Delay-coordinate maps and the spectra of Koopman operators

Suddhasattwa Das
Courant Institute of Mathematical Sciences, New York University, New York, New York, USA
E-mail: dass@cims.nyu.edu

Dimitrios Giannakis
Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York, USA
E-mail: dimitris@cims.nyu.edu

Abstract. The Koopman operator induced by a dynamical system is inherently linear and provides an alternate method of studying many properties of the system, including attractor reconstruction and forecasting. Koopman eigenfunctions represent the non-mixing component of the dynamics. They factor the dynamics, which can be chaotic, into quasiperiodic rotation on tori. Here, we describe a method in which these eigenfunctions can be obtained from a kernel integral operator, which also annihilates the continuous spectrum. We show that incorporating a large number of delay coordinates in constructing the kernel of that operator results, in the limit of infinitely many delays, in the creation of a map into the discrete spectrum subspace of the Koopman operator. This enables efficient approximation of Koopman eigenfunctions from high-dimensional data in systems with point or mixed spectra.

1. Introduction

The tasks of dimension reduction and forecasting of time series are very common in physical and engineering sciences, where the time-series studied are often partial observations of a nonlinear dynamical system. A classical example of such time series is data collected from the Earth’s climate system, where many of the active degrees of freedom are difficult to access via direct observations (e.g., subsurface ocean circulation). Moreover, the available observations typically mix together different physical processes operating on a wide range of spatial and temporal scales. For instance, in the climate system, the seasonal cycle and the El Niño Southern Oscillation (the latter evolving on interannual timescales) both have strong associated signals in sea surface temperature [1]. In this and many other applications, identifying dynamically important, coherent patterns of variability from the data can enhance our scientific understanding and predictive capabilities of complex phenomena.
Ergodic theory, and in particular its operator-theoretic formulation [2, 3], provides a natural framework to address these objectives. In this framework, the focus is on the action of the dynamical system on spaces of observables (functions of the state), as opposed to the dynamical flow itself. The advantage of this approach, first realized in the seminal work of Koopman [4], is that the action of a general dynamical system on spaces of observables is always linear. As a result, with appropriate regularity assumptions, the problem of identification and prediction of dynamically intrinsic coherent patterns can be formulated as an estimation problem for the spectrum of a linear evolution operator. In addition, for systems exhibiting ergodic behavior, spectral quantities such as eigenvalues and eigenfunctions can be statistically estimated from time-ordered data without prior knowledge of the state space geometry or the equations of motion. At the same time, spaces of observables are also infinite dimensional, so the issue of finite-dimensional approximation of (potentially unbounded) operators becomes relevant.

Starting from the techniques proposed in [5–7], the operator-theoretic approach to ergodic theory has stimulated the development of a broad range of techniques for data-driven modeling of dynamical systems. These methods employ either the Koopman [2, 6–18] or the Perron-Frobenius (transfer) operators [5, 19–22], which are duals to one another in appropriate function spaces. The goal common to these techniques is to approximate spectral quantities for the operator in question, such as eigenvalues, eigenfunctions, and spectral projections, from measured values of observables along orbits of the dynamics. To that end, a diverse range of approaches has been employed, including state space partitions [5, 19–22], harmonic averaging [6, 7, 23], iterative methods [8, 9], dictionary/basis representations [10–12, 15–17], delay-coordinate embeddings [12–15], and spectral-moment estimation [18].

Compared to observables identified by eigendecomposition techniques based on kernel integral operators that do not depend on the dynamics (e.g., covariance [24, 25] or heat operators [26–28], the latter of which have been popular in manifold learning applications), eigenfunctions of evolution operators are likely to offer higher physical interpretability and predictability, as they are determined from an operator intrinsic to the dynamical system. In particular, one of the key properties of Koopman or Perron-Frobenius eigenfunctions for ergodic dynamical systems is that they evolve periodically and with a single frequency (even if the underlying dynamical system is aperiodic), and thus have high predictability. This and a number of other attractive properties motivate the identification of such eigenfunctions of data.

