omega)$ . Then there exists constants  $h_M, C_M > 0$  such that for all  $g \in W^{t,2}(\Omega)$  and subsets  $\Omega_n$  with  $h_{\Omega_n, \Omega} \leq h_M$  we have*

$$\|U_f g - U_f^n g\|_{f^*(W^{s,2}(\Omega))} \leq C_\Omega h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)}. \quad (21)$$

If, furthermore,  $f \in C(\Omega, \Omega)$ , we have the pointwise estimate

$$|(U_f g)(x) - (U_f^n g)(x)| \leq \|U_f g - U_f^n g\|_{f^*(W^{s,2}(\Omega))}$$

for each  $x \in \Omega$ .

*Proof* From the definition of the pullback space  $f^*(W^{s,2}(\Omega))$  and the many zeros Theorem 10, we see that

$$\begin{aligned} \|U_f g - U_f^n g\|_{f^*(W^{s,2}(\Omega))} &\leq \|U_f\| \|(I - \mathbb{P}_{\Omega_n})g\|_{W^{s,2}(\Omega)} \\ &\leq \|U_f\| C_\Omega h_{\Omega_n, \Omega}^{t-s} \|(I - \mathbb{P}_{\Omega_n})g\|_{W^{t,2}(\Omega)}. \end{aligned}$$

The many zeros Theorem 10 can be applied since the projection  $\mathbb{P}_{\Omega_n} g$  is identical to the interpolant of  $g$  on  $\Omega_n$ , so that  $(I - \mathbb{P}_{\Omega_n})g|_{\Omega_n} = 0$ . Since  $\|U_f\| \leq 1$  and  $\|I - \mathbb{P}_{\Omega_n}\| \leq 1$ , the inequality in Equation 21 follows. The estimate above holds for any function  $f : \Omega \rightarrow \Omega$ . If, in addition,  $f$  is continuous, then we have

$$\mathfrak{K}_f(x, y) := \mathfrak{K}(f(x), f(y)) \leq \bar{k}^2.$$

Since  $\mathfrak{K}_f$  is the kernel of the RKH space  $R_\Omega(H_X)$ , this is sufficient to ensure that  $f^*(W^{s,2}(\Omega)) \hookrightarrow C(\Omega)$ . We then have

$$|(U_f g)(x) - (U_f^n g)(x)| \leq \|U_f g - U_f^n g\|_{C(\Omega)} \lesssim \|U_f g - U_f^n g\|_{f^*(W^{s,2}(\Omega))},$$

and the proof is complete.

Before proceeding to our discussion of data-dependent approximations, a few observations are in order to emphasize some of the nuances that distinguish Theorem 6 from its predecessors Theorems 4 and 5. The most obvious difference is, of course, that in Theorems 4 and 5 the samples are dense in a compact Riemannian manifold  $M$ , which allows for the possibility that the entire state space  $X$  of the underlying flow is a compact manifold  $M := X := \Omega$ . Note that the bounds on the pointwise error in this case are specified relative to the norm  $\|g\|_{H_M}$ , and the operator  $U_f$  defined as  $U_f g := g \circ f$  acts on functions  $g$  that are defined over all of the state space  $X := M = \Omega$ . On the other hand, in Theorem 6 the samples are dense in the proper subset  $\Omega \subset X := \mathbb{R}^d$ . The norm  $\|\cdot\|_{f^*(W^{s,2}(\Omega))}$  is defined for functions  $h := g \circ f$  with  $g \in W^{s,2}(\Omega)$  and  $f : \Omega \rightarrow \Omega$ . In other words the operator  $U_f$  in this theorem is understood as an operator  $U_f g := g \circ f$  for  $g \in W^{s,2}(\Omega)$ . Note that in this interpretation the operator  $U_f$  is an operator on functions supported on the proper subset  $\Omega \subset X := \mathbb{R}^d$ , in contrast to the operator  $U_f$  in Theorems 4 or 5 that acts on functions supported on the full state space  $X := M$ .

In writing the original discrete evolution law in Equation 1, we have stipulated only that  $f : X \rightarrow X$ . Connecting the error estimates in Theorem 6 to the original evolution law would require some additional assumptions about the underlying flow and the limiting set  $\Omega \subset X := \mathbb{R}^d$ . For instance, suppose that we know that  $\Omega$  is a positive invariant set for the flow, meaning that  $\phi_n \in \Omega$  implies that  $\phi_{n+1} \in \Omega$ . Under this assumption it must be the case that  $f(\Omega) \subseteq \Omega$ , and the restricted function  $f|_\Omega := R_\Omega f$  is a mapping  $f|_\Omega : \Omega \rightarrow \Omega$ . In this situation Theorem 6 can be understood as a statement about the approximation of the Koopman operator associated with the restriction  $f|_\Omega$  from the samples in  $\Xi$  that are dense in  $\Omega$ .

## 4.2 Data-Dependent Approximations $\mathfrak{U}_f^n$ and $\mathfrak{U}_{P_{\Omega_n} f}^n$

In this section we explore how the projection approximations  $U_f^n g := (P_{\Omega_n} g) \circ f$  introduced above are used to construct error estimates for the data-dependent approximations  $\mathfrak{U}_f^n$  and  $\mathfrak{U}_{P_{\Omega_n} f}^n$ .

### 4.2.1 Data-Dependent Approximations $\mathfrak{U}_f^n$ on a manifold $M$

*Example 6* In this example we review the EDMD method since it is one of the most popular and well-known algorithms that is studied in the context of Koopman theory. We will see that the data-driven approximation  $\mathfrak{U}_f^n g := P_{\Omega_n}((P_{\Omega_n} g) \circ f)$ , which is defined in terms of projections  $P_{\Omega_n}$  over the finite dimensional spaces  $H_{\Omega_n}$  in an RKH space, has a coordinate representation that is a special case of that defined in terms of the classical EDMD algorithm. In the last few years, the notation used to present the EDMD algorithm has evolved into a kind of common convention (see for example [21,31,41,42]) and we review in this section this notation. Suppose we have collected the  $m$  samples  $\{\phi_i, y_i\}_{1 \leq i \leq m}$  of the discrete dynamics described in the canonical Equation 1. The output is defined as  $y_i := f(\phi_i)$ . We define the matrices  $\Phi$  and  $Y$  of inputs and outputs, respectively, as

$$\begin{aligned}\Phi &:= [\phi_1 \ \phi_2 \ \cdots \ \phi_m] \in \mathbb{R}^{d \times m}, \\ Y &:= [y_1 \ y_2 \ \cdots \ y_m] \in \mathbb{R}^{d \times m}.\end{aligned}$$

In the notation that has become standard in descriptions of the EDMD algorithm, we define the  $i^{\text{th}}$  basis function  $\psi_i := \mathfrak{K}_{\xi} \in \mathcal{F}_n := \text{span}\{\psi_i \mid 1 \leq i \leq n\}$ , introduce the vector of basis functions as  $\psi(x) := \{\psi_1(x), \dots, \psi_n(x)\}^T \in \mathbb{R}^{n \times 1}$ , and set the data matrices

$$\begin{aligned}\Psi(\Phi) &:= [\psi(\phi_1) \ \psi(\phi_2) \ \cdots \ \psi(\phi_m)] = [\psi_i(\phi_j)] \in \mathbb{R}^{n \times m}, \\ \Psi(Y) &:= [\psi(y_1) \ \psi(y_2) \ \cdots \ \psi(y_m)] = [\psi_i(y_j)] \in \mathbb{R}^{n \times m}.\end{aligned}$$

The EDMD algorithm, viewed in terms of linear algebra, solves for the matrix  $A_{n,m} \in \mathbb{R}^{n \times n}$  that satisfies the minimization problem

$$\begin{aligned}A_{n,m} &:= \underset{A \in \mathbb{R}^{n \times n}}{\text{argmin}} \|A\Psi(\Phi) - \Psi(Y)\|_F^2 \\ &= \underset{A \in \mathbb{R}^{n \times n}}{\text{argmin}} \sum_{1 \leq i \leq m} \|A\psi(\phi_i) - \psi(y_i)\|_{\mathbb{R}^n}^2\end{aligned}$$

with  $\|\cdot\|_F$  the Frobenius norm on matrices. The solution of this minimization problem is given by the matrix  $A_{n,m} := \Psi(Y)\Psi^+(\Phi)$  with  $\Psi^+(\Phi)$  the Moore-Penrose pseudoinverse of the matrix  $\Phi$ . Finally, the EDMD algorithm defines a data-driven approximation  $\mathfrak{U}_{n,m}$  of the Koopman operator  $U_f$  that is given by

$$(\mathfrak{U}_{n,m} g)(x) := c^T A_{n,m} \psi(x) := \sum_{i,j,\ell} c_i \Psi_{i,\ell}(Y) \Psi_{\ell,m}^+(\Phi) \psi_m(x)$$

for each  $g := \sum_{1 \leq i \leq n} c_i \psi_i \in H_{\Omega_n}$ . It is not difficult to relate the data-driven approximation  $\mathfrak{U}_f^n g := P_{\Omega_n}((P_{\Omega_n} g) \circ f)$  to  $\mathfrak{U}_{n,m}$ . By definition  $\mathfrak{U}_f^n : H_X \rightarrow H_{\Omega_n}$ , and  $\mathfrak{U}_{n,m} : H_{\Omega_n} \rightarrow H_{\Omega_n}$ , and since these two operators have different domains, they cannot be the same operator. However, in some circumstances, we do have

$$\mathfrak{U}_{n,m} := \mathfrak{U}_f^n \Big|_{H_{\Omega_n}}.$$

To establish this relation, we choose the finite dimensional space  $\mathcal{F}_n := H_{\Omega_n}$ , the basis  $\psi_i := \mathfrak{K}_{\xi_i}$ , for  $\xi_i \in \Omega_n$  with  $1 \leq i \leq n$ , and set the number of samples  $m$  equal to the dimension of the approximant space  $H_{\Omega_n}$ ,  $m = n$ . The equivalence is immediate when we note that

$$\begin{aligned}\Psi(\Phi) &:= [\psi_i(\phi_j)] = [\mathfrak{K}(\xi_i, \xi_j)] = \mathbb{K}(\Omega_n) \\ \Psi(Y) &:= [\psi_i(y_j)] = [\mathfrak{K}_{\xi_i}(f(\xi_j))] = [\mathfrak{K}(\xi_i, y_j)],\end{aligned}$$

and compare the coordinate expression above for  $\mathfrak{U}_{n,m}$  to that of  $\mathfrak{U}_f^n$  in equation 18. In view of these relationships, we can view one contribution of this paper as defining a specific choice of the bases in the EDMD algorithm that enables precise rates of convergence over some important cases of regular subsets and smooth manifolds. These error estimates are given in terms of the fill distance of samples in the manifold. This result has no precedent in either the study of approximations of Koopman operators or the EDMD algorithm.

**Theorem 7** *Suppose that the hypotheses of Theorem 11 hold. Furthermore, suppose that the mapping  $f : M \rightarrow M$  is such that the pullback space  $f^*(H_M) \subseteq H_M$ . Then there are constants  $C_M, h_M > 0$  such that for all  $\Omega_n \subset \Omega$  that satisfy  $h_{\Omega_n, \Omega} \leq h_M$  we have the pointwise bound*

$$|(U_f g)(x) - (\mathfrak{U}_f^n g)(x)| \leq C_M h_{\Omega_n, M}^{t-s} (\|g\|_{W^{t,2}(M)} + \|U_f g\|_{W^{t,2}(M)})$$

for  $g \in W^{t,2}(M)$  and  $x \in M$ .

