2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

# Convergence Rates for Approximations of Deterministic Koopman Operators via Inverse Problems

Nathan Powell Ali Bouland John Burns Andrew Kurdila

***Abstract*—This paper derives rates of convergence of approximations of the deterministic Koopman operator using a framework based on estimating solutions of inverse problems. By restricting the domain of the Koopman operator, simple sufficient conditions are derived that ensure that the resulting Koopman operator is compact when acting on a suitable reproducing kernel Hilbert space (RKHS). Approximations of the Koopman operator, or its inverse, are derived in terms of Galerkin approximations of solutions to an associated inverse problem which depends on noisy data. The resulting bounds on accuracy of approximations to the Koopman operator then take a classical form: we obtain explicit representations of the contributions of the approximation error and the generalization error in terms of the reduced dimension and noise level. As the reduced dimension of the approximations increases, the approximation error decreases, while the generalization error increases. The generalization error increases with an increase in the noise level. In the case of a discrete evolution over a smooth, compact, connected, Riemannian manifold, we show that these two contributions to the error can be bounded in terms of the fill distance of centers of approximation and samples in the manifold.**

***Index Terms*—Koopman theory, inverse problems, nonlinear estimation**

## I. Introduction

This paper studies Koopman operators that are associated with the discrete, generally nonlinear system defined by the recursion

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

$$y_{n+1} = g(\phi_{n+1}) := (Ug)(\phi_n), \tag{2}$$

where $X$ is the state space, the state $\phi_n \in X$, the state propagation function $f : X \to S := f(X) \subseteq X$, the measurement $y_n \in \mathbb{R}$, and the function $g : X \to \mathbb{R}$. The Koopman operator $U := U_f : \mathcal{G} \to \mathcal{H}$ associated with the above discrete dynamics is defined to be the composition operator

$$Ug := g \circ f, \tag{3}$$

where $\mathcal{G}$ and $\mathcal{H}$ are function spaces. Although the Koopman operator $U := U_f$ depends on the fixed function $f$ that appears in the evolution law, we omit the subscript in this paper. In the literature on Koopman operators the function $g$ is referred to as the observable function, or simply the observable. Throughout this paper we assume that $(X, d_X)$ is a complete metric space. Examples where $X$ is the Euclidean space $X := \mathbb{R}^d$ are considered, as are some instances when $X$ is a smooth, compact, connected, Riemannian manifold. In cases where $X$ is not compact, as in the case when $X = \mathbb{R}^d$, we eventually restrict the analysis to a compact invariant subset of $X$.

There are many applications where the study of the Koopman operator $U$ yields important insights into the qualitative or quantitative properties of the original nonlinear dynamical system in (1). See [7], [31], [33] for good overviews of the general theory and several applications. To motivate this paper we just consider an iconic problem from forecasting or output prediction. Suppose we want to predict the output $y_{n+i}$ at the future discrete time $n + i$ from the current state $\phi_n$ at time $n$. Combining (1) and (2) above we find that

$$y_{n+i} = (U^i g)(\phi_n) = (g \circ \underbrace{f \circ \cdots \circ f}_{i \text{ times}})(\phi_n) = g(\phi_{n+i}).$$

This means that if we know or can build an approximation of the Koopman operator $U$, it can be used to estimate or predict the future output $y_{n+i}$ of the system from the current state $\phi_n$. The recursion above is said to determine a forward orbit $\{g_k\}_{k\in\mathbb{N}_0}$ of observable functions in some selected function space for the linear dynamic system $g_{n+1} = Ug_n$ with $g_0 := g$.

Even if the original nonlinear system governed by (1) evolves in a subset of a finite dimensional space $X$, the linear system $g_{n+1} = Ug_n$ for the observable functions is infinite dimensional. Any practical implementation or algorithm derived in terms of the Koopman operator must address how to construct approximations of the Koopman operator. This topic has been the focus of numerous investigations over the past few years. These include [2], [3], [6], [9], [11], [15]–[17], [22], [24], [27], [30], [32], [34], [36], [41], among others. The theoretical differences among these papers can be subtle. This paper is the first among the above that uses the theory of inverse problems for deterministic Koopman operators to derive error bounds that (1) expressly include noise contributions and (2) that are explicit in the reduced dimension $N$ and number of samples $M$ used for constructing approximations. Of all the above, the very recent work in [2], [28] seems most similar in spirit in that they too attack the problem as a type of inverse problem. But they consider stochastic systems and employ stochastic definitions of the Koopman operator. In addition, our analysis and approach makes explicit the role of the direct (Jackson) and

Andrew Kurdila, Ali Bouland, and Nathan Powell are with the Mechanical Engineering Department, and John Burns is with the Department of Mathematics, at Virginia Tech.

Corresponding author is Nathan Powell, nrpowell@vt.edu.

inverse (Bernstein) inequalities in producing error bounds for Koopman approximations, while theirs does not.

There is another way that the approach in this paper differs in structure from the above studies. In the definition in (3) above, the Koopman operator acts between function spaces $\mathcal{G}$ and $\mathcal{H}$. The functions contained in $\mathcal{G}$ and $\mathcal{H}$ in general will have different domains. Functions in $\mathcal{G}$ are defined over $S = f(X)$, while those in $\mathcal{H}$ are defined on all of $X$. Hypotheses that define a common setting can be found in the papers [2], [8], [12], [16], [17], [26]–[28], [42]. These assume that $f$ is nonsingular and measure preserving so that observable functions evolve in the Hilbert space $L^2_\mu(X)$, with $U : L^2_\mu(X) \to L^2_\mu(X)$. Although these assumptions simplify the configuration space of the orbit $\{g_k\}_{k\in\mathbb{N}_0}$ and make $U$ a linear isometry, they also imply the Koopman operator $U : L^2_\mu(X) \to L^2_\mu(X)$ is not compact. This can complicate approximations, as noted in [28] for stochastic problems, if it is desired to use eigenfunctions as the basis of approximations. The approach in this paper formulates the approximation of the Koopman operator in terms of different function spaces for its domain and range. The goal is to develop rates of convergence of Koopman approximations for single step predictions that are general, while minimizing assumptions about the function $f$ or the measure $\mu$ that describe the discrete dynamics and Koopman operator.