Yet, for systems of sufficient complexity, Koopman and Perron-Frobenius operators have significantly more complicated spectral behavior than kernel integral operators, generally exhibiting a continuous spectral component and/or non-isolated eigenvalues, which presents challenges to the construction of data-driven approximation techniques with spectral convergence guarantees. Indeed, to our knowledge, spectral convergence results for the data-driven approximation of Koopman eigenvalues and eigenfunctions have been limited to special cases such as quasiperiodic rotations on tori [15], or systems observed through measurement functions lying in finite-dimensional invariant subspaces.
The main contribution of the work described below is the construction of a data-driven approximation scheme for Koopman eigenvalues and eigenfunctions that provably converges for a broad class of ergodic dynamical systems and observation maps, encompassing many of the applications encountered in the physical and engineering sciences. Our approach will be based on a combination of ideas from delay-coordinate maps of dynamical systems [29], kernel integral operators for machine learning [26–28, 30, 31], and Galerkin approximation techniques for variational eigenvalue problems [32]. Using these tools, we will construct a compact kernel integral operator that commutes with the Koopman operator in an asymptotic limit of infinitely many delays, and employ the finite-dimensional common eigenspaces of these operators as Galerkin approximation spaces for the Koopman eigenvalue problem. We will show that orthonormal bases of these spaces can be stably and efficiently approximated from finitely many measurements taken near the attractor, and the resulting data-driven Galerkin schemes converge in the asymptotic limit of large data.

2. Assumptions and statement of main results

A common underlying assumption for statistical analysis techniques methods is ergodicity, which is discussed below.

**Ergodicity.** Ergodic dynamics have the property that long-time averages of observables are equivalent to expectation values with respect to the invariant, ergodic measure $\mu$ [33]. This property justifies the working principle that the global properties (with respect to $\mu$) of an observable $F$ can be obtained from a time series for $F$, namely, $F(x_0), \ldots, F(x_{N-1})$, where $x_0, \ldots, x_{N-1}$ is an unobserved trajectory on the state space of the dynamical system. For our purposes, ergodicity implies that $L^2$ inner products between observables can be approximated by calculating time-correlations. Also, our methods rely on integral operators, and these can be approximated as matrices under the ergodic hypothesis. We now make our assumptions more precise.

**Assumption 1.** Let $M$ be an $m$-dimensional manifold equipped with its Borel $\sigma$-algebra. $\{\Phi^t, t \in \mathbb{R}\}$ is a flow on $M$ with a compact, invariant, attracting set $X$ with a Borel invariant ergodic measure $\mu$ with support equal to $X$. $F \in C^0(M; \mathbb{R}^d)$ is a continuous measurement function through which we collect a time-ordered data set consisting of $N$ samples $F(x_0), F(x_1), \ldots, F(x_{N-1})$, each $F(x_n)$ lying in $d$-dimensional data space. Here, $x_n = \Phi^{n\Delta t}(x_0)$, and $\Delta t$ is a fixed sampling interval such that the map $\Phi^{\Delta t}$ is ergodic for the invariant measure $\mu$.

**The Koopman operator.** Central to all our following discussions will be the concept of the Koopman operator. Koopman operators [2,3,34] act on observables by composition with the flow map, i.e., by time shifts. The space $L^2(X, \mu)$ of square-integrable, complex-valued functions on $X$ will be our space of observables. Given an observable $f \in L^2(X, \mu)$ and time $t \in \mathbb{R}$, $U^t : L^2(X, \mu) \to L^2(X, \mu)$ is the operator
Delay-coordinate maps and the spectra of Koopman operators defined as
\[(U^t f) : x \mapsto f(\Phi^t(x)), \quad \text{for } \mu\text{-a.e. } x \in X.\]

\(U^t\) is called the Koopman operator at time \(t\) associated with the flow. For measure-preserving systems, \(U^t\) is unitary, and has a well-defined spectral expansion consisting in general of both point and continuous parts lying in the unit circle [7]. The problems of mode decomposition and non-parametric prediction can both be stated in terms of the Koopman operator [15]. We will now describe an important tool for studying Koopman operators, namely their eigenfunctions.