*Proof* Since  $f^*(H_M) \subset H_M$  we know that  $P_{\Omega_n} g \in H_M$  and  $P_{\Omega_n} g \circ f \in f^*(H_M) \subset H_M$ . By definition, we have

$$\begin{aligned} |(U_f g)(x) - (\mathfrak{U}_f^n g)(x)| &:= |(g \circ f)(x) - (P_{\Omega_n}((P_{\Omega_n} g) \circ f))(x)| \\ &\leq |(g \circ f)(x) - (P_{\Omega_n} g) \circ f(x)| + |((I - P_{\Omega_n})(P_{\Omega_n} g \circ f))(x)|. \end{aligned}$$

Under the hypotheses of this theorem, we have

$$H_M \approx W^{t,2}(M) \hookrightarrow W^{s,2}(M) \hookrightarrow C(M).$$

This means that

$$\begin{aligned} |(U_f g)(x) - (\mathfrak{U}_f^n g)(x)| &\leq \sup_{\eta \in M} |((g - P_{\Omega_n} g) \circ f)(x)| + \sup_{\eta \in M} |((I - P_{\Omega_n})(P_{\Omega_n} g \circ f))(\eta)| \\ &\leq \|(I - P_{\Omega_n})g\|_{C(M)} + \|(I - P_{\Omega_n})(P_{\Omega_n} g \circ f)\|_{C(M)} \\ &\leq \|(I - P_{\Omega_n})g\|_{C(M)} + \|(I - P_{\Omega_n})(P_{\Omega_n} g \circ f)\|_{W^{s,2}(M)} \end{aligned}$$

Now we apply the many zeros Theorem 3 to each of the right hand side terms above. We know that

$$\begin{aligned} (I - P_{\Omega_n})g|_{\Omega_n} &= 0 \\ (I - P_{\Omega_n})(P_{\Omega_n} g \circ f)|_{\Omega_n} &= 0 \end{aligned}$$

since the projection operator  $P_{\Omega_n}$  is identical to the interpolation operator on  $\Omega_n \subset M$ . By the many zeros theorem on the manifold  $M$ , we can write

$$\begin{aligned} |(U_f g)(x) - (\mathfrak{U}_f^n g)(x)| &\leq C_1 h_{\Omega_n, M}^{t-s} \|g\|_{W^{t,2}(M)} + C_2 h_{\Omega_n, M} \|P_{\Omega_n} g \circ f\|_{W^{t,2}(M)} \\ &\leq \max\{C_1, C_2\} h_{\Omega_n, M}^{t-s} (\|g\|_{W^{t,2}(M)} + \|U_f g\|_{W^{t,2}(M)}). \end{aligned}$$

#### 4.2.2 Data-Dependent Approximations $\mathfrak{U}_{P_{\Omega_n} f}^n$ on $\Omega \subset X := \mathbb{R}^d$

We next turn to the study of *data-dependent* approximations

$$(\mathfrak{U}_{P_{\Omega_n} f}^n g)(x) := (\mathbb{P}_{\Omega_n} g) \circ (\mathbb{P}_{\Omega_n} f)(x)$$

for each  $x \in \Omega \subset X := \mathbb{R}^d$ . We focus in this section on the case when the samples  $\Xi$  are dense in a sufficiently regular set  $\Omega \subset X := \mathbb{R}^d$ . We will show that effective error bounds on the convergence rate are derived if, in addition to the hypotheses outlined in Theorem 6, we assume additionally that the RKH space  $R_\Omega(H_X)$  is contained in the space of Lipschitz continuous functions.

**Theorem 8** *Suppose that hypotheses of Theorem 6 hold, additionally that we have the continuous embedding  $R_\Omega(H_X) \approx W^{t,2}(\Omega) \subseteq W^{s,2}(\Omega) \hookrightarrow C^{0,1}(\Omega)$ , and  $f \in (W^{t,2}(\Omega))^d$ . Then there are constants  $h_\Omega, C_\Omega > 0$  such that for all  $g \in W^{t,2}(\Omega)$  and  $\Omega_n \subset \Xi$  such that  $h_{\Omega_n, \Omega} < h_\Omega$  we have*

$$\left| (U_f^n g)(x) - (\mathfrak{U}_{P_{\Omega_n} f}^n g)(x) \right| \leq C_\Omega h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)} \|f\|_{(W^{t,2}(\Omega))^d}$$

for each  $x \in \Omega$ .

*Proof* Since  $g \in W^{t,2}(\Omega) \subseteq W^{s,2}(\Omega) \hookrightarrow C^{0,1}(\Omega)$ , we begin by writing the inequality

$$\begin{aligned} \left| (U_f^n g)(x) - (\mathfrak{U}_{P_{\Omega_n} f}^n g)(x) \right| &:= |(\mathbb{P}_{\Omega_n} g)(f(x)) - (\mathbb{P}_{\Omega_n} g)(\mathbb{P}_{\Omega_n}(f(x)))| \\ &\leq \|\mathbb{P}_{\Omega_n} g\|_{C^{0,1}(\Omega)} \|f(x) - (\mathbb{P}_{\Omega_n} f)(x)\|_{\mathbb{R}^d} \\ &\leq \|g\|_{W^{s,2}(\Omega)} \|(I - \mathbb{P}_{\Omega_n})f\|_{C(\Omega, \mathbb{R}^d)}. \end{aligned}$$

Since  $W^{s,2}(\Omega) \hookrightarrow C(\Omega)$ , we know that  $(W^{s,2}(\Omega))^d \hookrightarrow C(\Omega, \mathbb{R}^d)$ . It follows that

$$\left| (U_f^n g)(x) - (\mathfrak{U}_{P_{\Omega_n} f}^n g)(x) \right| \lesssim \|g\|_{W^{s,2}(\Omega)} \|(I - \mathbb{P}_{\Omega_n})f\|_{(W^{s,2}(\Omega))^d}.$$

Now the desired result follows from an application of the many zeros Theorem 10, applied to each component of the vector-valued function  $f$ .

Before concluding this section, we note that the results to Theorems 6 and 8 can be combined to obtain the total error bound

$$\left| (U_f g)(x) - (\mathfrak{U}_{P_{\Omega_n} f}^n g)(x) \right| \leq \left( \tilde{C}_\Omega + C_\Omega \|f\|_{(W^{t,2}(\Omega))^d} \right) h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)}$$

with  $\tilde{C}_\Omega$  the constant from Theorem 6 and  $C_\Omega$  the constant from Theorem 8.

## 5 Approximation of the Orbital Koopman Operator $\mathcal{U}_\phi : H_X \rightarrow \phi^*(H_X)$

With the analysis in Section 4 of the Koopman operator  $U_f : H_X \rightarrow f^*(H_X)$ , the study of approximations of the orbital Koopman operator  $\mathcal{U}_\phi : H_X \rightarrow \phi^*(H_X)$  proceeds similarly, but with a few caveats. Differences arise owing to the fact that  $\phi : T \rightarrow \mathbb{R}^d$  is a function of time, and correspondingly the pullback spaces  $\phi^*(H_X)$  or  $\phi^*(R_\Omega(H_X))$  contain functions of time. Approximations of the Koopman operator  $U_f$  are expressed in terms of the pullback spaces  $f^*(H_X)$ , or  $f^*(R_\Omega(H_X))$  that consists of a collection of spatial functions defined over  $X$ .

As before, we construct two types of general approximations to  $\mathcal{U}_\phi$ . We set

$$\mathcal{U}_\phi^n g := \mathcal{U}_\phi P_{\Omega_n} g := P_{\Omega_n} g \circ \phi, \quad (22)$$

which is entirely analogous to the definition of the  $U_f^n$  in Equation 17. The explicit realization of  $\mathcal{U}_\phi^n$  is written as

$$(\mathcal{U}_\phi^n g)(t) := \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) g(\xi_i) \mathfrak{K}_{\xi_j}(\phi(t))$$

for each  $t \in T$  and  $g \in H_X$ . Just about all of the comments and qualifications made about the approximation  $U_f^n$  of the Koopman operator  $U_f$  can be made regarding the approximation  $\mathcal{U}_\phi^n$  of the orbital koopman operator  $\mathcal{U}_\phi$ . The coordinate expression above for  $U_\phi^n := P_{\Omega_n} g \circ \phi$  coincides with that for the approximation that is defined  $\mathcal{U}_\phi^n := \mathbb{P}_{\Omega_n} g \circ \phi$  in the case that  $g \in R_\Omega(H_X)$  and  $t \mapsto \phi(t) \in \Omega \subset X$ . This latter definition is used when samples  $\Xi$  are dense in  $\Omega$ , but not dense in the entire state space  $X$ . Again, in either interpretation the approximation  $\mathcal{U}_\phi^n$  cannot be realized in practice unless a closed form expression for the trajectory  $t \mapsto \phi(t)$  is known. Still, error estimates for  $\mathcal{U}_\phi^n$  are useful for understanding the ideal performance of  $\mathcal{U}_\phi^n$  in the event that  $t \mapsto \phi(t)$  is known. Also, we briefly outline below the fact that error bounds for approximations  $\mathcal{U}_\phi^n$  of the orbital Koopman operator  $\mathcal{U}_\phi$  are entirely analogous to those for approximations  $U_f^n$  of the Koopman operator  $U_f$ .

On the other hand, data-driven approximations of the orbital Koopman operators  $\mathcal{U}_\phi$  can be more complicated than those for  $U_f$ . The data-driven approximation of  $\mathcal{U}_\phi g = g \circ \phi$  requires an approximation of the spatial function  $g$ , and an approximation of the time-dependent function  $t \mapsto \phi(t)$ .

To construct approximations of the time-dependent function, we introduce an additional RKH space  $H_T$  induced by an admissible kernel  $p : T \times T \rightarrow \mathbb{R}$ ,

$$H_T = \overline{\text{span}\{p_t \mid t \in T\}}.$$

Up until this point, we have assumed that  $T = [0, \infty)$ , but now we (again) overload this notation and assume that  $T \subset [0, \infty)$  is a bounded subset contained in  $[0, \infty)$ . We define

$$T_n := \{t_i \mid \phi(t_i) \in \Omega_n\}$$

to be the finite set of times corresponding to samples in  $\Omega_n \subset \Omega$ , and we denote by

$$H_{T_n} := \text{span}\{p_{t_i} \mid t_i \in T_n\}$$

the finite dimensional approximant spaces contained in  $H_T$ . The construction of approximations of time-varying functions over the bounded set  $T$  is carried out using the  $H_T$ -orthogonal projections  $Q_{T_n} : H_T \rightarrow H_{T_n}$ , which again coincide with the unique interpolation operator over  $H_{T_n} \subset H_T$ . We overload the notation and extend the definition of the projection operator  $Q_{T_n}$  so that it applies entrywise to vector-valued functions,

$$Q_{T_n} \phi := \{Q_{T_n} \phi_1, \dots, Q_{T_n} \phi_d\}^T.$$

We limit our discussion of the data-driven approximation of the Koopman operator  $\mathcal{U}_\phi$  to the case that the samples  $\Xi$  are dense in a proper subset  $\Omega \subset X := \mathbb{R}^d$ , but are not dense in the entire state space. The data-driven approximation  $\mathcal{U}_{Q_{T_n}}^n$  of the orbital Koopman operator  $\mathcal{U}_\phi g := g \circ \phi$  with  $g \in R_\Omega(H_X)$  is now defined to be

$$U_{Q_{T_n} \phi}^n := \mathcal{U}_{Q_{T_n} \phi} \mathbb{P}_{\Omega_n} g := (\mathbb{P}_{\Omega_n} g) \circ (Q_{T_n} \phi)$$

for each  $t \in T \subset [0, \infty)$ . This operator has the realization

$$\left( U_{Q_{T_n} \phi}^n g \right)(t) = \sum_{1 \leq i, j \leq n} \mathbb{K}_{j,i}^{-1}(\Omega_n) g(\xi_i) \mathfrak{K}_{\xi_j} \left\{ \begin{array}{c} \sum_{1 \leq \ell, m \leq n} \mathbb{P}_{m,\ell}^{-1}(T_n) \phi_1(t_m) p_{t_\ell}(t) \\ \vdots \\ \sum_{1 \leq \ell, m \leq n} \mathbb{P}_{m,\ell}^{-1}(T_n) \phi_d(t_m) p_{t_\ell}(t) \end{array} \right\}$$

with the Grammian matrix  $\mathbb{P}(T_n) := [p(t_i, t_j)]_{t_i, t_j \in T_n}$ . By inspection, this coordinate expression can be formed in terms of the samples  $\Omega_n := \{\xi_i := \phi(t_i) \mid t_i \in T_n\}$  and  $T_n := \{t_\ell \mid \phi(t_\ell) \in \Omega_n\}$ .

### 5.1 Approximations $\mathcal{U}_\phi^n$

From the similarity of the definitions of the approximation  $\mathcal{U}_\phi^n$  of the orbital Koopman operator  $\mathcal{U}_\phi$  to that of the approximation  $U_f^n$  of  $U_f$ , it is not surprising that the associated error bounds are derived in nearly identical fashion. Each of the Theorems 1, 4, 5, and 6 are readily transcribed to generate error bounds for  $\mathcal{U}_\phi^n$  by replacing  $U_f \rightsquigarrow \mathcal{U}_\phi$ ,  $U_f^n \rightsquigarrow \mathcal{U}_\phi^n$ ,  $f^*(H_X) \rightsquigarrow \phi^*(H_X)$ , and  $f^*(W^{s,2}(\Omega)) \rightsquigarrow \phi^*(W^{s,2}(\Omega))$ . With these changes the pointwise estimates for  $x \in \Omega$  are also replaced by pointwise estimates for  $t \in T$ . We do not write modified versions of Theorems 1 through 6 since their form and proofs would be largely redundant. Below we just summarize these results and leave the details to the reader.