## II. Summary of New Results

### A. Theoretical Results

This paper examines approximations of the Koopman operator as a solution to an inverse problem. We use the symbol $\mathbb{U}$ to represent the Koopman operator $U$ for a very special choice of the domain and range function spaces $\mathcal{G}$ and $\mathcal{H}$. We specifically outline conditions that ensure the Koopman operator $\mathbb{U} : R(S) \to L^2_\mu(X)$ is compact. Here $R(S)$ is the restriction to $S$ of functions in a reproducing kernel Hilbert space (RKHS) $H(X)$ defined over the whole configuration space $X$. Two equations,

$$\mathbb{U}g = y \quad \text{and} \quad \mathbb{U}g^\delta = y^\delta, \tag{4}$$

are introduced to study approximations of the Koopman operator $\mathbb{U}$ from the viewpoint of inverse problems. Here $g$ solves the noise-free equation, and $g^\delta$ solves the noisy equation when using perturbed data $y^\delta$. Approximations of the Koopman operator are defined in terms of Petrov-Galerkin approximations $g^\delta_N$ for both the noisy and noise free case. The analysis in this paper derives the rates of convergence of approximations of the Koopman operator having the following form:

$$\|\mathbb{U}g^\delta_N - \mathbb{U}g\|_{L^2_\mu(X)} \lesssim \mathcal{O}\left(\frac{1}{\rho_B(N)}\delta + \rho_J(N)\right).$$

In this equation $\rho_B(N)$ and $\rho_J(N)$ are rate functions that converge to zero as $N \to \infty$. The rate functions $\rho_J$ and $\rho_B$ are associated with the Bernstein and Jackson inequalities, respectively, that hold for the subspaces of approximants. This bound for approximation of Koopman operators from noisy data has the general structure commonly found in inverse problems [13], [23], [29]. The term $\delta/\rho_B(N)$ grows, while the term $\rho_J(N)$ decreases, as $N \to \infty$. As we carefully explain later in the paper, the uncertainty $\delta$ can arise from using $M$ quadrature points to approximate integrals that arise in the equations that form Petrov-Galerkin approximations. In practice then, we obtain error bounds that are explicit in the reduced spatial dimension $N$ and in the number of samples $M$. We also show that under the additional assumption that the subset $S$ is a Riemmannian manifold $\mathcal{M}$ of dimension $d$, we can obtain more precise rates of convergence in terms of the fill distance $h_{\Xi_N,\mathcal{M}}$ of samples $\Xi_N$ over the manifold $\mathcal{M}$,

$$\|\mathbb{U}g^\delta_N - \mathbb{U}g\|_{L^2_\mu(X)} \lesssim \mathcal{O}\left(\frac{1}{h^r_{\Xi_N,\mathcal{M}}}\delta + h^p_{\Xi_N,\mathcal{M}}\|g\|_{H^p(\mathcal{M})}\right),$$

for exponents $r$ and $p$ that relate to the smoothness of the kernel and smoothness of the function $g$. Note again that the first term $\frac{1}{\rho_B(N)}\delta$ grows as $N \to \infty$, and the second term decreases, $h^r_{\Xi_N,\mathcal{M}} \to 0$ as $N \to \infty$. The latter bound depends crucially on Bernstein inequalities derived very recently for types of smooth Riemannian manifolds and their subsets [20], [21].

We demonstrate this qualitative behavior in Figure (1).

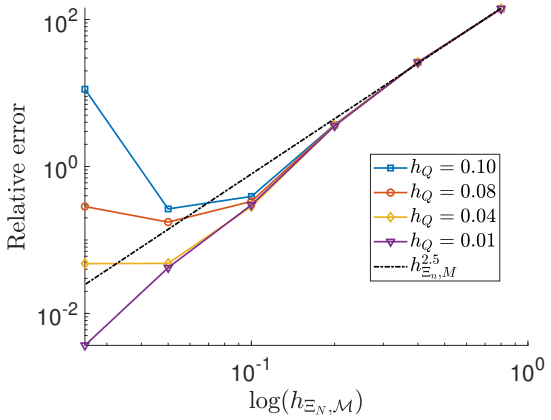


Fig. 1. A depiction of the rates of convergence of approximation of the Koopman operator. On the right of the plot, over relatively larger fill distances, the error is dominated by the approximation error and thus decreases at the expected rate associated with the Jackson inequality rate function $\rho_J(N) \approx \mathcal{O}(h^{2.5}_{\Xi,\mathcal{M}})$. As the fill distances decrease toward the left region of the plot, the error bound is dominated by the Bernstein inequality term $\frac{1}{\rho_B(N)}\delta$. This generalization error grows with larger perturbations characterized by larger quadrature errors. In this figure $h_Q$ is the fill distance of the point quadrature centers, so that $\delta \sim O(1/h_Q)$.

### B. Practical Considerations

We also discuss how the theoretical rates discussed above can be related to the popular extended dynamic mode decomposition (EDMD) method [24], [26], [42] when kernel bases are used to approximate the Koopman operator. When quadrature rules are used to approximate the Petrov-Galerkin approximations in this paper, the resulting Equations (17) are *exactly* the form generated by the EDMD algorithm. As a consequence, the theory above can be viewed as a means of generating error bounds for the EDMD algorithm when

kernel bases are used. The EDMD equations depend on the reduced dimension $N$ and on the number of quadrature points $M$. We can interpret the effect of increasing the number of quadrature points $M$ as analogous to increasing the number of snapshots $M$ used in the EDMD algorithm. The qualitative form of the error plot in Figure (1), which is so common in inverse problems, suggests an important principle in using the EDMD algorithm. The reduced dimension $N$ and the number of snapshots $M$ when using the EDMD algorithm should be selected in a way that balances or equilibrates the two contributing sources of error. The qualitative error behavior, and its implications, is well-known in the study of inverse problems [13], [23].

## III. Symbols, Notation, Background

The symbols $\mathbb{R}$ and $\mathbb{R}^+$ denote the real numbers and nonnegative real numbers, respectively, and we denote the positive integers by $\mathbb{N}$. If $U \subseteq V$ for two normed vector spaces $U$ and $V$, we say that the inclusion is continuous if there is a constant $c > 0$ such that $\|g\|_V \leq c\|g\|_U$ for all $g \in U$. When $U$ is continuously embedded in $V$, we denote this relationship as $U \hookrightarrow V$. We say that the $U$ is densely embedded in $V$ if the inclusion $U \subseteq V$ is continuous and in addition the set $U$ is dense in $V$ in the norm on $V$. When $U$ is densely embedded in $V$ we depict this relationship as $U \hookrightarrow\!\!\!\!\rightarrow V$. If several such relationships hold for different spaces and we want to be explicit, we write $U \overset{i}{\hookrightarrow} V$ or $U \overset{i}{\hookrightarrow\!\!\!\!\rightarrow} V$ where $i : u \in U \mapsto u = i(u) \in V$ is the canonical injection of $U$ into $V$.