**Koopman eigenfunctions.** Every eigenfunction \(z\) of \(U^t\) satisfies the following equation for some \(\omega \in \mathbb{R}\):
\[U^t z = \exp(i\omega t)z.\]  

Koopman eigenfunctions are particularly useful for prediction and dimension reduction in dynamical systems. This is because, as seen in (1), the knowledge of an eigenfunction \(z\) at time \(t = 0\) enables accurate predictions of \(z\) up to any time \(t\), since \(U^t\) operates on \(z\) as a multiplication operator by a time-periodic, single-frequency multiplication factor. Moreover, it is possible to construct a dimension reduction map, sending the high-dimensional data \(F(x) \in \mathbb{R}^d\) to the vector \((z_1(x), \ldots, z_l(x)) \in \mathbb{C}^l\), where \(l \ll d\), and the \(z_1, \ldots, z_l\) are Koopman eigenfunctions corresponding to rationally independent frequencies \(\omega_1, \ldots, \omega_l\) [7, 12, 15]. In this representation, the \(z_k\) can be thought of as “coordinates” corresponding to distinct physical processes operating at the timescales \(2\pi/\omega_k\). Also of interest (and in some cases easier to compute) are the projections of the observation map \(F\) onto the Koopman eigenfunctions, called Koopman modes [7]. Data-driven techniques for computing Koopman eigenvalues, eigenfunctions, and modes that have been explored in the past include methods based on generalized Laplace analysis [6, 7], dynamic mode decomposition (DMD) [8–10, 35], extended DMD (EDMD) [11, 17], Hankel matrix analysis [10, 13, 14], and data-driven Galerkin methods [12, 15, 16]. The latter approach, as well as the related work in [36], additionally address the problem of nonparametric prediction of observables and probability densities.

Let \(D\) be the closed subspace of \(L^2(X, \mu)\) spanned by the eigenfunctions of \(U^t\), and \(D^\perp\) its orthogonal complement. Systems in which \(D\) contains non-constant functions and \(D^\perp\) is non-empty are called mixed-spectrum systems. The \(L^2\) space of a general system admits the \(U^t\)-invariant decomposition [7]
\[L^2(X, \mu) = D \oplus D^\perp.\]  

**Kernel integral operators.** The method that we will describe in this paper relies heavily on kernel integral operators. A kernel is a function \(k : M \times M \to \mathbb{R}\), measuring the similarity between pairs of points on \(M\). Kernel functions can be of various designs, and are meant to capture the nonlinear geometric structures of data; see for example [26, 27, 37]. One advantage of using kernels is that they can be defined so as to operate directly on the data space, e.g., \(k(x, y) = \kappa(F(x), F(y))\) for some function \(\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}\) of appropriate regularity. Defined in this manner, \(k\) can be evaluated
using measured quantities $F(x)$ without explicit knowledge of the underlying state $x$. Associated with a square-integrable kernel $k \in L^2(X \times X, \mu \times \mu)$ is a compact integral operator $K : L^2(X, \mu) \mapsto L^2(X, \mu)$ such that

$$Kf(x) := \int_X k(x, y)f(y)\,d\mu(y).$$

In some cases, we will make the following assumptions on kernels.

**Assumption 2.** The kernel $k : M \times M \mapsto \mathbb{R}$ is positive and continuous. Moreover, for every compact set $S \subseteq M$, there exists a constant $l > 0$ such that $k|_{S \times S} \geq l$.

**Overview of approach.** We will address the eigenvalue problem for $U^t$ by solving an eigenvalue problem for a kernel integral operator $P_Q$, which is accessible from data, and in a suitable limit, commutes with $U^t$. Since commuting operators have common eigenspaces, this will allow us to compute eigenfunctions of $U^t$ through expansions in eigenbases obtained from $P_Q$. In what follows, we will construct $P_Q$ by first defining $K$ from (3) using a kernel function operating on delay-coordinate mapped data [29] with $Q$ delays, and then performing a Markov normalization of that operator via the procedure introduced in the diffusion maps algorithm [27].

In our following main result, we show that in the limit of infinitely many delays, $Q \to \infty$, the eigenspaces of $P_Q$ corresponding to nonzero eigenvalues are also eigenspaces of $U^t$.