For instance, if the hypotheses of Theorem 1 hold, we conclude that

$$\|\mathcal{U}_\phi g - \mathcal{U}_\phi^n g\|_{\phi^*(H_X)} \leq \|(I - P_{\Omega_n})g\|_{H_X}$$

for all  $g \in H_X$ , and further if  $\phi : T \rightarrow \mathbb{R}^d$  is continuous that we have the pointwise bound

$$|(\mathcal{U}_\phi g)(t) - (\mathcal{U}_\phi^n g)(t)| \leq \|(I - P_{\Omega_n})g\|_{H_\Omega} \rightarrow 0$$

as  $n \rightarrow \infty$  for  $t \in T$  and  $g \in H_\Omega$ . If the hypotheses of Theorem 4 apply, we conclude that

$$|(\mathcal{U}_\phi g)(t) - (\mathcal{U}_\phi^n g)(t)| \leq C_M h_{\Omega_n, M}^k \|g\|_{H_M}$$

for each  $t \in T$ , while the left hand side is bounded by

$$|(\mathcal{U}_\phi g)(t) - (\mathcal{U}_\phi^n g)(t)| \leq C_M h_{\Omega_n, M}^{t-d/2} \|g\|_{H_M}$$

if the hypotheses of Theorem 5 are true. Finally, the analysis in Theorem 6 can be extended to the case at hand to show that

$$\|\mathcal{U}_\phi g - \mathcal{U}_\phi^n g\|_{\phi^*(W^{s,2}(\Omega))} \leq C_M h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)},$$

for all  $g \in W^{s,2}(\Omega)$ , under the suitably modified hypotheses of that theorem. Using the fact that the pointwise error in time is less than

### 5.2 Data-Dependent Approximations $\mathcal{U}_{Q_{T_n}\phi}^n$

The derivation of error bounds for the data-dependent approximations  $\mathcal{U}_{Q_{T_n}\phi}^n$  is modeled on the approximations in Theorems 7 and 8 above, but with changes to account for the approximation of  $g$  and  $\phi$  in different finite dimensional spaces of functions.

**Theorem 9** *Suppose that the set of samples  $\Xi$  is dense in the set  $\Omega \subset X := \mathbb{R}^d$ , the set of times  $\bigcup_n T_n$  is dense in the compact set  $T \subset \mathbb{R}^+$ , the kernel  $\mathfrak{K} : X \times X \rightarrow \mathbb{R}$  induces the native space  $H_X$  over  $X$ ,  $g \in R_\Omega(H_X) \hookrightarrow C^{0,1}(\Omega)$ , and the admissible kernel  $p : T \times T \rightarrow \mathbb{R}$  induces the RKH space  $H_T \approx W^{r,2}(T) \hookrightarrow W^{u,2}(T)$  with  $1/2 \leq u \leq \lceil r \rceil - 1$ . Then there are numbers  $C_T, h_T > 0$  such that for all  $T_n \subset T$  such that  $h_{T_n, T} < h_T$  and  $\phi \in (W^{r,2}(T))^d$  we have*

$$\left| (\mathcal{U}_\phi^n g)(t) - (\mathcal{U}_{Q_{T_n}\phi}^n g)(t) \right| \leq C_T h_{T_n, T}^{r-u} \|g\|_{R_\Omega(H_X)} \|\phi\|_{(W^{r,2}(\Omega))^d}$$

for each  $t \in T$ . If, furthermore,  $R_\Omega(H_X) \approx W^{t,2}(\Omega) \subseteq W^{s,2}(\Omega)$  with  $d/2 \leq s \leq \lceil t \rceil - 1$  and  $\Omega \subset X$  is a compact set with a Lipschitz boundary, we have the total error estimate

$$\left| (\mathcal{U}_\phi g)(t) - (\mathcal{U}_{Q_{T_n}\phi}^n g)(t) \right| \leq \left( C_\Omega h_{\Omega_n, \Omega}^{t-s} + C_T h_{T_n, T}^{r-u} \|\phi\|_{(W^{r,2}(T))^d} \right) \|g\|_{W^{t,2}(\Omega)}$$

for each  $t \in T$ .

*Proof* We start as in the proof of Theorem 8 by writing

$$\begin{aligned} \left| (\mathcal{U}_\phi^n g)(t) - (\mathcal{U}_{Q_{T_n}\phi}^n g)(t) \right| &:= |(\mathbb{P}_{\Omega_n} g)(\phi(t)) - (\mathbb{P}_{\Omega_n} g)(Q_{T_n}(\phi(t)))|, \\ &\leq \|\mathbb{P}_{\Omega_n} g\|_{C^{0,1}(\Omega)} \|\phi(x) - (Q_{T_n} \phi)(x)\|_{\mathbb{R}^d}, \\ &\leq \|g\|_{R_\Omega(H_X)} \|(I - Q_{T_n})\phi\|_{C(\Omega, \mathbb{R}^d)}, \\ &\leq \|g\|_{R_\Omega(H_X)} \|(I - Q_{T_n})\phi\|_{W^{u,2}(\Omega)}. \end{aligned}$$

Now the results follows the same argument as in the proof of Theorem 6. We know that the  $H_T$ -orthogonal projection onto  $H_{T_n}$  coincides with the interpolant on the subspace  $H_{T_n}$ , so that  $(I - Q_{T_n})\phi$  is equal to zero when restricted to  $T_n$ . We apply the many zeros Theorem 10 to each entry to obtain the desired inequality.

Using the fact that the error  $\left| (\mathcal{U}_\phi g)(t) - (\mathcal{U}_{Q_{T_n}\phi}^n g)(t) \right|$  is bounded from above by  $\|\mathcal{U}_\phi g - \mathcal{U}_\phi^n g\|_{\phi^*(W^{s,2}(\Omega))}$ , we conclude that  $\left| (\mathcal{U}_\phi g)(t) - (\mathcal{U}_{Q_{T_n}\phi}^n g)(t) \right| \leq C_M h_{\Omega_n, \Omega}^{t-s} \|g\|_{W^{t,2}(\Omega)}$ , for all  $g \in W^{s,2}(\Omega)$ . Using triangle inequality, we arrive at the second result of the theorem.

## 6 Analytical And Numerical Examples

*Example 7* In this paper we have studied the orbital Koopman  $\mathcal{U}_\phi$  operator for some semidynamical systems in continuous time, and we have noted that it is not the usual Koopman operator defined for time dependent systems. In this example we give one particular example that motivates the study of the orbital Koopman operator. In a series of papers by the authors [56, 73] the authors have investigated the RKH space embedding method for adaptive estimation of systems that are governed by uncertain nonlinear ODEs that have the form

$$\dot{\phi}(t) = Ax(t) + Bf(\phi(t))$$

where  $A \in \mathbb{R}^{d \times d}$  is known,  $B \in \mathbb{R}^{d \times 1}$  is known, and the function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  unknown. It is shown in [56] that one form for the estimator equations of the RKH embedding method can be written as

$$\begin{aligned} \dot{\hat{\phi}}(t) &= A\hat{\phi}(t) + B\mathcal{E}_t\mathcal{U}_\phi\hat{f}(t) \\ \dot{\hat{f}}(t) &= \gamma(B\mathcal{E}_t\mathcal{U}_\phi)^*(\phi(t) - \hat{\phi}(t)) \end{aligned} \quad (23)$$

with  $\hat{\phi}(t) \in \mathbb{R}^d$  the state estimate of  $\phi(t)$ , the function  $\hat{f}(t) \in H_X$  an estimate of  $f$ ,  $\mathcal{U}_\phi : H_X \rightarrow \phi^*(H_X)$  the orbital Koopman operator, and  $\mathcal{E}_t : \phi^*(H_X) \rightarrow \mathbb{R}$  the evaluation operator at  $t \in \mathbb{R}^+$ . These equations define an evolution in the infinite dimensional space  $\mathbb{R}^d \times H_X$ , so they define an example of a distributed parameter system. As discussed in [56], approximate trajectories are generated by replacing  $\mathcal{U}_\phi$  with the approximations  $\mathcal{U}_f^n$  or  $\mathcal{U}_\phi^n$ . The results of this paper can therefore be interpreted as a source of ways for generating estimates of convergence rates for the RKH embedding method.

*Example 8* In this example we return to Example 1 and study approximations  $\mathcal{U}_\phi^n$  of the orbital Koopman operator associated with the orbit  $\Gamma^+(\phi_0)$  of the trajectory  $t \mapsto \phi(t) \in \mathbb{R}^2$  through the initial condition  $\phi_0 = \{-0.2, 1.6\}^T \in \mathbb{R}^2$ . We define the limiting set  $\Omega$  to be an arc of the motion over the open interval  $(a, b) \subset \mathbb{R}^+$ , that is,

$$\Omega = \bigcup_{t \in (a, b)} \phi(t).$$

We define a set of samples  $\Xi := \bigcup_{n \in \mathbb{N}} \Omega_n$  where for each  $n \in \mathbb{N}$  we choose

$$\Omega_n := \left\{ \xi_{n,i} := \phi(t_{n,i}) \mid a < t_{n,1} < t_{n,2} < \dots < t_{n,n} < b. \right\}$$

Note that we have assumed that all of the finite samples have been selected from the one trajectory  $t \mapsto \phi(t)$  that passes through the initial condition  $\phi_0 \in \mathbb{R}^d$ . We allow (but do not require) that the discrete sample sets  $\Omega_n$  are nested, so that  $\Omega_n \subset \Omega_{n+1}$ . We require, however, that the fill distance  $h_{\Omega_n, \Omega} \rightarrow 0$  as  $n \rightarrow \infty$ . Suppose that we choose the kernel  $\mathfrak{K} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  that is defined over the entire real plane. For instance, we can choose  $\mathfrak{K}$  to be Sobolev-Matern kernel  $\mathfrak{K}_m$  with  $m > 1$ , the Abel kernel  $\mathfrak{K}_{\alpha,1,2}$ , the  $\ell^1$ -exponential kernel  $\mathfrak{K}_{\alpha,1,1}$ , or the Gaussian kernel  $\mathfrak{K}_{\alpha,2,2}$ , with all of these kernels defined on  $X \times X := \mathbb{R}^2 \times \mathbb{R}^2$ . Whatever the specific choice of the kernel  $\mathfrak{K}$ , it defines the RKH space  $H_X$  of functions over the state space  $X := \mathbb{R}^2$ : functions in  $H_X$  are defined on all of  $X := \mathbb{R}^2$ . We also define finite dimensional spaces of approximants

$$H_{\Omega_n} := \text{span} \left\{ \mathfrak{K}_{\xi_{n,i}} \mid \xi_{n,i} \in \Omega_n \right\}.$$

With these definitions, we know that

$$H_{\Omega_n} \subset H_\Omega \subseteq H_X$$

all contain functions that are defined over all of the state space  $X := \mathbb{R}^2$ . From Theorem 1, we know that

$$\|\mathcal{U}_\phi g - \mathcal{U}_\phi^n g\|_{\phi^*(H_X)} \leq \|(I - P_{\Omega_n})g\|_{H_X}$$

for all  $g \in H_X$ . It is trivial that for any of the above choices of the kernel  $\mathfrak{K}$ , the map  $x \mapsto \mathfrak{K}_x \in H_X$  is one-to-one. If we can additionally ensure that the kernel  $\mathfrak{K}$  separates all the  $d_X := d_{\mathbb{R}^2}$ -closed subsets of  $X := \mathbb{R}^2$ , then we can conclude from Theorem 2 that  $\|(I - P_{\Omega_n})g\|_{H_\Omega} \rightarrow 0$  as  $n \rightarrow \infty$  for each  $g \in H_\Omega$ . As noted in [58], the Abel kernel  $\mathfrak{K}_{\alpha,1,2}$  and the  $\ell^1$ -exponential kernel separates every closed subset of  $\mathbb{R}^2$ , so both of these kernels generate convergent approximations in  $H_\Omega$ . Also, the Matern-Sobolev kernel for  $m > d/2 = 1$  induces a native space  $H_X$  that is equivalent to the Sobolev space  $W^{m,2}(\mathbb{R}^2)$ . But as discussed more fully in [74], the Sobolev space  $W^{m,2}(\mathbb{R}^2)$  contains  $C_0^\infty(\mathbb{R}^2)$ , and therefore it has a rich family of smooth bump functions. The Sobolev space  $W^{s,2}(\mathbb{R}^2)$  therefore separates all the  $d_{\mathbb{R}^2}$ -closed subsets of  $X = \mathbb{R}^2$ . This implies the convergence of the Koopman approximations in  $H_\Omega$  in this case also. However, it is known that the Gaussian kernel does not separate all the