### A. Reproducing Kernel Hilbert Spaces

In this paper the mapping $\mathfrak{K} : X \times X \to \mathbb{R}$ is an admissible kernel that induces the reproducing kernel Hilbert space (RKHS) $(H(X), (\cdot,\cdot)_H)$ of real-valued functions over the set $X$. The theory and definitions for the RKHS $H(X)$ holds for a general set $X$, and further assumptions are made on $X$ as they are needed. The RKHS $H(X)$ is also commonly known as the native space generated by the kernel $\mathfrak{K}$. When we say that the real-valued kernel $\mathfrak{K}$ is admissible, we mean that it is continuous, symmetric, and positive semidefinite. Since the kernel $\mathfrak{K}$ is continuous, it is also known as a Mercer kernel in the theory of native spaces. It is positive semidefinite in the sense that for any collection of $N$ points $\{x_1, \cdots, x_N\} \subset X$ the matrix $[\mathfrak{K}(x_i, x_j)]$ is positive semidefinite. The kernel $\mathfrak{K}$ is positive definite if the matrix $[\mathfrak{K}(x_i, x_j)]$ is positive definite for any selection of $N$ points. We denote by $\mathfrak{K}_x(\cdot) := \mathfrak{K}(x, \cdot)$ the kernel section centered at $x \in X$. Since kernel sections are often used to build approximations, we also refer to the kernel section $\mathfrak{K}_x : X \to \mathbb{R}$ as the kernel basis function centered at $x \in X$. The native space $H(X)$ is the closed linear span of the kernel sections, $H(X) = \overline{\text{span}\,\{\mathfrak{K}_x \mid x \in X\}}$. Intuitively, we can view $H(X)$ as the space that is the superposition of kernel basis functions $\mathfrak{K}_x$, which are defined in terms of the fixed template $\mathfrak{K}_x$, as $x$ is moved throughout $X$. Often, the kernel section $\mathfrak{K}_x$ is an example of a radial basis function [40]. When $H(X)$ is such a native space, any closed subspace $V \subseteq H(X)$ is also an RKHS for the kernel $\mathfrak{K}_V : X \times X \to \mathbb{R}$ defined by $\mathfrak{K}_V(x, y) := (\Pi_V \mathfrak{K}_x, \Pi_V \mathfrak{K}_y)_H$ where $\Pi_V : H(X) \to V$ is the $H(X)$-orthogonal projection onto $V$ [5].

### B. Function Spaces

For any metric space $Z$ the vector space of continuous, real-valued functions on $Z$ is represented by $C(Z)$. We write $C_b(Z)$ for the Banach space of bounded continuous functions over $Z$ with the uniform norm $\|f\|_{C_b(Z)} := \sup_{z \in Z} |f(z)|$. In the usual way, when $(Z, \alpha)$ is a measure space, we define the Lebesgue space $L^p_\alpha(Z)$ to be the Banach spaces of real-valued functions over $Z$ having norms

$$\|f\|_{L^p_\alpha(Z)} := \begin{cases} \sup_{z \ \alpha\text{-a.e. } \in Z} |f(z)| & \text{if} \quad p = \infty, \\ \left(\int_Z |f(z)|^p \alpha(dz)\right)^{1/p} & \text{if} \quad 1 \leq p < \infty. \end{cases}$$

When the measure $\alpha$ is Lebesgue measure, we just write $L^2(Z) := L^2_\alpha(Z)$.

This paper also makes systematic use of certain types of real Sobolev spaces. For a Lebesgue measurable subset $S \subset \mathbb{R}^d$, the Sobolev space $W^m(S)$ for an integer $m > 0$ is defined to be the collection of all functions $g \in L^2(S)$ such that the norm below is finite,

$$\|f\|^2_{W^m(S)} := (f, f)_{W^m(S)} := \sum_{|\alpha|=0}^{m} (D^\alpha f, D^\alpha f)^2_{L^2(S)} < \infty, \tag{5}$$

where the sum is over all nonnegative multiindices $\alpha := [\alpha_1, \alpha_2, \ldots, \alpha_d] \in \mathbb{N}^d$ with $|\alpha| = \sum_{i=1}^d |\alpha_i|$, and

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

The Sobolev space $W^r(S)$ for $r \in \mathbb{R}^+$ is defined as the interpolation space between $W^m(S)$ and $W^{m+1}(S)$ for $m = \lfloor r \rfloor \in \mathbb{N}$, and we denote the by $W^{-r}(S) := (W^r(S))^*$ the topological dual space of $W^r(S)$ [1], [10].

Sobolev spaces $W^r(\mathcal{M})$ over smooth, connected, compact Riemannian manifolds $\mathcal{M}$ are defined in similar fashion, but in a way that is intrinsic to the manifold $\mathcal{M}$ [4], [39]. Suppose that $\nabla$ is the connection or covariant derivative operator, $\Delta$ is the associated Laplacian operator, and $\mu$ is the volume measure on the manifold $\mathcal{M}$. The Sobolev space $W^r(\mathcal{M})$ consists of all the functions in $L^2_\mu(\mathcal{M})$ such that $(I + \Delta)^{r/2} f \in L^2_\mu(\mathcal{M})$. The norm is given by

$$\|f\|^2_{W^r(\mathcal{M})} := (f, f)_{W^r(\mathcal{M})} = \|(I + \Delta)^{r/2} f\|^2_{L^2_\mu(\mathcal{M})}.$$

As above, the negative index Sobolev space $W^{-r}(\mathcal{M})$ is defined to be the dual space $W^{-r}(\mathcal{M}) = (W^r(\mathcal{M})^*$ for $r \geq 0$. When $r$ is a positive integer this definition can be interpreted intuitively like the Sobolev spaces $W^r(S)$ above for $S \subset \mathbb{R}^d$. The space $W^m(\mathcal{M})$ for a positive integer $m > 0$ consists of all functions in $L^2_\mu(\mathcal{M})$ such that their covariant derivatives through order $m$ are contained in $L^2_\mu(\mathcal{M})$. In this case the norm above is equivalent to

$$\|f\|^2_{W^m(\mathcal{M})} := \sum_{0 \leq k \leq m} (\nabla^k f, \nabla^k f)^2_{L^2_\mu(\mathcal{M})},$$

and this expression reduces to the expression above in (5) when $X = \mathbb{R}^d = \mathcal{M}$ [19], [39].