**Theorem 1.** Under Assumption 1, there exists a real, self-adjoint, ergodic, compact Markov operators $P : L^2(X, \mu) \mapsto L^2(X, \mu)$, which commutes with $U^t$, and is a limit of operators $P_1, P_2, \ldots$ (also real, self-adjoint, ergodic, compact, and Markov) in the $L^2(X, \mu)$ operator-norm topology. The operators $P_Q$ have Markov kernels $p_Q : M \times M \mapsto \mathbb{R}$ satisfying the conditions in Assumption 2, and determined from delay-coordinate mapped observations $F(x), F(\Phi^{\Delta t}(x)), \ldots, F(\Phi^{(Q-1)\Delta t}(x))$ with $Q$ delays. Moreover, the kernel $p : M \times M \mapsto \mathbb{R}$ of $P$ lies in $L^\infty(X \times X, \mu \times \mu)$, and $p_Q$ converges to $p$ in $L^p(X \times X, \mu \times \mu)$ norm with $1 \leq p < \infty$.

Theorem 2 below is a continuation of Theorem 1, and can be used to conclude some useful properties of the operator $P$.

**Theorem 2.** Let Assumption 1 hold, and $T$ be a kernel integral operator with a real-valued, symmetric kernel $\tau \in L^2(X \times X, \mu \times \mu)$ such that $T$ commutes with $U^t$ (e.g., $T = P$ from Theorem 1). Then,

(i) $\tau$ lies in the tensor product subspace $D \otimes D$, and is invariant under the flow $\Phi^t \times \Phi^t$.

(ii) $D$ and $D^\perp$ are invariant under $T$. Moreover, $\text{ran} \, T$ is a subspace of $D$, $D^\perp$ is a subspace of $\text{ker} \, T$, and both $\text{ran} \, T$ and $\text{ker} \, T$ are invariant under $U^t$.

Moreover, if $\text{ran} \, T$ contains non-constant functions:

(iii) There exists a measurable map $\pi : X \to \mathbb{T}^D$ for some $D \in \mathbb{N}$, whose components consist of joint eigenfunctions of $T$ and $U^t$, such that $\pi$ factors $\Phi^t$ into a rotation.
on the torus by a vector $\bar{\omega} \in \mathbb{R}^D$, i.e., $\pi(\Phi^t(x)) = \pi(x) + \bar{\omega}t \mod 1$ for $\mu$-a.e. $x \in X$.

(iv) There exists a choice of dimension $D$ from (iii) and a symmetric kernel $\hat{\tau} \in L^2(T^D \times T^D, \text{Leb})$ on the $D$-torus, such that $\tau(x, y) = \hat{\tau}(\pi(x), \pi(y))$ for $\mu \times \mu$-a.e. $(x, y) \in X \times X$.

Note that Theorems 1 and 2 hold for operators acting on $L^2$ spaces only. To be able to say more about the behavior of these operators on spaces of continuous functions, an additional assumption on the Koopman eigenfunctions and the observation map will be needed.

**Assumption 3.** All Koopman eigenfunctions $z_1, z_2, \ldots$ are continuous. Moreover, the discrete component $F_D$ of $F$ from (2) is expressible as $F_D = \sum_{k \in \mathbb{N}} a_k z_k$, for some $\{a_k\}_{k \in \mathbb{N}} \in \ell^1$.

We then have:

**Theorem 3.** Let Assumptions 1 and 3 hold. Then, the kernel $p$ of the operator $P$ from Theorem 1 is uniformly continuous on a full-measure, dense subset of $X \times X$. As a result:

(i) $P$ maps $L^2(X, \mu)$ into the space of $\mu$-a.e. continuous functions on $X$.

(ii) $P$ compactly maps $C^0(X)$ into itself.

(iii) The norms of the operators $P$ in (i) and (ii) are bounded above by $\|p\|_{L^\infty(X \times X)}$.

(iv) For every $f \in C^0(X)$, $P_Q f$ is a sequence of continuous functions converging $\mu$-a.e. to $Pf$.

**Corollary 4** (spectral convergence). Under the assumptions of Theorem 1, the following hold:

(i) For every nonzero eigenvalue $\lambda$ of $P$ with multiplicity $\alpha$ and every neighborhood $S \subset \mathbb{R}$ of $\lambda$ such that $\text{spec}(P) \cap S = \{\lambda\}$, there exists $Q_0 \in \mathbb{N}_0$ such that for all $Q > Q_0$, $\text{spec}(P_Q) \cap S$ contains $\alpha$ elements converging as $Q \to \infty$ to $\lambda$.