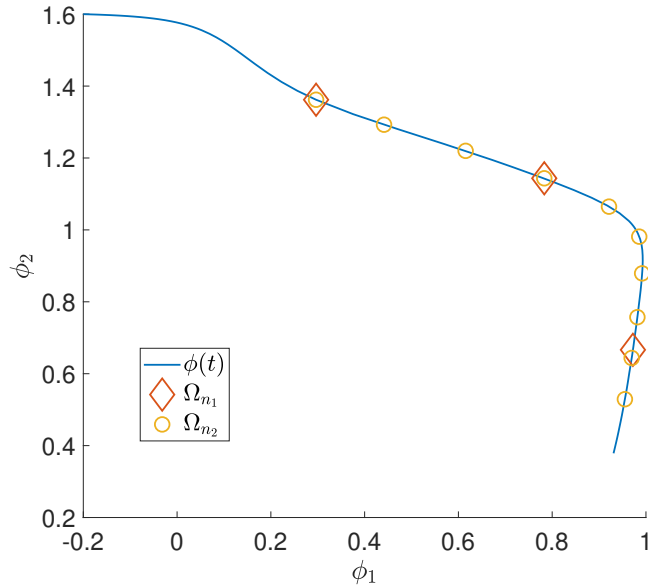


Fig. 2: From initial condition  $\phi_0 = \{-0.2, 1.6\}^T$ , the input trajectory  $\phi(t)$  is generated over 7 ns. Two estimates are generated using two different collection of centers  $\Omega_{n_1}$  and  $\Omega_{n_2}$ , which are placed quasi-uniformly along the orbit from times  $t_1 = 1$  ns to  $t_2 = 6$  ns.

$d_{\mathbb{R}^2}$ -closed subsets of  $X := \mathbb{R}^2$ . [58] We would require additional analysis to conclude convergence in this case since Theorem 2 does not apply.

Here we examine the numerical results of approximating a function  $g(t) = \sin(\phi_1(t) + \phi_2(t))^3$ . Using the Matern-Sobolev kernel with  $m = 3/2$ . As mentioned in [52], measuring time in nanoseconds and the currents  $\phi_1$  and  $h(x_1)$  in mA with parameters,  $C, L$ , and  $u$  as given in Example 1, the equations of motion for the state dynamics are given by

$$\dot{\phi}(t) = \begin{Bmatrix} 0.5(-h(\phi_1(t)) + \phi_2(t)) \\ 0.2(-\phi_1(t) - 1.5\phi_2(t) + 1.2) \end{Bmatrix},$$

$$\phi(0) = \{-0.2, 1.6\}^T,$$

As mentioned in [52], this system has two stable and one unstable equilibrium. From initial condition  $\phi_0$ , the trajectory converges to the stable equilibrium at  $\{0.884, 0.21\}^T$ . The simulation is run over 7 ns. Two estimates of  $\mathcal{U}_\phi g$  are calculated. The first and second estimates uses 3 and 13 kernel centers, respectively, spaced quasi-uniformly along the orbit. Both estimates begin placing centers from times  $t_1 = 1$  ns to  $t_2 = 7$  ns as demonstrated in Figure 2. Figure 3 illustrates the behavior of the estimates  $\mathcal{U}_\phi^{n_1}$  and  $\mathcal{U}_\phi^{n_2}$  of the orbital Koopman operator  $\mathcal{U}_\phi$  acting on  $g$ . From the figure, it is clear that the error between the estimate and the true operator acting on  $g$  is low near the centers placed between  $t_1$  and  $t_2$ . Additionally, we can see that the estimate  $\mathcal{U}_\phi^{n_2} g$  using the larger number of centers has a much smaller error than  $\mathcal{U}_\phi^{n_1} g$  since  $h_{\Omega_{n_2}, \Omega} < h_{\Omega_{n_1}, \Omega}$ .

To better understand the effects of the fill distance on the reduction in error, we examine the bounds on the error as given by orbital Koopman operator's version of Theorem 6 (see Section 5). In this case, the dimension of the ambient space is  $d = 2$ , and the dimension of the restricted set along the orbit is  $k = 1$ . For each estimate, kernels are placed quasi-uniformly throughout the orbit from the initial time of  $t_1 = 0$  to the final run time of  $t = 9$  ns. When using Sobolev-Matern kernel of order  $m = 3/2 > 1$  and choosing  $s = m - (d - k)/2 > 0$  and  $s \leq \lceil t \rceil - 1$ , we can bound the error with a reduction in the fill distance  $h_{\Omega_n, \Omega}$  as seen in Figure 4.

*Example 9* We next turn to the study of the Lotka-Volterra Equations in Example 3, which we interpret as an example that determines a flow on a manifold  $M$ . We prepare this example with a careful discussion of a continuous semiflow that can be associated with the governing equations.

In writing the original equations, we are given a set of nonlinear ODEs over the state space  $X := \mathbb{R}^2$ . It is relatively straightforward to show that every initial condition  $\phi_0$  in the open first quadrant generates a continuous solution of these equations, one that is defined for all  $t \in \mathbb{R}^+$ . The solutions depend continuously on the initial conditions. These properties suffice to determine a continuous semigroup  $\{S(t)\}_{t \in \mathbb{R}^+}$  in the usual way [75] from the ODEs, provided we can define a suitable domain of initial conditions. We turn to this question next and seek to characterize a positive invariant set contained in the open first quadrant. There is a single equilibrium in the open first quadrant located at  $\phi_e := \{1, 2\}$ . Moreover, it is also known and easy to verify [53] that the trajectory  $t \mapsto \phi(t)$  passing through an initial condition  $\phi_0$  in the open first quadrant leaves the function

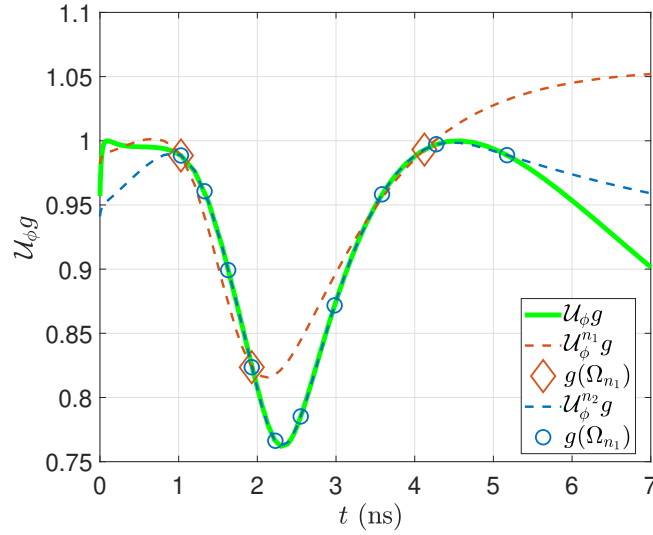


Fig. 3: An illustration of the behavior of the estimates  $\mathcal{U}_\phi^{n_1}$  and  $\mathcal{U}_\phi^{n_2}$  of the orbital Koopman operator  $\mathcal{U}_\phi$  acting on  $g$ . The error between the estimate and the true operator acting on  $g$  is smaller near the centers placed between  $t_1$  and  $t_2$ . From the figure, it is also evident that the error in the estimate  $\mathcal{U}_\phi^{n_2} g$  is smaller than that of the  $\mathcal{U}_\phi^{n_1} g$  since  $h_{\Omega_{n_2}, \Omega} < h_{\Omega_{n_1}, \Omega}$ .

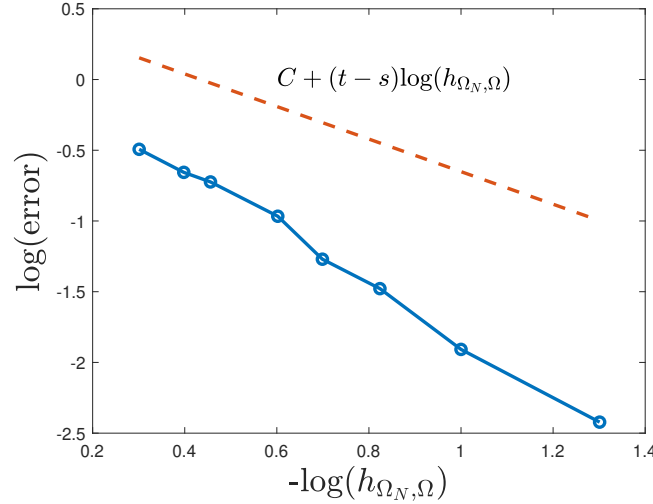


Fig. 4: An illustration of the convergence of the estimate  $\mathcal{U}_\phi^n$  to the orbital Koopman operator  $\mathcal{U}_\phi$  acting on  $g$ . Here the error is given by  $\sup_{t \in T} |\mathcal{U}_\phi g(t) - \mathcal{U}_\phi^n g(t)|$ . The dashed line represents the convergence rate.

$V$  defined by

$$V(\phi) := \ln \phi_1 - \phi_1 + 2 \ln \phi_2 - \phi_2$$

constant. That is, we have  $V(\phi(t)) := V(\phi_0)$  for all  $t \in \mathbb{R}^+$  for any initial condition in the open first quadrant. We can define a compact set  $N \subset \mathbb{R}^2$  as the closed sublevel set

$$N := N_\alpha := \{x \in \mathbb{R}^2 \mid -V(x) \leq \alpha\}$$

for any constant  $\alpha$ , and this set is invariant under the solutions of the ODEs. The fact that  $N$  is closed follows from the fact that it is the inverse image of a closed set under a continuous mapping. If we set

$$S(t)\phi_0 := \phi(t)$$

for each initial condition  $\phi_0 \in N$ , we obtain a continuous semigroup  $\{S(t)\}_{t \in \mathbb{R}^+}$  over the compact, closed (complete) metric space  $(N, d_N) := (N, d_{\mathbb{R}^2})$ . Here  $d_{\mathbb{R}^2}$  is the usual Euclidean distance function restricted to the set  $N$ .

Next we define how a dynamical system on a manifold  $M$  is defined from the semiflow  $\{S(t)\}_{t \in \mathbb{R}^+}$  on  $N$ . The Jacobian of  $V$  is given by

$$DV := \begin{bmatrix} \frac{1}{\phi_1} - 1 & \frac{2}{\phi_2} - 1 \end{bmatrix},$$



and the determinant  $|DV|$  is nonzero for every point in the open first quadrant except the equilibrium  $\phi_e$ , where it is zero. By the implicit function theorem this means that the set  $M \subset N$ , where  $M := M(\phi_0) := \{x \in \mathbb{R}^2 \mid V(x) = V(\phi_0)\}$  with  $\phi_0 \in N$  and  $\phi_0 \neq \phi_e$  is a smooth, connected, regularly embedded submanifold of  $\mathbb{R}^2$ . Any smooth manifold can be equipped with a Riemannian metric, and here we just assume that such a metric has been chosen. We define the mapping  $T(t) : M \rightarrow M$  by restriction,  $T(t) := S(t)|_M$ . The operators  $\{T(t)\}_{t \in \mathbb{R}^+}$  satisfy the algebraic semigroup properties since the semigroup  $\{S(t)\}_{t \in \mathbb{R}^+}$  does and  $M$  is invariant under  $\{S(t)\}_{t \in \mathbb{R}^+}$ . Finally, since  $M$  is a regularly embedded submanifold of  $\mathbb{R}^2$ , we know that there are two constants  $C_1, C_2 > 0$  such that

$$C_1 \|x - y\|_{\mathbb{R}^2} \leq d_M(x, y) \leq C_2 \|x - y\|_{\mathbb{R}^2}$$

for all  $x, y \in M$ . [59] The fact that the metric  $d_M$  on  $M$  is equivalent to the usual Euclidean metric restricted to  $M$  suffices to show that  $\{T(t)\}_{t \in \mathbb{R}^+}$  is a continuous semiflow on the complete metric space  $(M, d_M)$ . Of course, it is clear from this analysis that the manifold  $M$  is just the orbit  $\Gamma^+(\phi_0)$  equipped with a Riemannian metric.