## IV. Koopman Operator Theory

### A. The Pullback Space and Space of Restrictions

Let the admissible kernel $\mathfrak{K} : X \times X \to \mathbb{R}$ define the RKHS space $H(X)$ of real-valued functions over the set $X$. Let $S$ be the range of $f : X \to S \subset X$, that is, $S = f(X)$. We define the kernel $\mathfrak{R} := \mathfrak{K}|_{S\times S}$ as the restriction of $\mathfrak{K}$ to $S \times S$. The copies of the kernel section $\mathfrak{R}_s := \mathfrak{R}(s, \cdot)$ as $s$ varies over $S$ defines the RKHS $R(S)$ of functions that are defined over $S$,

$$R(S) := \overline{\text{span}\,\{\mathfrak{R}_s \mid s \in S\}}.$$

The pullback space $P(X)$ is the range of the Koopman operator, with $P(X) = U(R(S))$. The pullback kernel $p : X \times X \to \mathbb{R}$ that defines the RKHS $P(X)$ is given by

$$p(x, y) = \mathfrak{R}(f(x), f(y)) := \mathfrak{K}|_{S\times S}(f(x), f(y))$$

for all $x, y \in X$. The pullback kernel $p$ defines an inner product on $P(X)$ that induces a norm that is equivalent to the expression

$$\|g\|_P \approx \inf\{\|r\|_R \mid g = r \circ f = Ur, r \in R(S)\}.$$

An immediate consequence of this fact is that $U : R(S) \to P(X)$ is a nonexpansive bounded linear operator.

Some of the most basic, but important, properties of the pullback space $P(X)$, the space of restrictions $R(S)$, and the Koopman operator $U$ are summarized in the following theorem.

**Theorem 1.** *With the definitions above, the adjoint $U^* : P(X) \to R(S)$ is an isometric isomorphism. Let $f : X \to S$ be continuous, and suppose that $(X, \mu)$ and $(S, \nu)$ are finite measure spaces. If the kernel $\mathfrak{K} : X \times X \to \mathbb{R}$ is uniformly bounded on the diagonal in the sense that there is a positive constant $\bar{k}$ such that*

$$\mathfrak{K}(x, x) = (\mathfrak{K}_x, \mathfrak{K}_x)_H = \|\mathfrak{K}_x\|_H \le \bar{k}^2, \quad \textit{for all } x \in X,$$

*then we have the continuous inclusions*

$$P(X) \hookrightarrow C(X) \hookrightarrow\!\!\!\rightarrow L^2_\mu(X),$$
$$R(S) \hookrightarrow C(S) \hookrightarrow\!\!\!\rightarrow L^2_\nu(S).$$

*Proof.* When $\mu, \nu$ are finite measures, it is well known that $C(X) \hookrightarrow L^2_\mu(X)$ and $C(S) \hookrightarrow L^2_\mu(S)$, and these embeddings are compact. The fact that $U^*$ is an isometric isomorphism between $R(S)$ and $P(X)$ can be deduced from arguments in a few standard references such as in Section 5.4 of [35] or Section 2.2.2 of [38]. Using the fact that $k(.,.)$ is bounded in the diagonal, it is immediate that that $P(X) \hookrightarrow C(X)$ and $R(S) \hookrightarrow C(S)$. □

## V. A Compact Koopman Operator $\mathbb{U}$

The analysis in this paper is made possible by restricting the domain of the Koopman operator in a way that ensures the operator is compact. Let $i, j$ be the canonical injections $P(X) \overset{i}{\hookrightarrow} C(X)$ and $C(X) \overset{j}{\hookrightarrow\!\!\!\rightarrow} L^2_\mu(X)$. We define the Koopman operator $\mathbb{U}$ as

$$\mathbb{U}g := j \circ i \circ Ug \quad \text{for all } g \in R(S).$$

When we unroll all these operators, the definition of $\mathbb{U}$ has the same functional form as that of $(Ug)(x) = g(f(x))$, with

$$(\mathbb{U}g)(x) = g(f(x)) \quad \text{for all } x \ \mu\text{-a.e.} \in X,$$

but $\mathbb{U} : R(S) \to L^2_\mu(X)$. In other words, $\mathbb{U}$ is the Koopman operator that results when viewing $U$ as a mapping from $R(S)$ into $L^2_\mu(X)$. Based on the result of the theorem below, we have the fact that the operator $\mathbb{U}$ is now a compact operator whose adjoint is an integral operator.

**Theorem 2.** *Suppose that the hypotheses of Theorem (1) hold and that $X$ is compact. The operator $\mathbb{U} : R(S) \to L^2_\mu(X)$ is linear, and compact. Its adjoint $\mathbb{U}^* : L^2_\mu(X) \to R(S)$ is given by the integral equation*

$$(\mathbb{U}^*h)(s) := \int_X \mathfrak{K}(s, f(\xi))h(\xi)\mu(d\xi) \quad \textit{for all } s \in S.$$

*The adjoint $\mathbb{U}^*$ is determined by an integral operator defined in terms of the unsymmetric kernel $u(s, x) := \mathfrak{K}(s, f(x))$, with $u(\cdot, \cdot) : S \times X \to \mathbb{R}$.*

*Proof.* The Koopman operator $U$ is a bounded operator from $R(S)$ into $P(X)$. But, as discussed in [37], the canonical injection $j \circ i$ of $P(X)$ into $L^2_\mu(X)$ is compact since the pullback kernel is bounded uniformly on the diagonal. Since $\mathbb{U}$ is the product of bounded operators and a compact operator, it is compact. Next we turn to deriving a formula for the adjoint $\mathbb{U}^*$. First note that for any $h \in L^2_\mu(X)$, the $R(S)$-valued function

$$\xi \mapsto \mathfrak{R}(\cdot, f(\xi))h(\xi) \in R(S) \quad \text{for all } \xi \in X$$

satisfies the integral bound

$$\int \|\mathfrak{R}(\cdot, f(\xi))h(\xi)\|^2_{R(S)}\mu(d\xi) \le \bar{k}\int |h(\xi)|^2\mu(d\xi).$$

This means that for any $h \in L^2_\mu(\xi)$, the Bochner integral

$$\int \mathfrak{R}(\cdot, f(\xi))h(\xi)\mu(d\xi) = \int \mathfrak{K}_{f(\xi)}h(\xi)\mu(d\xi)$$

is a well-defined element of $R(S)$. By the definition of the adjoint $\mathbb{U}^*$ of $\mathbb{U}$, we have

$$\begin{aligned}(\mathbb{U}g, h)_{L^2_\mu(X)} &= \int g(f(\xi))h(\xi)\mu(d\xi),\\ &= \int \left(g, \mathfrak{K}_{f(\xi)}\right)_{R(S)} h(\xi)\mu(d\xi),\\ &= \left(g, \int \mathfrak{K}_{f(\xi)}h(\xi)\mu(d\xi)\right)_{R(S)} = (g, \mathbb{U}^*h)_{R(S)}.\end{aligned}$$

The last two lines follow from the fact that the Bochner integral commutes with a bounded linear operator. □

## VI. Koopman Inverse Problems

In this section we build approximations $g_N$ of $g$, where $g$ is the solution of the operator equation

$$\mathbb{U}g = y, \tag{6}$$

or of the related operator equation

$$\mathbb{T}g := \mathbb{U}^*\mathbb{U}g = h, \tag{7}$$

for $h := \mathbb{U}^*y$. We can view such approximations $g_N$ as approximations of the action of the pseudoinverse $\mathbb{U}^\dagger g$.