(ii) Let $\Pi$ be any projector to the eigenspace $W_\lambda$ of $P$ at eigenvalue $\lambda$. Let also $\Pi_Q$ be any projector to the union of the eigenspaces of $P_Q$ corresponding to the eigenvalues in $\text{spec}(P_Q) \cap S$. Then, as $Q \to \infty$, $\Pi_Q$ converges strongly to $\Pi$. Moreover, the gap between $W_\lambda$ and $\text{ran} \Pi_Q$, defined in (19) below, converges to zero.

Theorems 1–3 and Corollary 4 are proved in Section 5. A result analogous to Theorem 1, but restricted to smooth manifolds, smooth observation maps, and Koopman operators with pure point spectrum and smooth eigenfunctions, was presented in [15]. Theorem 1 generalizes this result to non-smooth state spaces and Koopman operators with mixed spectra. With this result, the eigenvalues and eigenfunctions of $P_Q$ consistently approximate those of $P$, and the latter can be used in turn to approximate the eigenvalues and eigenfunctions of $U^t$ or its generator (defined in Section 3 ahead) through stable Galerkin techniques [12, 15]. One such technique will be presented in Section 6 utilizing diffusion regularization of the generator.
Physical measures. A point $x \in M$ is said to be in the basin of the measure $\mu$ with respect to the discrete-time map $\Phi^\Delta t$ if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\Phi^n \Delta t(x)) = \int_X f(y) \, d\mu(y), \quad \forall f \in C^0(M). \quad (4)$$

The basin $B_\mu$ of an invariant ergodic measure $\mu$ always includes $\mu$-a.e. point in the support of $\mu$ (in this case $X$), and is a forward-invariant set. An important property that we need the invariant measure $\mu$ to have is that it is physical [38]. Moreover, we will require that the dynamics has a suitable absorbing ball property. These assumptions can be summarized as follows:

**Assumption 4.** The set $B_\mu$ of points satisfying (4) has positive Lebesgue measure, i.e., the measure $\mu$ is physical. Moreover, there exists a subset $\mathcal{V} \subseteq B_\mu$, also of positive Lebesgue measure, such that for every $x_0 \in \mathcal{V}$ there exists a compact set $\mathcal{U}$ (which may depend on $x_0$, and necessarily includes $X$), such that the orbit $x_n = \Phi^n \Delta t(x_0)$ enters $\mathcal{U}$ and never leaves it.

Assumption 4 guarantees that even if $X$ is a zero Lebesgue-measure set in $M$, so long as the initial point $x_0$ lies in $\mathcal{V}$ (and not necessarily in $X$), these techniques converge. Examples where Assumption 4 is satisfied include:

(i) Ergodic flows on compact manifolds, in which case $\mathcal{U} = \mathcal{V} = \overline{B_\mu} = M = X$.
(ii) Certain classes of contractive flows on potentially noncompact manifolds (e.g., the Lorenz 63 (L63) system on $M = \mathbb{R}^3$ [39] studied in Section 8 ahead).
(iii) Certain classes of dissipative partial differential equations possessing inertial manifolds and physical measures [40, 41].

Assumption 4 is not required for our main theoretical results (Theorems 1–3 below) to hold, but will be required later for the convergence of our data-driven approximation techniques. It guarantees that even if $X$ is a zero Lebesgue-measure set in $M$, so long as the initial point $x_0$ lies in $\mathcal{V}$ (and not necessarily in $X$), these techniques converge. Examples where Assumption 4 is satisfied include:

(i) Ergodic flows on compact manifolds, in which case $\mathcal{U} = \mathcal{V} = \overline{B_\mu} = M = X$.
(ii) Certain classes of contractive flows on potentially noncompact manifolds (e.g., the Lorenz 63 (L63) system on $M = \mathbb{R}^3$ [39] studied in Section 8 ahead).
(iii) Certain classes of dissipative partial differential equations possessing inertial manifolds and physical measures [40, 41].