We could use the definition of the semiflow  $\{T(t)\}_{t \in \mathbb{R}^+}$  above to define the orbital Koopman operator  $\mathcal{U}_\phi$  for any initial condition  $\phi_0 \in M$ , and apply our approximation results to the associated orbital Koopman operator  $\mathcal{U}_\phi$ . However, we have already carried this out for an orbital Koopman operator in the last example, and now we consider the case of a discrete flow defined in terms of a mapping  $f : M \rightarrow M$  and approximate the usual Koopman operator  $U_f$ . We can associate a discrete flow with the semigroup  $\{T(t)\}_{t \in \mathbb{R}^+}$  on  $M$  by sampling. Suppose that  $t_i := (i - 1) * h$  for some fixed time step  $h > 0$ . We set

$$\begin{aligned} \phi_{i+1} &= T(h)\phi_i, \\ &= f(\phi_i), \end{aligned}$$

with  $f(\phi) := T(h)\phi$ , and we obtain a discrete evolution of the type studied in Equation 1, but now the system evolves on a manifold  $M$ . By construction the function  $f : M \rightarrow M$ .

Note that even though the governing ODEs constitute a classical case study, the function  $f : M \rightarrow M$  is not known in closed form. In this sense, the approximation of the Koopman operator  $U_f$  is a nontrivial question even for this rather straightforward system. In the remainder of this example we employ a *intrinsic* method of approximation. We choose a kernel  $\mathfrak{K} : M \times M \rightarrow \mathbb{R}$ , one that is defined explicitly over the manifold, to construct approximations of the Koopman operator.

We will estimate the Koopman operator in terms of the Sobolev spaces  $W^{m,2}(M)$ . Recall that if the dimension of  $M$  is  $k$ , then  $W^{m,2}(M)$  is continuously embedded in  $C^0(M)$  if  $m > k/2$ . Since  $M$  is a one dimensional manifold, choosing  $m > 1/2$  will suffice to guarantee that  $W^{m,2}(M)$  is a RKH space. We choose the Sobolev-Matern kernel  $\mathfrak{K}_m : M \times M \rightarrow \mathbb{R}$  for some  $m > 1/2$ . For some standard manifolds, the Sobolev-Matern kernels are known in closed form. For the problem at hand, since the manifold  $M$  is not one of the standard examples, determining the Sobolev-Matern kernel is difficult. We must solve for the fundamental solution of the order  $2m$  elliptic differential operator  $\mathcal{L} := \sum_{j=0}^m (\nabla^j)^* \nabla^j$ , which is the  $m^{\text{th}}$  order Laplace-Beltrami operator on the manifold  $M$ . In this equation  $\nabla$  is the covariant derivative of the Riemannian manifold. On solving for the Sobolev-Matern kernel  $\mathfrak{K}_m : M \times M \rightarrow \mathbb{R}$ , we can apply Theorem 7 to find that the pointwise error is bounded like

$$|(U_f g)(x) - (\mathfrak{U}_f^m g)(x)| \leq C_M h_{\Omega_n, M}^{m-s} (\|g\|_{W^{m,2}(M)} + \|U_f g\|_{W^{m,2}(M)}).$$

Before proceeding to the next example, we emphasize some key points about this intrinsic approximation method. To *realize this rate of convergence in applications*, we would first have to solve for the Matern-Sobolev kernel over the manifold  $M$ . Even though we know a great deal about the manifold  $M$  in this case, like the fact that it is a level set of a smooth function, the solution for the Matern-Sobolev kernel is nontrivial since the volume measure over the manifold is difficult to derive. In other problems the situation can be much worse. The closed form representation of the manifold  $M$  is usually not known. In such cases it is impossible to construct the Sobolev-Matern kernels over a manifold  $M$ , even in principle. Even if the manifold is known, the calculation of the solution of the Matern-Sobolev kernels can be prohibitively complicated. This would amount to 1) finding an explicit (finite) atlas for the manifold, 2) using the coordinate charts to express the representations of the Laplace-Beltrami operator in each chart domain, 3) and solving the coordinate expressions for the kernel in each chart domain. This is a highly nontrivial exercise in all but the simplest domains.

This would appear to be a significant barrier to using the bounds derived in the paper in some important (data-driven) applications: completing these calculations might require heroic efforts. Fortunately, we discuss an extrinsic example next which does not suffer from these issues.

*Example 10* From the last example, it is clear that the theorems derived in the paper *in principle* define strong bounds on the error in approximations of different types of Koopman operators. However, it is also clear that coming up with realizations of the kernels that achieve these rates of convergence can be problematic. Using the same equations as the previous example, we now describe an *extrinsic* method for constructing Koopman approximations that realize the error bounds described in the paper.

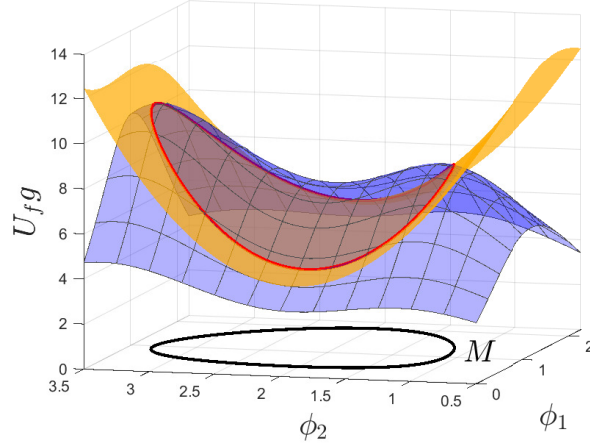


Fig. 5: A comparison between the approximation  $\mathfrak{U}_f^n g$  and the infinite-dimensional Koopman operator  $U_f g$  over the manifold  $M$  of the underlying Lotka-Volterra dynamics. Here we can see the approximation  $\mathfrak{U}_f^n g$  represented by the blue mesh intersects the Koopman operator  $U_f g$  represented by the orange surface at the points in red collected above the manifold.

The discrete dynamics defined in terms of the function  $f := T(h) : M \rightarrow M$  gives rise to the Koopman operator  $U_f$ . The manifold  $M$  is a smooth, one dimensional, closed, connected, Riemannian manifold that is regularly embedded in the Euclidean space  $X := \mathbb{R}^2$ . It turns out that, since  $X := \mathbb{R}^2$  is such a common and widely used manifold, the Sobolev-Matern kernel  $\mathfrak{K}_m$  of order  $m$  over  $X := \mathbb{R}^2$  is well known. The Sobolev-Matern kernel of order  $m$  over  $X := \mathbb{R}^d$  is given in [59] as

$$\mathfrak{K}_m(x, y) := \frac{2^{1-(m-d/2)}}{\Gamma(m-d/2)} \|x - y\|_{\mathbb{R}^d}^{m-d/2} B_{m-d/2}(\|x - y\|_{\mathbb{R}^d})$$

with  $B_{m-d/2}$  the Bessel function of order  $m - d/2$ . This kernel is easily calculated for any  $x, y \in \mathbb{R}^d$ , and in particular it can be computed over a nontrivial or unknown submanifold  $M$  of  $X := \mathbb{R}^d$  since it is known in closed form on the larger set  $X$ .

As discussed in Section 2.2, the restricted kernel  $r_M(x, y) := \mathfrak{K}_m|_{M \times M}(x, y)$  defines the kernel for the RKH space  $R_M(H_X)$ . We want to choose the kernel  $\mathfrak{K}_m$  with sufficiently high order to ensure that  $R_M(H_X)$  is equivalent to a Sobolev space  $W^{s,2}(M)$  for an appropriate degree of smoothness  $s > 0$ . Recall that if the kernel  $\mathfrak{K}_m$  over  $X := \mathbb{R}^d$  has a Fourier transform that satisfies the algebraic decay condition of order  $m$  in Equation 26 with  $m > d/2$ , then  $H_X \approx W^{m,2}(\mathbb{R}^d)$ . But the Sobolev-Matern kernel  $\mathfrak{K}_m$  defines the Sobolev space  $W^{m,2}(\mathbb{R}^d)$ : it automatically satisfies the decay condition in Equation 26 by definition. Moreover, when  $m > d/2$ , the Sobolev embedding Theorem ensures that  $W^{m,2}(\mathbb{R}^d) \hookrightarrow C^0(\mathbb{R}^d)$  and the trace map  $T$  is just the restriction map on these continuous functions. Recall from Proposition 2 that the trace (or restriction) map is a continuous mapping of  $W^{m,2}(\mathbb{R}^d)$  onto  $W^{s,2}(M)$  when  $s := m - (d - k)/2 > 0$ . With all of these observations, we conclude that if  $m > d/2$  and  $s = m - (d - k)/2 > 0$ , then  $R_M(H_X) \approx W^{s,2}(M)$ . For the problem at hand this means that we must choose  $m > 1/2$  and  $s = m - 1/2 > 0$ .

For this example, the trajectory was generated from an orbit starting at initial conditions  $\phi_0 = \{2, 2\}^T$  over 12.8 seconds. The approximation was constructed using the Sobolev-Matern kernel of order  $m = 5/2$ . The dimension of the approximation space is determined by the number of kernels  $n = 11$ , which are selected along the manifold quasi-uniformly. Figure 5 shows the approximation  $\mathfrak{U}_f^n$  and Koopman operator  $U_f$  acting on a function  $g$  over the input defined by the Lotka-Volterra dynamics. From the figure, it is clear that the error is minimal along the manifold  $M$  where samples are collected. We also examine the rates of convergence given by Theorem 7 in Figure 6. The number of centers range from  $n = 5$  to  $n = 26$  based on the desired fill distance. The error is bounded by a function of the fill distance, which decreases as the kernel centers are increased.

*Example 11* In this example we study the Van Der Pol oscillator described in Equations 2. First, the governing system of ODEs in Example 2 generate a continuous, bounded solution through any initial condition  $\phi_0 \in X := \mathbb{R}^2$ , and the solutions depend continuously on the initial conditions. Standard methods in dynamical system theory [72, 75] ensures that the governing equations define a continuous semigroup  $\{S(t)\}_{t \in \mathbb{R}^+}$  on the entire state space  $X := \mathbb{R}^2$ . The positive limit set  $\omega^+(\phi_0)$  is same for any choice of initial condition  $\phi_0 \in X$ . That is, for any pair  $\phi_0, \psi_0 \in X$ , we have  $\omega^+(\phi_0) = \omega^+(\psi_0)$ . Since the forward orbit is precompact in  $X := \mathbb{R}^2$ , the positive limit set  $\omega^+(\phi_0)$  is compact, connected, and positively invariant. Moreover, following essentially the same set of steps that are taken in Example 9, we can likewise show that  $M := \omega^+(\phi_0)$  is a smooth, one

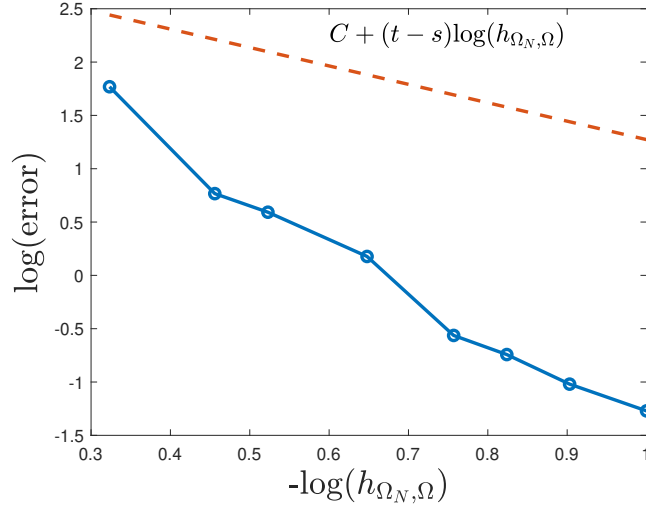


Fig. 6: Examining the rates of convergence given by Theorem 7. The number of centers range from  $n = 5$  to  $n = 26$  based on the desired fill distance. The error  $\sup_{x \in M} |(U_f g)(x) - (\mathfrak{U}_f^n g)(x)|$  is bounded by a function of the fill distance.

dimensional, regularly embedded submanifold of  $X := \mathbb{R}^2$ , and that the restricted semigroup  $T(t) := S(t)|_M$  is a  $d_M$ -continuous semigroup over  $(M, d_M)$ . Similar bounds as derived in Example 9 could be established in the current case, for any initial condition  $\phi_0 \in M$ .