### A. Petrov-Galerkin Approximations

As noted in the introduction, a few references [24]–[26], [42] have used the fact that the well-known extended mode dynamic decomposition (EDMD) algorithm can be understood as the equations resulting from a Petrov-Galerkin approximation of the Koopman operator. Here we analyze such approximations for specific choices of scattered bases in a native space. We let $\Xi_N \subseteq S$ be a set of centers used to define bases. To introduce finite dimensional spaces, we set $\mathcal{U}_N := \{u_{\xi_1}, \ldots, u_{\xi_N}\}$ and $W_N := \mathbb{U}^*(\mathcal{U}_N) := \text{span}\{w_{\xi_1}, \ldots, w_{\xi_N}\}$ where

$$\mathfrak{R}_{\xi_n}(\cdot) := \mathfrak{R}(\xi_n, \cdot) = \mathfrak{K}(\xi_n, \cdot) \in R(S), \tag{8}$$

$$u_{\xi_n}(\cdot) := \mathfrak{R}(\xi_n, f(\cdot)) = (\mathbb{U}\mathfrak{R}_{\xi_n})(\cdot) \in L^2_\mu(X), \tag{9}$$

$$w_{\xi_n}(\cdot) := (\mathbb{U}^* u_{\xi_n})(\cdot) \in \mathcal{R}(\mathbb{T}) \subseteq W(S) \subseteq R(S). \tag{10}$$

We then have the inclusions

$$\begin{aligned} &R_N \subseteq R(S), \\ &\mathcal{U}_N \subseteq \mathcal{U} \subseteq L^2_\mu(X), \\ &W_N \subseteq \mathcal{R}(\mathbb{T}) \subseteq W(S) := \mathcal{R}(\mathbb{U}^*). \end{aligned}$$

Associated with these spaces we define families of uniformly bounded projection operators that are used to build approximations. We define the projections $\mathbb{P}_N : R(S) \to R_N$, and when we say that this family is uniformly bounded we mean there is a constant $C > 0$ such that $\|\mathbb{P}_N\| := \|\mathbb{P}\|_{R(S)\to R(S)} \leq C$ for all $N \geq 0$.

We begin by defining the Galerkin [23], [29] approximation $g^\delta_N \in R_N$ of (7) that satisfies

$$g^\delta_N = \left(\mathbb{P}_N \mathbb{T}|_{R_N}\right)^{-1} \mathbb{P}_N h^\delta,$$

which corresponds when $\mathbb{P}_N$ is the $R(S)$-orthogonal projection onto $R_N$ to the variational equations

$$\left(\mathbb{T}g^\delta_N, \mathfrak{R}_{\xi_n}\right)_{R(S)} = \left(h^\delta, \mathfrak{R}_{\xi_n}\right)_{R(S)} \qquad \text{for all } 1 \leq n \leq N.$$

When we write $g^\delta_N(\cdot) := \sum_{n=1}^N \mathfrak{R}_{\xi_n}(\cdot)\theta_n$, this expression generates the matrix equations

$$\begin{aligned} \sum_{j=1}^N \int \mathfrak{K}(\xi_n, f(x))\mathfrak{K}(f(x), \xi_j)\mu(dx)\theta_j = \\ \int \mathfrak{K}(\xi_n, f(x))y(x)\mu(dx) \\ \text{for } 1 \leq n \leq N. \end{aligned} \tag{11}$$

Still another interpretation of these equations is possible. Define $z = \mathbb{U}g$ and seek a solution $z \in L^2_\mu(X)$ of

$$\mathbb{U}^* z = h. \tag{12}$$

The Petrov-Galerkin approximation $z^\delta_N \in \mathcal{U}_N \subset L^2_\mu(X)$ of the solution $z$ of (12) then satisfies the equation

$$z_N := \left(\mathbb{P}_N \mathbb{U}^*|_{\mathcal{U}_N}\right)^{-1} \mathbb{P}_N h^\delta. \tag{13}$$

In variational form, when $\mathbb{P}_N$ is the $R(S)$-orthogonal projection onto $R_N$, the solution $z^\delta_N \in \mathcal{U}_N \subseteq L^2_\mu(X)$ satisfies

$$\left(\mathbb{U}^* z^\delta_N, \mathfrak{R}_{\xi_n}\right)_{R(S)} = \left(h^\delta, \mathfrak{R}_{\xi_n}\right)_{R(S)} \qquad 1 \leq n \leq N, \tag{14}$$

or equivalently,

$$\left(z^\delta_N, u_{\xi_n}\right)_{L^2_\mu(X)} = h^\delta(\xi_n) \qquad 1 \leq n \leq N. \tag{15}$$

These operator equations yield the same matrix equations as in (11). It should also be noted that once the above (15) are solved for $\{\theta_j\}_{j=1}^N$, expressions for $g^\delta_N$ immediately follow since

$$\mathbb{U}g^\delta_N = z^\delta_N = \sum_{j=1}^N \mathfrak{K}(\xi_j, f(\cdot))\theta_j \Rightarrow g^\delta_N = \sum_{j=1}^N \mathfrak{K}(\xi_j, \cdot)\theta_j, \tag{16}$$

where $g^\delta_N$ is an approximation of the solution $g$ of (7).

### B. Quadrature Rules and the EDMD Algorithm

The Petrov-Galerkin equations in (11) are determined by the choice of centers $\Xi_N$ and basis that makes up $R_N$. The integral (11) cannot be implemented directly in many data-driven applications since it depends on the unknown function $f$. We can address this problem by using quadrature rules to approximate the unknown integral, and thereby obtain matrix equations that are realizable from observations alone. The quadrature points are denoted $\{x_m\}_{m=1}^M$, and each $x_m$ is assumed to be contained in a set $\Omega_m$ where the family of sets satisfies $X = \bigcup_{m=1}^M \Omega_m$. In data-driven applications we assume that these quadrature points $\{x_m\}_{m=1}^M$ are gathered along the trajectory of the original system, and that these points fill up a domain of interest over time. Applying the one point quadrature rules to (11) yields the equations

$$\begin{aligned} \sum_{j=1}^N \sum_{m=1}^M \mathfrak{K}(\xi_n, f(x_m))\mathfrak{K}(f(x_m), \xi_j)\mu(\Omega_m)\theta_j \\ = \sum_{m=1}^M \mathfrak{K}(\xi_n, f(x_m))y(x_m)\mu(\Omega_m) \\ \text{for } 1 \leq n \leq N. \end{aligned} \tag{17}$$

It is easy to see that these equations are precisely the same as those generated by the EDMD algorithm for a reduced basis of dimension $N$ and using $M$ snapshots, where in our case the snapshots are the quadrature points.