The following result shows that under Assumptions 1–4, the nonzero eigenvalues of $P_Q$ and the corresponding (continuous) eigenfunctions can be approximated to any degree of accuracy by a data-driven operator $P_{Q,N}$, acting on the finite-dimensional Hilbert space $L^2(\mathcal{U}, \mu_N)$ associated with the sampling probability measure $\mu_N = \sum_{n=0}^{N-1} \delta_{x_n} / N$, and constructed from time-ordered measurements $F(x_0), \ldots, F(x_{N-1})$ of the observable $F$. 
Theorem 5. Under Assumptions 1–4 and for any starting state $x_0 \in \mathcal{V}$, the following hold:

(i) Every eigenfunction of $P_{Q,N} : L^2(\mathcal{U},\mu_N) \mapsto L^2(\mathcal{U},\mu_N)$ at nonzero eigenvalue extends to a continuous eigenfunction of an analogous integral operator $P''_{Q,N} : C^0(\mathcal{U}) \mapsto C^0(\mathcal{U})$ to $P_{Q,N}$ acting on continuous functions, corresponding to the same eigenvalue.

(ii) Every eigenfunction of $P_Q : L^2(X,\mu) \mapsto L^2(X,\mu)$ at nonzero eigenvalue extends to a continuous eigenfunction of an analogous integral operator $P''_Q : C^0(\mathcal{U}) \mapsto C^0(\mathcal{U})$ to $P_Q$ acting on continuous functions, corresponding to the same eigenvalue.

(iii) As $N \to \infty$, $P''_{Q,N}$ converges in spectrum to $P''_Q$ in the sense of Corollary 4.

Theorem 5 is proved in Section 7. Using the eigenvalues and eigenfunctions of $P_{Q,N}$, it is possible to construct data-driven Galerkin schemes for the Koopman eigenvalue problem, which are structurally identical to its counterparts formulated in terms of the eigenvalues and eigenfunctions of $P$, and converge in a suitable joint limit of large data ($N \to \infty$), infinitely many delays ($Q \to \infty$), and infinite Galerkin approximation space dimension. Figure 1 shows numerical results obtained via this approach from data generated by two dynamical systems mixed spectra, described in (38) and (39), respectively. In both examples, we start with a $C^\infty$ vector field $\vec{V}$ on a manifold $M$. In the first example, $M = X = T^4$, so $\mathcal{U} = X = M$; in the second example, $M = \mathbb{R}^3 \times S^1$ and $X = X_{Lor} \times S^1 \subset M$, where $X_{Lor}$ is the Lorenz 63 attractor embedded in $\mathbb{R}^3$.

Remark 6. One class of dynamical systems that this work does not address is those with purely continuous spectrum (i.e., $\mathcal{D} = \text{span}\{1_X\}$, where $1_X$ is the function equal to 1 everywhere on $X$). In fact, Theorem 2 shows that for such systems, in the limit of infinitely many delays, every non-constant observable will lie in the nullspace of the integral operator $P$. See Section 8 for further discussion. Moreover, in certain cases of the observation map $F$, the kernel of $P$ can become the whole space other than the constant functions, even if $\mathcal{D}$ contains non-constant functions. Corollary 38 in Section 8 establishes necessary and sufficient “observability” conditions under which, in certain product dynamical systems, $P$ is not a trivial operator.

Outline of the paper. In Section 3, we review some important concepts from the spectral theory of dynamical systems. In Section 4, we construct the integral operator $P_Q$, which is the key tool of our methods and is also the operator described in Theorems 1–3. Next, we prove these theorems and Corollary 4 in Section 5. In Section 6, we present a Galerkin method for the eigenvalue problem for the Koopman generator, with a small amount of diffusion added for regularization, formulated in the eigenbasis of $P$. In Section 7, we introduce the data-driven realization of $P_Q$, and establish the spectral convergence properties stated in Theorem 5, along with the convergence properties of the associated data-driven Galerkin scheme for the generator. In Section 8, the methods are applied to two mixed-spectrum flows, followed by a discussion of the results.
Delay-coordinate maps and the spectra of Koopman operators