However, in this example, we want to explore the case when initial conditions are outside of the positive limit set and generate substantially different trajectories. We are interested in precise descriptions of relationships between Koopman approximations that arise from samples  $\Xi(\phi_0)$  and  $\Xi(\psi)$  that are collected from different orbits  $\Gamma^+(\phi_0)$  and  $\Gamma^+(\psi_0)$ , when  $\phi_0 \neq \psi_0$  and the initial conditions is outside of  $M := \omega^+(\phi_0) = \omega^+(\psi_0)$ .

As in Example 9, we induce a discrete dynamical system by discretization of the semiflow  $\{S(t)\}_{t \in \mathbb{R}^+}$ , with  $\phi_{n+1} = S(h)\phi_n = f(\phi_n)$  and  $f : X \rightarrow X$  an unknown mapping. We subsequently define  $U_f g := g \circ f$  in the usual way for  $g \in H_X$ . We denote by  $\omega_f^+(\phi_0)$  and  $\omega_f^+(\psi_0)$  the positive limit sets for the discrete flow to distinguish them from the positive limit set  $M$  of the semigroup  $\{S(t)\}_{t \in \mathbb{R}^+}$ . We know that the positive limit sets of the discrete flows are contained in that of the semigroup in continuous time,

$$\begin{aligned} \omega_f^+(\phi_0) &\subseteq \omega^+(\phi_0) = M, \\ \omega_f^+(\psi_0) &\subseteq \omega^+(\psi_0) = M. \end{aligned}$$

Note that the limiting set of the two collections of samples are given by

$$\begin{aligned} \Omega_{\phi_0} &:= \overline{\Xi(\phi_0)} = \Xi(\phi_0) \cup \omega_f^+(\phi_0) \subseteq \Xi(\phi_0) \cup \omega^+(\phi_0) = \Xi(\phi_0) \cup M, \\ \Omega_{\psi_0} &:= \overline{\Xi(\psi_0)} = \Xi(\psi_0) \cup \omega_f^+(\psi_0) \subseteq \Xi(\psi_0) \cup \omega^+(\psi_0) = \Xi(\psi_0) \cup M, \end{aligned}$$

We denote by

$$\begin{aligned} U_{f, \phi_0}^n g &:= (P_{\Omega_{\phi_0, n}} g) \circ f, \\ U_{f, \psi_0}^n g &:= (P_{\Omega_{\psi_0, n}} g) \circ f \end{aligned}$$

the projection-based approximations generated by the samples  $\Xi(\phi_0)$  and  $\Xi(\psi_0)$ , respectively. It is then immediately apparent that

$$\|U_{f, \phi_0}^n g - U_{f, \psi_0}^n g\|_{f^*(H_X)} \leq \|U_f\| \left\{ \|(I - P_{\Omega_{\phi_0, n}})g\|_{H_X} + \|(I - P_{\Omega_{\psi_0, n}})g\|_{H_X} \right\},$$

since both  $U_{f, \phi_0}^n$  and  $U_{f, \psi_0}^n$  are maps from  $H_X$  into  $f^*(H_X)$ . Now we can apply Theorem 2 to conclude that  $\|U_{f, \phi_0}^n g - U_{f, \psi_0}^n g\|_{f^*(H_X)} \rightarrow 0$  as  $n \rightarrow \infty$  for each  $g \in H_{\Omega_{\phi_0} \cap \Omega_{\psi_0}}$ . In other words, the projection-based approximations of the Koopman operator asymptotically agree for all functions  $g \in H_{\Omega_{\phi_0} \cap \Omega_{\psi_0}}$ . Not only does this agree with our intuition, but it describes explicitly the space in which approximations converge.

## 7 Conclusions

In this paper, we study the approximation of the Koopman operator in RKH space associated with semiflows in a complete metric space. The paper explores the relationship between the rate of convergence of the approximation error for a specified subspace of observables and how the data samples fill the set that supports a motion of the semiflow. In particular, we first derive a general error bound for the projection-based approximations of the Koopman operator in an RKH space. A separating property is proven to be sufficient for the convergence of the projection-based Koopman operator approximation. When the RKH space is induced by a Sobolev-type kernel, an upper bound of approximation error is derived in terms of the fill distance. The data-dependent approximation of Koopman operators is discussed next, as well as its relationship to the existing EDMD algorithm. The convergence of the data-dependent approximation is proven based on the convergence of projection-based approximation. The rates of convergence are derived for the cases when the data samples fill a bounded compact manifold, or a bounded subset, respectively. The results above are derived for Koopman operators associated with discrete-time semiflows. For continuous-time semiflows, the convergence rate of the orbital Koopman operators is derived in terms of how the time indices at which data samples are taken fill the entire time interval. Extensive analytical and numerical examples are presented to illustrate the theoretical results.

## Statements & Declarations

### Funding

The authors declare that no funds, grants, or other support were received for carrying out the research presented in this manuscript.

### Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

### Data Availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## A Sobolev Spaces

### A.1 Sobolev Spaces over Subsets $\Omega \subseteq X := \mathbb{R}^d$

The Sobolev spaces over a subset  $\Omega \subseteq X := \mathbb{R}^d$  are defined in terms of the space  $L^2(\Omega) := L^2_\mu(\Omega)$  that consists of functions over  $\Omega$  that are square integrable with respect to a measure  $\mu$ . The norm  $\|\cdot\|_{L^2(\Omega)}$  is induced by the familiar inner product

$$(f, g)_{L^2(\Omega)} := \int_{\Omega} f(\xi)g(\xi)\mu(d\xi),$$

with which  $L^2(\Omega)$  is a Hilbert space. It is most commonly the case that for subsets  $\Omega \subset X := \mathbb{R}^d$ , we choose  $\mu(\cdot)$  to be Lebesgue measure. The most intuitive definition of a Sobolev space can be stated when the order is a nonnegative integer. For integer  $k \geq 0$ , the Sobolev space  $W^{k,2}(\Omega)$  for a subset  $\Omega \subseteq X := \mathbb{R}^d$  consists of those functions in  $L^2(\Omega)$  for which all weak derivatives  $D^\alpha f \in L^2(\Omega)$  with  $D^\alpha(\cdot) = \partial^{|\alpha|}(\cdot)/\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}$  for all integer multiindices  $\alpha = \{\alpha_1, \dots, \alpha_d\} \in \mathbb{N}_0^d$  and  $|\alpha| = \sum_{1 \leq i \leq d} \alpha_i$ . With the inner product

$$(f, g)_{W^{k,2}(\Omega)} := \sum_{0 \leq |\alpha| \leq k} (D^\alpha f, D^\alpha g)_{L^2(\Omega)}$$

the Sobolev space  $W^{k,2}(\Omega)$  is a Hilbert space.

The Sobolev spaces for integer order  $k$  are readily understood: they measure the smoothness of a function in terms of the number of weak derivatives. It turns out that these spaces are a special case of the scale of Sobolev spaces  $W^{r,2}(\Omega)$  that are indexed by a positive real number  $r > 0$ . The spaces  $W^{r,2}(\Omega)$  are defined as an interpolation space between an integer order space  $W^{k,2}(\Omega)$  and  $L^2(\Omega)$ . While we do not need to get into the details of the construction of these spaces, we must use them in certain proofs. For example, in this paper we will often need to talk about the restriction or trace mapping that evaluates a function in a Sobolev space  $W^{k,2}(\Omega)$  over some subset  $\tilde{\Omega} \subset \Omega$ . This topic is described below in the discussion of trace mappings and restrictions in Section A.3. The trace theorems rely on the definition of fractional Sobolev spaces, which are examples of spaces in the real scale of Sobolev spaces  $W^{r,2}(\Omega)$  for real  $r \geq 0$ . See Proposition 2 for instance. Also, many of the approximation theorems in Sobolev spaces are expressed in terms of Sobolev spaces having a positive real, and sometimes fractional, smoothness index. See the many zeros Theorems 10 and 11.

## A.2 Sobolev Spaces over Riemannian Manifolds $M$

In this section we describe how the definitions above are modified to define Sobolev spaces over subsets  $\Omega$  of manifolds  $M$ . A detailed account that is based on intrinsic characterizations can be found in [63, 76, 77], and equivalent definitions in terms of a finite, fixed collection of coordinate charts can be found in [59, 78]. We always assume in this paper that  $M$  is a connected, complete Riemannian manifold with a positive injectivity radius and bounded geometry. A detailed account of these assumptions can be found in [77], page 284, as well as [65] or [76]. If the manifold  $M$  is a smooth, connected, compact Riemannian manifold, or if  $M := \mathbb{R}^d$  these conditions are satisfied.

Suppose that  $M$  is a Riemannian manifold with metric  $g$  and the inner product  $\langle \cdot, \cdot \rangle_{g,p}$  on the tangent space  $T_p M$  at the point  $p \in M$ . By definition the volume measure  $d\mu$  on  $M$  can be written in terms of a set of coordinates  $(x^1, \dots, x^d)$  as  $d\mu(x) := \sqrt{\det(g)} dx^1 \dots dx^d$ . The norm  $\|f\|_{L^p(\Omega)}$  for a subset  $\Omega \subseteq M$  is selected to be the same as the definition above, but the measure is chosen to be volume measure on the manifold  $M$ . For integer  $k > 0$  the Hilbert space  $W^{k,2}(\Omega)$  consists of the functions in  $L^2(\Omega) := L^2_\mu(\Omega)$  that have a bounded norm that is defined in terms of the inner product

$$(f, g)_{W^{k,2}(\Omega)} := \sum_{0 \leq |j| \leq k} \int_{\Omega} \langle \nabla^j f, \nabla^j g \rangle_{g,p} d\mu(p), \quad (24)$$

which is written in terms of the covariant derivative operator  $\nabla$  and the volume measure  $\mu$  on the manifold  $M$ . As in Section A.1, the Sobolev space  $W^{r,2}(\Omega)$  for a real number  $r > 0$  is defined as an interpolation space between an integer order Sobolev space  $W^{k,2}(\Omega)$  and  $L^2(\Omega)$ . [70, 77, 79]

## A.3 Embedding and Trace Theorems

This section gives a brief account of the embedding and trace theorems for Sobolev spaces that are used in this paper. We need these results to establish relationships between certain RKH spaces and Sobolev spaces and to make sense of the notion of restricting a Sobolev function to the boundary, a subset of a domain, or even a manifold. Embedding theorems describe relationships between spaces while trace theorems make precise the evaluation of functions on boundaries and subdomains. It turns out that there are a plethora of embedding theorems associated with Sobolev spaces: these theorems describe when a Sobolev space is contained in another well-known function space. In this short section we review some of the most common embedding theorems for Sobolev spaces, restricting attention to the specific embeddings used in this paper. The reader should see [70, 79, 80] for a detailed treatment of more general embeddings as well as trace theorems.

*Common Embedding Theorems:* Perhaps the simplest embedding theorems for Sobolev spaces that we use in the paper follow directly from the definition of the Sobolev space  $W^{k,2}(\Omega)$  for integer  $k \geq 0$ , and these hold when  $\Omega$  is a subset of either  $X := \mathbb{R}^d$  or a manifold  $M$ . We have  $W^{m,2}(\Omega) \hookrightarrow W^{k,2}(\Omega)$  for any nonnegative integers with  $m \geq k$ . It is a standard fact about Sobolev spaces that for real  $r > 0$ ,  $W^{r,2}(\Omega) \hookrightarrow W^{s,2}(\Omega)$  for any real indices  $r \geq s \geq 0$ .

Elements in a Sobolev space  $W^{r,2}(\Omega)$  with  $\Omega \subseteq \mathbb{R}^d$ , or  $W^{r,2}(M)$  with  $M$  a manifold, are subsets of  $L^2(\Omega)$ , or  $L^2(M)$ , respectively. This means that members of these Sobolev spaces are equivalence classes of functions whose members may not have well-defined values at particular points  $x \in \Omega$  or  $x \in M$ . In contrast, when the admissible kernel  $\mathfrak{K}$  is continuous, the RKH space  $H_X$  only contains continuous functions for which pointwise evaluation is always well-defined. For deriving approximation errors, we are interested in conditions under which it makes sense to evaluate a function in a Sobolev space at a point.