### C. A General Theorem on Rates of Convergence

As in many discussions of convergence of approximations of solutions to inverse problems, the direct and inverse inequalities of approximation theory [10] play a critical role. These direct and inverse inequalities are also known as the Jackson and Bernstein inequalities, respectively. These inequalities are defined in terms of a rate function $\rho : \mathbb{N} \to \mathbb{R}^+$ that converges to zero as $N \to 0$, $\lim_{N\to\infty} \rho(N) = 0$. In the problem at hand we say that the pair $W(S), R(S)$, the family of finite dimensional spaces $\{W_N\}_{N\in\mathbb{N}} \subset W(S)$, and uniformly bounded projection operators $\mathbb{P}_N : R(S) \to W_N$ satisfies the inverse or Bernstein inequality with the rate function $\rho_B(N) \to 0$ if for any $w_N \in M_N$ we have

$$\|w_N\|_{W(S)} \leq \frac{1}{\rho_B(N)} \|w_N\|_{R(S)}. \tag{18}$$

Let $\Pi_N$ denote the $L^2_\mu(X)$-orthogonal projection onto $\mathcal{U}_N$. We say that the family $\{\mathcal{U}_N\}_{N\in\mathbb{N}} \subseteq \mathcal{U}$ and the pair of Banach spaces $\mathcal{U}, L^2_\mu(X)$ with $\mathcal{U} \hookrightarrow L^2_\mu(X)$ satisfies a direct or Jackson inequality with rate function $\rho_J(N) \to 0$ if we have

$$\|(I - \Pi_N)f\|_{L^2_\mu(X)} \leq \rho_J(N)\|f\|_{\mathcal{U}}. \tag{19}$$

The study of approximation spaces that satisfy these two inequalities is a classical topic of approximation theory. Well-known examples include spline spaces, finite element spaces, piecewise polynomial spaces, and native spaces [10]. The primary results in this paper rely on recent derivations of these inequalities for certain types of native spaces in [18], [19].

The primary convergence results in this paper are derived from the theorem below.

**Theorem 3.** *Suppose that $\mathbb{U}$ and $\mathbb{U}^*$ are injective, $h = \mathbb{U}^* b$, $\|h - h^\delta\|_{R(S)} \leq \delta$, and denote by $g$ and $g^\delta$ the solutions of (12) for the right hand sides $h$ and $h^\delta$, respectively. Let $z^\delta_N = \mathbb{U} g^\delta_N$ in (13) solve the Petrov-Galerkin Equations (15) for the noisy right hand side $h^\delta$ when $\mathbb{P}_N : R(S) \to W_N \subseteq W(S) := \mathcal{R}(\mathbb{U}^*) \subseteq R(S)$ with $\|\mathbb{P}_N\|_{\mathcal{L}(R(S))} \leq C$ for some constant $C > 0$. Finally, suppose that that the Bernstein and Jackson inequalities in (18) and (19), respectively hold. If the Galerkin approximations are convergent, there are two positive constants $C_B, C_J > 0$ such that the following error bounds hold on estimates of the action of the Koopman operator*

$$\|\mathbb{U} g^\delta_N - \mathbb{U} g\|_{L^2_\mu(X)} \leq C_B \frac{1}{\rho_B(N)} \delta + C_J \rho_J(N) \|g\|_{R(S)}.$$

*Proof.* The proof of this theorem is modeled on the proof of Theorem 3.11 of [23], or from the combination of Theorems 17.1 and 17.2 of [29]. Here we have introduced the language of the Jackson and Bernstein inequalities, which are well-known in approximation theory [10]. The full proof is lengthy, and exceeds the limits of this very brief paper. □

## VII. Numerical Examples

In this section, we illustrate the rates of convergence for an evolution law on a manifold $\mathcal{M}$. It is a one-dimensional submanifold of the torus in $\mathbb{R}^3$ with major radius $R_1 = 1$ and minor radius $R_2 = 1/3$. The submanifold $\mathcal{M} \subseteq \mathbb{R}^3$ is parameterized by the following expression

$$\mathcal{M} := \Big\{ (u, v, w) \in \mathbb{R}^3 \mid u = (R_1 + R_2 \cos(6\theta)) \cos(\theta), \\ v = (R_1 + R_2 \cos(6\theta)) \sin(\theta), \\ w = R_2 \sin(6\theta) \Big\} \tag{20}$$

for $\theta \in [0, 2\pi)$, and it is trivial to define a second compatible coordinate chart that covers the submanifold. If we define the time varying parameter $\theta(t) := t$ for $t \in \mathbb{R}^+$, a semigroup $S(t) : \mathcal{M} \to \mathcal{M}$ on $\mathcal{M}$ can be generated using the coordinate chart in (20), along with the second coordinate chart. A corresponding discrete evolution law over $\mathcal{M}$ for time step $h$ can be written in terms of the function $f : \mathcal{M} \to \mathcal{M}$ defined from

$$\phi_{n+1} = f(\phi_n) := S(h)\phi_n. \tag{21}$$

For this example $h = 0.01$ seconds. We define the RKHS using Wendland's compactly supported kernel

$$\mathfrak{K}(r) = (1 - \frac{r}{\ell})^6_+ \left(3 + 18\frac{r}{\ell} + 35\left(\frac{r}{\ell}\right)^2\right) \tag{22}$$

where the radial basis function (RBF) $\mathfrak{K}(r) = \mathfrak{K}(\|\xi - x\|_{\mathbb{R}^D}) = \mathfrak{K}(\xi, x) = \mathfrak{K}_\xi(x)$ can be defined in terms of its center $\xi$. In the problem at hand, the hyperparameter $\ell$ is sometimes referred to as the support radius and corresponds to the "spread" of the kernel. For this example, we selected $\ell = 1.7$. The kernel centers $\Xi_N$ are selected quasi-uniformly along the manifold with a spacing roughly equivalent to a desired fill distance $h_{\Xi_N, \mathcal{M}}$. Generating the desired rates of convergence requires a nested sequence of approximation spaces $H_{N_1}(\mathcal{M}) \subseteq H_{N_2}(\mathcal{M})$. The nested sequential spaces were generated for an increasing number of centers, $N$, by halving the fill distance of the previous approximation space in the sequence.