The Sobolev embedding theorem is a powerful tool that establishes sufficient conditions for the evaluation of functions in a Sobolev space at points of the domain. Versions of this theorem hold for Sobolev spaces over a subset  $\Omega \subseteq \mathbb{R}^d$  as well as over subsets of a manifold  $M$ . When the domain  $\Omega \subseteq \mathbb{R}^d$  has a Lipschitz boundary, the Sobolev embedding theorem states that  $W^{r,2}(\Omega) \hookrightarrow C_b(\Omega)$  when  $r > d/2$ , [80] page 85. An analogous results holds when  $M$  is a  $d$ -dimensional, connected, smooth, Riemannian manifold having positive radius of injectivity and bounded geometry. In this case  $W^{r,2}(M) \hookrightarrow C_b(M)$ , as discussed in [63], page 1748, or [64], page 7. Again, if  $M$  is a compact Riemannian manifold or  $M := \mathbb{R}^d$ , these properties are satisfied.

It is also possible to establish a stronger embedding than that guaranteed by the Sobolev embedding theorem described above. It is sometimes the case that a Sobolev space is embedded in the space  $C^{k,\lambda}(\overline{\Omega})$  of  $\lambda$ -Holder continuous functions for  $0 < \lambda \leq 1$ . We define the norm on  $C^{k,\lambda}(\overline{\Omega})$  as

$$\|f\|_{C^{k,\lambda}(\overline{\Omega})} := \|f\|_{C^k(\overline{\Omega})} + \sum_{|\alpha|=k} \sup_{\substack{x \neq y, \\ x, y \in \Omega}} \frac{\|D^\alpha f(x) - D^\alpha f(y)\|}{\|x - y\|^\lambda} = \|f\|_{C^k(\Omega)} + |f|_{C^{k,\lambda}(\overline{\Omega})}. \quad (25)$$

If the domain  $\Omega$  has a Lipschitz boundary, we have the continuous embedding

$$W^{k+m,p}(\Omega) \hookrightarrow C^{k,\lambda}(\overline{\Omega})$$

when  $0 < \lambda \leq m - d/p$ . See page 10 and Chapter 4 in [80]. Note that, in particular, the above continuous embedding implies that  $W^{m,2}(\Omega) \hookrightarrow C^{0,1}(\overline{\Omega})$  for  $m \geq d/2 + 1$ , with  $C^{0,1}(\overline{\Omega})$  the Lipschitz continuous functions. This embedding holds for the entire Euclidean space  $\Omega := \mathbb{R}^d$  also. This embedding plays an important role in the error bound derivation for data-driven approximations of the Koopman operator.

*Relationships between RKH Spaces and Sobolev Spaces:* When the Sobolev embedding theorem holds for a space  $W^{r,2}(\Omega)$  or  $W^{r,2}(M)$  with  $r > d/2$ , the Sobolev space is actually a RKH space. Since we have

$$|\mathcal{E}_x f| = |f(x)| \leq \|f\|_{C(\Omega)} \leq C \|f\|_{W^{s,2}(\Omega)},$$

each evaluation functional  $\mathcal{E}_x : W^{r,2}(\Omega) \rightarrow \mathbb{R}$  is bounded on the Hilbert space  $W^{r,2}(\Omega)$ . This is necessary and sufficient for a Hilbert space to be a RKH space. We have shown this fact for a domain  $\Omega \subseteq \mathbb{R}^d$ , but the argument for a subset of a manifold  $M$  is identical.

The above argument is one way to determine that a Sobolev space is an example of an RKH space. There is yet another, stronger, way to establish a relationship between a Sobolev space and an RKH space. Let  $H_X$  be the native space of the kernel  $\mathfrak{K}$  over  $X := \mathbb{R}^d$ , and denote its Fourier transform as  $\hat{\mathfrak{K}} := \mathcal{F}(\mathfrak{K})$ . We say that the Fourier transform  $\hat{\mathfrak{K}}$  has algebraic decay if

$$\hat{\mathfrak{K}}(\xi) \approx (1 + \|\xi\|_2^2)^{-\tau} \quad (26)$$

for some  $\tau > d/2$ . When this is true, then  $H_X \approx W^{\tau,2}(\mathbb{R}^d)$ . [57, 59]

The fact that the Fourier transform has algebraic decay can also be used to infer relationships between native spaces  $R_\Omega(H_X)$  and Sobolev spaces on proper subdomains  $\Omega \subset \mathbb{R}^d$ , provided that the subdomain is suitably regular. For instance, if  $\mathfrak{K} \in L^1(\mathbb{R}^d)$ , the Fourier transform of the kernel  $\mathfrak{K}$  has algebraic decay of order  $\tau > d/2$ , and  $\Omega \subset \mathbb{R}^d$  has a Lipschitz boundary, then

$$R_\Omega(H_X) \approx W^{\tau,2}(\Omega),$$

see Corollary 10.48, [57]. In the event that  $M$  is a compact, connected,  $k$ -dimensional, regularly embedded, Riemannian submanifold  $M$  of  $X := \mathbb{R}^d$ , from [59], Theorem 5, we know that

$$R_M(H_X) \approx W^{\tau-(d-k)/2,2}(M) \quad (27)$$

when the kernel  $\mathfrak{K}$  has algebraic decay of exponent  $\tau > d/2$ .

*Trace Theorems:* In this paper we will sometimes consider flows that evolve on a submanifold  $M$  contained in  $\mathbb{R}^d$ , and in this case we are interested in approximating functions over the manifold. Again, as above, a function in the Sobolev space may not have a well-defined value at a point on the manifold, when it is a measure zero subset of  $\mathbb{R}^d$ . The problem of defining a value on a set of measure zero is a classical problem in the study of Sobolev spaces, one that is addressed using the trace operator. If a function in the Sobolev space happens to be continuous, the trace operator  $T$  over a subset of the domain is just the restriction of the function to that subset. The trace operator is then extended by continuity to all functions in the Sobolev space for certain ranges of the smoothness parameter. Trace theorems describe the mapping properties of the trace operator and are essential to understanding the values of Sobolev functions on subsets, boundaries, curves, and manifolds. In this paper we will use the following trace theorem on submanifolds  $M$  contained in  $\mathbb{R}^d$ .

**Proposition 2 (Proposition 2, [59])** *Let  $1 \leq k \leq d$  and  $\tau > (d-k)/2$ . Let  $M$  be a smooth,  $k$ -dimensional, compact, embedded submanifold of  $\mathbb{R}^d$ . The trace operator  $T$  extends to a continuous operator mapping  $W^{\tau,2}(\mathbb{R}^d)$  onto  $W^{\tau-(d-k)/2,2}(M)$ . Further, there is a reverse embedding, i.e., there is a bounded linear extension operator  $\mathbb{E}_M : W^{\tau-(d-k)/2,2}(M) \rightarrow W^{\tau,2}(\mathbb{R}^d)$  such that  $T_M(\mathbb{E}_M f) = (\mathbb{E}_M f)|_M = f$  for all  $f \in W^{\tau-(d-k)/2,2}(M)$ .*

This theorem tells us that when we take the trace of functions in  $W^{\tau,2}(\mathbb{R}^d)$  on the submanifold, the resulting functions defined over the manifold have a reduced smoothness: they are elements of  $W^{\tau-(d-k)/2,2}(M)$ .

#### A.4 The ‘‘Many Zeros’’ Approximation Theorems in Sobolev Spaces

A primary resource used in this paper for deriving rates of convergence of Koopman operators are the ‘‘many zeros’’ theorems of scattered data approximation. [57, 63, 64] We use versions of these theorems for the case when  $\Omega$  is a regular subset  $\Omega \subseteq \mathbb{R}^d$ , or when  $M$  is a manifold. We first summarize a form that applies to a subset  $\Omega \subseteq \mathbb{R}^d$ . It is a simplified version of the ‘‘many zeros theorem’’ that can be found in many places such as [59] Proposition 9, page 1764. Also see a similar result in [57] Corollary 11.33, page 201 that applies when  $\Omega$  is open, bounded, satisfies the interior cone condition, and has a Lipschitz boundary.

**Theorem 10** *Suppose that  $\Omega \subset X := \mathbb{R}^d$  is compact and has a Lipschitz boundary. Let  $t \in \mathbb{R}$  with  $t > d/2$ ,  $s \in \mathbb{N}_0$  and  $0 \leq s \leq \lceil t \rceil - 1$ . Then there are constants  $h_\Omega, C_\Omega > 0$  such for all  $\Omega_n \subset \Omega$  such that the fill distance  $h_{\Omega_n, \Omega} \leq h_\Omega$  and  $u \in W^{t,2}(\Omega)$  that satisfies  $u|_{\Omega_n} = 0$  we have*

$$\|u\|_{W^{s,2}(\Omega)} \leq C_\Omega h_{\Omega_n, \Omega}^{t-s} \|u\|_{W^{t,2}(\Omega)}.$$

The above theorem also holds for certain types of smooth manifolds: the version below is a simplified form [63], page 1765.

**Theorem 11** *Suppose that  $M$  is a compact, connected, smooth,  $d$ -dimensional manifold. Let  $t \in \mathbb{R}$  with  $t > d/2$ ,  $s \in \mathbb{N}_0$  with  $0 \leq s \leq \lceil t \rceil - 1$ . Then there are constants  $h_M, C_M > 0$  such that for all  $\Omega_n \subset M$  such that the fill distance  $h_{\Omega_n, M} \leq h_M$  and for all  $u \in W^{t,2}(M)$  that satisfies  $u|_{\Omega_n} = 0$  we have*

$$\|u\|_{W^{s,2}(M)} \leq C_M h_{\Omega_n, M}^{t-s} \|u\|_{W^{t,2}(M)}.$$

#### A.5 Pointwise Approximation Theorems

The theorems above are powerful tools for deriving approximation rates in Sobolev spaces. In either of the above two theorems, if we take  $s = 0$ , we see that the error rates are established in  $L^2(\Omega)$  or  $L^2(M)$ . However, there are many alternative lines of attack to establish approximation rates from bases defined by scattered samples. One important class of these estimates establish pointwise bounds in the norms  $L^\infty(\Omega)$  or  $L^\infty(M)$ . A discussion of the importance of pointwise bounds can be found in [64]. Below we summarize one simple pointwise bound that is used later in deriving approximation rates of Koopman operators for semiflows over manifolds. This theorem is found in [63], Corollary (3.9), page 1749.

**Theorem 12** *Suppose that  $M$  is a  $d$ -dimensional, compact, connected, Riemannian manifold without boundary, let  $\mathfrak{K} : M \times M \rightarrow \mathbb{R}$  be a positive definite kernel that induces the native space  $H_M$ , and suppose further that  $H_M$  is equivalent to the Sobolev space  $W^{r,2}(M)$  for  $r > d/2$ . Then there are constants  $h_M, C_M > 0$  such that for all  $g \in W^{r,2}(M)$  and  $\Omega_n \subset M$  such that  $h_{\Omega_n, M} < h_M$  we have*

$$|g(x) - P_{\Omega_n} g(x)| \leq C_M h_{\Omega_n, M}^{r-d/2} \|g\|_{W^{r,2}(M)}$$

for all  $x \in M$ .

While the above theorem relies on the equivalence of the native space  $H_M$  with a Sobolev space, the following general theorem is also available from [57], page 321, Theorem 17.21:

**Theorem 13** *Let  $M$  be a compact  $C^\ell$  Riemannian manifold, the kernel  $\mathfrak{K} \in C^{2k}(M \times M)$  that induces  $H_M$  be positive definite, and  $\ell > 2k$ . Then there exists  $h_M, C_M > 0$  such that for all  $g \in H_M$  and  $\Omega_n \subset M$  with  $h_{\Omega_n, M} < h_M$  we have*

$$|g(x) - P_{\Omega_n} g(x)| \leq C_M h_{\Omega_n, M}^k \|g\|_{H_M}.$$

for all  $x \in M$ .

We also note that there are versions [63–65] of the many zeros theorems that are stated in terms of spaces  $W^{s,p}(\Omega)$  for  $0 \leq s < \infty$  and  $1 \leq p \leq \infty$ , and these can be used to derive pointwise bounds also by choosing  $s = 0, p = \infty$ .