As mentioned in [40], when defined over $\mathbb{R}^D$, the native RKHS generated by the Wendland kernel above is equivalent to a Sobolev space, with $H(\mathbb{R}^D) \approx W^4(\mathbb{R}^D)$. By

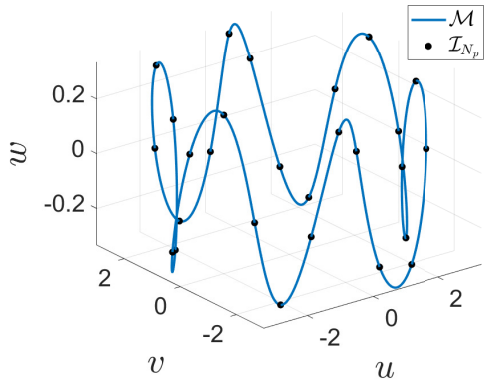


Fig. 2. The submanifold $\mathcal{M}$ defined parametrically by (20) along with the 27 interpolation points used to interpolate the nonlinear function $p(u, v, w)$ with the function $f_s$ of a known smoothness.

restricting it to a $d$-dimensional manifold $\mathcal{M}$, the restricted native space loses smoothness by an amount $\frac{(D-d)}{2}$, so the restricted native space $R(\mathcal{M}) \approx W^{4-(D-d)/2}(\mathcal{M})$. In this example, $D = 3$ and $d = 1$ so that the native space is $H(\mathcal{M}) \approx W^3(\mathcal{M})$.

In order to demonstrate the rates of convergence empirically, the target function to be approximated must be of a particular smoothness. Matérn kernels defined as

$$\kappa_\nu(r) = \sigma^2 \frac{2^{1-(\nu-D/2)}}{\Gamma(\nu - D/2)} r^{(\nu-D/2)} \mathcal{K}_{(\nu-D/2)}(r). \quad (23)$$

are a set of functions that belong to $W^{s_0}(\mathbb{R}^D)$ when $\nu = (s_0 + D/2)/2$. The Matérn-kernel is parameterized by normalization constant $\sigma$ and $\nu$, which defines the order of the modified Bessel function of the second kind $\mathcal{K}_\nu$ and the exponent of the Euclidean norm $\|x_i - x\|_{\mathbb{R}^D}$. Following the numerical example given in [14], we define the target function

$$f_{s_0}(x) = \sum_{i=1}^{N_p} c_i \kappa_{\frac{1}{2}(s_0+D/2)}(\|x - x_i\|_{\mathbb{R}^D}) \quad (24)$$

as a linear combination of Matérn kernels with a finite set of points $\Xi_{N_p} = \{x_1, x_2, ..., x_{N_p}\}$. As described in [14] this function has smoothness that satisfies the strict equality $s_0 = 2\nu - \frac{D-d}{2}$. In this case, we choose $s_0 = 4$. The generated target functions over $\mathcal{M}$ belong to $W^{s,2}(\mathcal{M})$, $s < 2\nu - \frac{D}{2} - \frac{d}{2}$. Note again that when we restrict functions to the manifold $\mathcal{M}$ the functions lose additional smoothness, $d/2$. The coefficients $\{c_i\}_{i=1}^{N_p}$ of the target function are calculated by interpolating the following smooth, nonlinear function

$$p(u, v, w) = \frac{1}{8}(u^5 - 10u^3v^2 + 5uv^4)(u^2 + v^2 - 60w^2) \quad (25)$$

at the points $\Xi_{N_p}$. Here, we chose a set of points that corresponded to a quasi-uniform distribution of $N_p = 27$ points along $\mathcal{M}$ as seen in Figure (2). The target function is plotted over $\mathcal{M}$ in Figure (3).

The analysis of this example now follows from Theorem (3) using the Bernstein and Jackson inequalities for smooth, connected, compact, Riemannian manifolds in [18]–[21] that have appeared relatively recently.

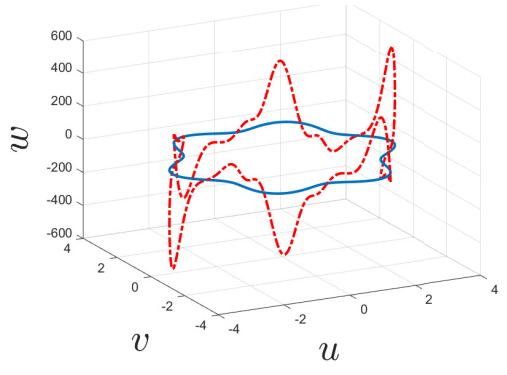


Fig. 3. An illustration of the target function $f_s$ used in the numerical experiment. Here the target function is given by the dashed red line as points, $(u, v, w + f_s)$ over the manifold $\mathcal{M}$ given by the solid blue curve.

The Galerkin approximation is defined in terms of integrals which cannot be calculated because $f$ and the measure $\mu$ are generally unknown. We address this issue by using a one-point quadrature rule with quadrature points $x_m \in \Omega_m$. With this approach, estimates are susceptible to error of order of the fill distance of the quadrature points, which we denote $h_Q$. Here we examined the rates of convergence for a nested sequence of spaces defined by the kernel centers $\Xi_N$ using different quadrature fills. Following the results of Theorem (3), the quadrature error acts as a perturbation $\delta \approx \mathcal{O}(h_Q)$. The rates of convergence for different quadrature errors are given in Figure (1). The relative error is estimated by calculating the maximum error over the orbit evaluated at a resolution equal to the finest selected fill distance. From the figure, we can see that over the larger fill distances to the right half of the plot, the error is dominated by the approximation error $\left\|g - g_N^\delta\right\|$. It decreases at the expected rate associated with the Jackson inequality term $\rho_J(N) \approx \mathcal{O}(h_{\Xi,\mathcal{M}}^{2.5})$. Note that the measured error is an approximation that is lower than the predicted error because the supremum error over the orbit would require the evaluation of an infinite number of points.

For smaller fill distances with a larger number $N$ of kernel centers, the error bound is dominated by the Bernstein inequality term $\frac{1}{\rho_B(N)}\delta$. This qualitative behavior can be observed on the left half of the plot in Figure (1) with the increase in error as the fill distances get smaller. Note that larger quadrature fills $h_Q$ result in larger perturbations $\delta \approx \mathcal{O}(h_Q)$ of the data set. This results in more rapid growth of the dominant error bound $\frac{1}{\rho_B(N)}\delta$ for smaller fill distances as seen on the left side of the plot in Figure (1). Here we also note that the quadrature fill $h_Q$ can be reduced by collecting more quadrature points. This is qualitatively similar to collecting more samples to reducing the variance associated with error from standard approximations from statistical learning theory.