## References

1. Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLOS ONE*, 11(2):1–19, 02 2016.
2. Bingni W. Brunton, Lise A. Johnson, Jeffrey G. Ojemann, and J. Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016.
3. Natasza Marrouch, Joanna Slawinska, Dimitrios Giannakis, and Heather L Read. Data-driven koopman operator approach for computational neuroscience. *Annals of Mathematics and Artificial Intelligence*, 88(11):1155–1173, 2020.
4. Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013.
5. Fredrik Raak, Yoshihiko Susuki, and Takashi Hikihara. Data-driven partitioning of power networks via koopman mode analysis. *IEEE Transactions on Power Systems*, 31(4):2799–2808, 2016.
6. Joshua L Proctor and Philip A Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International health*, 7(2):139–145, 2015.
7. Joanna Slawinska, Eniko Szekely, and Dimitrios Giannakis. Data-driven koopman analysis of tropical climate space-time variability, 2017.
8. B.O. Koopman. Hamiltonian systems and transformations in hilbert space. *Proc. Natl. Acad. Sci.*, 17(5):315–318, 1931.
9. Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator Theoretic Aspects of Ergodic Theory*. Springer, 2015.
10. Andrzej Lasota and Michael C. Mackey. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer, 1994.
11. Dimitrios Giannakis, Abbas Ourmazd, Joanna Slawinska, and Zhizhen Zhao. Spatiotemporal Pattern Extraction by Spectral Analysis of Vector-Valued Observables. *JOURNAL OF NONLINEAR SCIENCE*, 29(5):2385–2445, OCT 2019.
12. Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *APPLIED AND COMPUTATIONAL HARMONIC ANALYSIS*, 47(2):338–396, SEP 2019.
13. Patrick Gels, Stefan Klus, Jens Eisert, and Christof Schuette. Multidimensional Approximation of Nonlinear Dynamical Systems. *JOURNAL OF COMPUTATIONAL AND NONLINEAR DYNAMICS*, 14(6), JUN 2019.
14. Hao Zhang, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. Online Dynamic Mode Decomposition for Time-Varying Systems. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 18(3):1586–1609, 2019.
15. Samuel Rudy, Alessandro Alla, Steven L. Brunton, and J. Nathan Kutz. Data-Driven Identification of Parametric Partial Differential Equations. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 18(2):643–660, 2019.
16. Anthony M. Degennaro and Nathan M. Urban. SCALABLE EXTENDED DYNAMIC MODE DECOMPOSITION USING RANDOM KERNEL APPROXIMATION. *SIAM JOURNAL ON SCIENTIFIC COMPUTING*, 41(3):A1482–A1499, 2019.
17. Kathleen P. Champion, Steven L. Brunton, and J. Nathan Kutz. Discovery of Nonlinear Multiscale Systems: Sampling Strategies and Embeddings. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 18(1):312–333, 2019.
18. Soledad Le Clairche and Jose M. Vega. Spatio-Temporal Koopman Decomposition. *JOURNAL OF NONLINEAR SCIENCE*, 28(5):1793–1842, OCT 2018.
19. M. A. Khodkar and Pedram Hassanzadeh. Data-driven reduced modelling of turbulent Rayleigh-Benard convection using DMD-enhanced fluctuation-dissipation theorem. *JOURNAL OF FLUID MECHANICS*, 852, AUG 6 2018.
20. Dimitrios Giannakis, Anastasiya Kolchinskaya, Dmitry Krasnov, and Joerg Schumacher. Koopman analysis of the long-term evolution in a turbulent convection cell. *JOURNAL OF FLUID MECHANICS*, 847:735–767, JUL 25 2018.
21. Stefan Klus, Feliks Nuske, Peter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schuette, and Frank Noe. Data-Driven Model Reduction and Transfer Operator Approximation. *JOURNAL OF NONLINEAR SCIENCE*, 28(3):985–1010, JUN 2018.
22. Senka Macesic, Nelida Crnjari-Zic, and Igor Mezić. Koopman Operator Family Spectrum for Nonautonomous Systems. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 17(4):2478–2515, 2018.
23. Zlatko Drmac, Igor Mezić, and Ryan Mohr. DATA DRIVEN MODAL DECOMPOSITIONS: ANALYSIS AND ENHANCEMENTS. *SIAM JOURNAL ON SCIENTIFIC COMPUTING*, 40(4):A2253–A2285, 2018.
24. Erik M. Bollt, Qianxiao Li, Felix Dietrich, and Ioannis Kevrekidis. On Matching, and Even Rectifying, Dynamical Systems through Koopman Operator Eigenfunctions. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 17(2):1925–1960, 2018.
25. Maziar S. Hemati, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. *THEORETICAL AND COMPUTATIONAL FLUID DYNAMICS*, 31(4):349–368, AUG 2017.
26. Jia-Chen Hua, Farzad Noorian, Duncan Moss, Philip H. W. Leong, and Gemunu H. Gunaratne. High-dimensional time series prediction using kernel-based Koopman mode regression. *NONLINEAR DYNAMICS*, 90(3):1785–1806, NOV 2017.
27. Alessandro Alla and J. Nathan Kutz. NONLINEAR MODEL ORDER REDUCTION VIA DYNAMIC MODE DECOMPOSITION. *SIAM JOURNAL ON SCIENTIFIC COMPUTING*, 39(5):B778–B796, 2017.
28. J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. DYNAMIC MODE DECOMPOSITION Data-Driven Modeling of Complex Systems Preface. In *DYNAMIC MODE DECOMPOSITION: DATA-DRIVEN MODELING OF COMPLEX SYSTEMS*, volume 149 of *Other Titles in Applied Mathematics*, pages IX+. 2016.
29. J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. In *DYNAMIC MODE DECOMPOSITION: DATA-DRIVEN MODELING OF COMPLEX SYSTEMS*, volume 149 of *Other Titles in Applied Mathematics*. 2016.
30. Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic Mode Decomposition with Control. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 15(1):142–161, 2016.
31. Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *JOURNAL OF NONLINEAR SCIENCE*, 25(6):1307–1346, DEC 2015.
32. Benjamin Peherstorfer and Karen Willcox. Dynamic data-driven reduced-order models. *COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING*, 291:21–41, JUL 1 2015.
33. Shaowu Pan and Karthik Duraisamy. Data-Driven Discovery of Closure Models. *SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS*, 17(4):2381–2413, 2018.
34. Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.
35. Qianxiao Li, Felix Dietrich, Erik M. Bollt, and Ioannis G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.
36. Erik M. Bollt, Qianxiao Li, Felix Dietrich, and Ioannis Kevrekidis. On matching, and even rectifying, dynamical systems through koopman operator eigenfunctions. *SIAM Journal on Applied Dynamical Systems*, 17(2):1925–1960, 2018.

37. Sören Hanke, Sebastian Peitz, Oliver Wallscheid, Stefan Klus, Joachim Böcker, and Michael Dellnitz. Koopman operator-based finite-control-set model predictive control for electrical drives, 2018.
38. Andrew Kurdila and Parag Bobade. Koopman theory and approximation spaces. *Arxiv*, 2018.
39. Stefan Klus, Peter Koltai, and Christof Schutte. On the numerical approximation of the perron–frobenius and koopman operator. *arXiv*, 2016.
40. Omer San, Romit Maulik, and Mansoor Ahmed. An artificial neural network framework for reduced order modeling of transient flows. *COMMUNICATIONS IN NONLINEAR SCIENCE AND NUMERICAL SIMULATION*, 77:271–287, OCT 2019.
41. Milan Korda and Igor Mezic. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28:687–710.
42. Stefan Klus, Ingmar Schuster, and Krikamol Muandet. Eigendecompositions of transfer operators in reproducing kernel hilbert spaces. *Nonlinear Science*, 30:283–315.
43. Suddhasattwa Das, Dimitrios Giannakis, and Joanna Slawinska. Reproducing kernel hilbert space compactification of unitary evolution groups. *arXiv*, 2019.
44. Suddhasattwa Das and Dimitrios Giannakis. Koopman spectra in reproducing kernel hilbert spaces. *arXiv preprint, arXiv:1801.07799v8*, 2019.
45. Romeo Alexander and Dimitrios Giannakis. Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *arXiv*, 2019.
46. Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*.
47. L. Györfy, M. Kohler, A. Kryzak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin, 2002.
48. Ronald DeVore, Gerard Kerkycharian, Dominique Picard, and Vladimir Temlyakov. Approximation methods for supervised learning. *Foundations of Computational Mathematics*, 6(1):3–58, February 2006.
49. V.N. Temlyakov. Approximation in learning theory. *Constructive Approximation*, 27(1):33–74, 2008.
50. Steve Smale and Ding-Xuan Zhou. Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26(2):153–172, August 2007.
51. Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39(1):1–49, 2001.
52. Hassan K Khalil. *Nonlinear systems*. Upper Saddle River, 2002.
53. Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, 2006.
54. Parag Bobade, Suprotim Majumdar, Savio Pereira, Andrew J Kurdila, and John B Ferris. Adaptive estimation for nonlinear systems using reproducing kernel hilbert spaces. *Advances in Computational Mathematics*, 45(2):869–896, 2019.
55. Parag Bobade, Suprotim Majumdar, Savio Pereira, Andrew J Kurdila, and John B Ferris. Adaptive estimation in reproducing kernel hilbert spaces. In *2017 American Control Conference (ACC)*, pages 5678–5683. IEEE, 2017.
56. Jia Guo, Sai Tej Paruchuri, and Andrew J Kurdila. Persistence of excitation in continuously embedded reproducing kernel hilbert space. 2020.
57. Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.
58. Ernesto De Vito, Lorenzo Rosasco, and Alessandro Toigo. Learning sets with separating kernels. *Applied and Computational Harmonic Analysis*, 37(2):185–217, 2014.
59. Edward Fuselier and Grady B Wright. Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. *SIAM Journal on Numerical Analysis*, 50(3):1753–1776, 2012.
60. Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
61. Saburo Saitoh and Yoshihiro Sawano. *Theory of reproducing kernels and applications*. Springer, 2016.
62. Vern I Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge University Press, 2016.
63. Thomas Hangelbroek, Francis J Narcowich, and Joseph D Ward. Kernel approximation on manifolds i: bounding the lebesgue constant. *SIAM Journal on Mathematical Analysis*, 42(4):1732–1760, 2010.
64. Thomas Hangelbroek, F Narcowich, Christian Rieger, and J Ward. An inverse theorem for compact lipschitz regions in  $\mathbb{R}^d$  using localized kernel bases. *Mathematics of Computation*, 87(312):1949–1989, 2018.
65. Thomas Hangelbroek, Francis J Narcowich, and Joseph D Ward. Polyharmonic and related kernels on manifolds: interpolation and approximation. *Foundations of Computational Mathematics*, 12(5):625–670, 2012.
66. Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2005.
67. Franco Brezzi and Michel Fortin. *Mixed and hybrid finite element methods*, volume 15. Springer Science & Business Media, 2012.
68. Jia Guo, Sai Tej Paruchuri, and Andrew J Kurdila. Approximations of the reproducing kernel hilbert space (rkhs) embedding method over manifolds. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1596–1601. IEEE, 2020.
69. Holger Wendland. Multiscale analysis in sobolev spaces on bounded domains. *Numerische Mathematik*, 116(3):493–517, 2010.
70. Ronald A DeVore and George G Lorentz. *Constructive approximation*, volume 303. Springer Science & Business Media, 1993.
71. JA Walker. Abstract dynamical systems and evolution equations. In *Dynamical Systems and Evolution Equations*, pages 85–136. Springer, 1980.
72. Stephen H Saperstone. *Semidynamical systems in infinite dimensional spaces*, volume 37. Springer Science & Business Media, 2012.
73. Parag Bobade, Dimitra Panagou, and Andrew J Kurdila. Multi-agent adaptive estimation with consensus in reproducing kernel hilbert spaces. In *2019 18th European Control Conference (ECC)*, pages 572–577. IEEE, 2019.
74. Andrew J. Kurdila, Jia Guo, and Sai Tej Paruchuri. Persistence of excitation in reproducing kernel hilbert spaces, positive limit sets, and smooth manifolds. 2019.
75. G. P. Szegő and N.P. Bhatia. *Stability Theory of Dynamical Systems*. Springer, 2002.
76. Emmanuel Hebey. *Sobolev Spaces on Riemannian Manifolds*. Springer-Verlag, 1996.
77. Hans Triebel. *Theory of Function Spaces, Volume 2*. Birkhauser, 1992.
78. Ernesto De Vito, Nicole Mücke, and Lorenzo Rosasco. Reproducing kernel hilbert spaces on manifolds: Sobolev and diffusion spaces. 2019.
79. Colin Bennett and Robert C Sharpley. *Interpolation of operators*. Academic press, 1988.
80. R. A. Adams and John Fournier. *Sobolev spaces*, volume 140. Elsevier, 2003.