## VIII. Conclusion

In this paper, the problem of approximating a deterministic Koopman operator in a data-driven scenario is investigated. A theoretical framework based on estimating solutions of inverse problems is used to derive rates of convergence of approximations of the Koopman operator. The analysis begins by restricting the domain of the Koopman operator to ensure that the Koopman operator is compact. The Koopman operator $\mathbb{U}$ is approximated using the linear system $\mathbb{U}g = y$ for $y(\cdot) \in L^2_\mu(X)$, where the function $y(x)$ is the system output at the next time step when the state is located at $x \in X$. The method of approximation entails the use of Petrov-Galerkin projection, coupled with the use of a quadrature methods to approximate integrals appearing in the approximations.

The primary result of this analysis is the derivation of bounds on rates of convergence of the approximations of the Koopman operator in terms of two contributions as described in Theorem (3). These two error terms are derived from

suitable forms of the Bernstein and Jackson inequalities, respectively. For a fixed uncertainty level $\delta$, as is typical in inverse problems, the term that includes $\rho_B(N)$ increases as $N \to \infty$, while the term that includes $\rho_J(N)$ decreases. Under some common assumptions regarding the discrete evolution law, strong rates of convergence are derived in the paper that are based on the recent development of Bernstein and Jackson inequalities for scattered bases in Sobolev spaces $W^r(\mathcal{M})$ over a compact, smooth Riemannian manifolds $\mathcal{M}$ of dimension $d$.

## REFERENCES


[1] Robert A. Adams and John J.F. Fournier. *Sobolev Spaces*. Academic Press, 2003.

[2] Giovanni S Alberti, Ernesto De Vito, Matti Lassas, Luca Ratti, and Matteo Santacesaria. Learning the optimal tikhonov regularizer for inverse problems. *arXiv preprint arXiv:2106.06513*, 2021.

[3] Alessandro Alla and J. Nathan Kutz. Nonlinear Model Order Reduction Via Dynamic Mode Decomposition. *Siam Journal On Scientific Computing*, 39(5):B778–B796, 2017.

[4] Thierry Aubin. *Nonlinear Analysis on Manifolds. Monge-Ampère Equations*. Springer-Verlag, 1982.

[5] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

[6] 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.

[7] Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern koopman theory for dynamical systems. *SIAM Review*, 64(2), 2022.

[8] Suddhasattwa Das and Dimitrios Giannakis. Koopman spectra in reproducing kernel hilbert spaces. *arXiv preprint, arXiv:1801.07799v8*, 2019.

[9] 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.

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

[11] Zlatko Drmac, Igor Mezic, and Ryan Mohr. Data Driven Modal Decompositions: Analysis And Enhancements. *Siam Journal On Scientific Computing*, 40(4):A2253–A2285, 2018.

[12] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator theoretic aspects of ergodic theory*, volume 272. Springer, 2015.

[13] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[14] 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.

[15] Patrick Gelss, Stefan Klus, Jens Eisert, and Christof Schuette. Multidimensional Approximation of Nonlinear Dynamical Systems. *Journal Of Computational And Nonlinear Dynamics*, 14(6), JUN 2019.

[16] Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied And Computational Harmonic Analysis*, 47(2):338–396, SEP 2019.

[17] 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.

[18] T. Hangelbroek, F.J. Narcowich, and J.D. Ward. Kernel approximation on manifolds i: Bounding the lebesgue constant. *SIAM J. Math. Anal.*, 42:1732–1760, 2010.

[19] T. Hangelbroek, F.J. Narcowich, and J.D. Ward. Polyharmonic and related kernels on manifolds: Interpolation and approximation. *Foundations of Computational Mathematics*, 12:625–670, 2012.

[20] 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.

[21] Thomas Hangelbroek, Francis J Narcowich, Christian Rieger, and Joseph D Ward. Direct and inverse results on bounded domains for meshless methods via localized bases on manifolds. *Contemporary Computational Mathematics-A Celebration of the 80th Birthday of Ian Sloan*, pages 517–543, 2018.

[22] 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.

[23] Andreas Kirsch et al. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer, 2011.

[24] Stefan Klus, Peter Koltai, and Christof Schutte. On the numerical approximation of the perron–frobenius and koopman operator. *arXiv*, 2016.

[25] 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.

[26] Milan Korda and Igor Mezic. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28:687–710, 2018.

[27] Vladimir Kostic, Pietro Novelli, Carlo Ciliberto Andreas Maurer, Lorenzo Rosasco, and Massimiliano Pontil. Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces. 2022.

[28] Vladimir Kostic, Pietro Novelli, Andreas Maurer, Carlo Ciliberto, Lorenzo Rosasco, and Massimiliano Pontil. Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces. *arXiv preprint arXiv:2205.14027*, 2022.

[29] Rainer Kress, V Maz'ya, and V Kozlov. *Linear integral equations*, volume 82. Springer, 1989.

[30] Andrew Kurdila and Parag Bobade. Koopman theory and linear approximation spaces. *Arxiv*, 2018.

[31] Andrzej Lasota and Michael C Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics*, volume 97. Springer Science & Business Media, 1998.

[32] Senka Macesic, Nelida Crnjari-Zic, and Igor Mezic. Koopman Operator Family Spectrum for Nonautonomous Systems. *Siam Journal On Applied Dynamical Systems*, 17(4):2478–2515, 2018.

[33] Alexandre Mauroy, Y Susuki, and I Mezić. *Koopman operator in systems and control*. Springer, 2020.

[34] Sai Tej Paruchuri, Jia Guo, Michael Kepler, Tim Ryan, Haoran Wang, Andrew J Kurdila, and Daniel Stilwell. Intrinsic and extrinsic approximation of koopman operators over manifolds. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1608–1613. IEEE, 2020.

[35] Vern I Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge University Press, 2016.

[36] Nathan Powell, Bowei Liu, and Andrew J Kurdila. Koopman methods for estimation of animal motions over unknown, regularly embedded submanifolds. *arXiv preprint arXiv:2203.05646*, 2022.

[37] Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11:905–934, 2010.

[38] Saburou Saitoh and Yoshihiro Sawano. *Theory of reproducing kernels and applications*. Springer, 2016.

[39] Ernesto De Vito, Nicole Mucke, and Lorenzo Rosasco. Reproducing kernel hilbert spaces on manifolds: Sobolev and diffusion spaces. *arXiv.org*, pages 1–31, 2019.

[40] Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[41] 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.

[42] 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, 2015.