Figure 1: Eigenfunctions of $P_Q$ and the associated matrix representation of the generator $V$ for the torus-based flow $\Phi^t_{T_3 \times R_\omega}$ in (38) (top panels) and the L63-based flow $\Phi^t_{Lor \times R_\omega}$ in (39) (bottom panels). The eigenfunctions $\phi_i$ of $P_Q$ from (15) have been computed from single orbits of these two systems using a large number of delays, $Q = 2000$, and plotted as time series along these orbits in the middle and right-hand panels. These time series are periodic with frequencies equal to integer multiples of the rotation frequency $\omega$, each frequency having multiplicity 2. This behavior is in agreement with the result from Theorem 1 that $U^t$ and $P_Q$ commute in the limit $Q \to \infty$. The left-hand panels show the elements $V_{ij} = \langle \phi_i, V \phi_j \rangle$ of the matrix representation of the generator in the $\{\phi_i\}$ basis. Note that $V_{ji} = -V_{ij}$ since $V$ is a skew-symmetric operator. Moreover, because the first eigenfunction $\phi_0$ of $P_Q$ is the constant function and $V \phi_0 = 0$, the first column and row only have zero entries. The $2 \times 2$ block-diagonal form of the matrices indicates that each of the eigenfunction pairs $(\phi_1, \phi_2), (\phi_3, \phi_4), \ldots$ spans an eigenspace of $V$.

3. Overview of spectral methods for dynamical systems

In this section, we review some concepts from the spectral theory of dynamical systems and establish some facts about Koopman eigenfunctions. Henceforth, we use the notations $\langle f, g \rangle = \int_X f^* g \, d\mu$ and $\|f\| = \langle f, f \rangle^{1/2}$ to represent the inner product and norm of $L^2(X, \mu)$, respectively.

Generator of a flow. By continuity of the flow $\Phi^t$, the family of operators $U^t$ is a
strongly continuous, 1-parameter group of unitary transformations of the Hilbert space $L^2(X, \mu)$. By Stone’s theorem [42], any such family has a generator $V$, which is a skew-adjoint operator with a dense domain $D(V) \subset L^2(X, \mu)$, defined as

$$Vf := \lim_{t \to 0} \frac{1}{t} (U^t f - f), \quad f \in D(V).$$

(5)

The operators $U^t$ and $V$ share the same eigenfunctions; in particular, $z \in D(V)$ with $U^t z = e^{i\omega t} z$ satisfies

$$Vz = i\omega z.$$

In light of (5) and the above relation, we can interpret the quantity $\omega \in \mathbb{R}$ as a frequency intrinsic to the dynamical system (which we sometimes refer to as an “eigenfrequency”).

Vector fields as generators. If we start with a vector field $\vec{V}$ on $M$, then under appropriate regularity conditions (for example, $\vec{V}$ is locally Lipschitz continuous and satisfies suitable growth bounds at infinity), this vector field induces a $C^1$ flow $\Phi^t : M \mapsto M$ defined for all $t \in \mathbb{R}$. Suppose that there is a compact invariant set $X \subseteq M$ with an ergodic invariant measure $\mu$. This set $X$ is not necessarily a submanifold, and may not even have any differentiability properties. Nevertheless, $(X, \Phi^t, \mu)$ is an ergodic dynamical system with an associated strongly-continuous, unitary group of Koopman operators $U^t$. Acting on $C^1(M)$ functions restricted to $X$, the generator $V$ of this group coincides with the vector field $\vec{V}$, the latter viewed as an operator on $C^1(M)$. For example, in quasiperiodic systems, $X = M = \mathbb{T}^m$, $\vec{V}$ generates a rotation, and $\mu$ is equivalent to the Lebesgue volume measure. On the other hand, for the Lorenz attractor (see (37)), $M = \mathbb{R}^3$, $\vec{V}$ is smooth and dissipative, $X$ is a compact subset with non-integer fractal dimension [43], and $\mu$ is supported on $X$.

Eigenfunctions as factor maps. We state the following properties of a Koopman eigenfunction $z$ of an ergodic dynamical system.

(i) If $z$ corresponds to a nonzero eigenfrequency $\omega$, then it has zero mean with respect to the invariant measure $\mu$. This can be concisely expressed as $\langle 1, z \rangle = 0$.

(ii) The flow $\Phi^t$ is semi-conjugate to the irrational rotation by $\omega t$ on the unit circle, with $z$ acting as a semiconjugacy map. This follows directly from (1). Since the eigenfunctions are $L^2$ equivalence classes, the semiconjugacy is measure-theoretic (holds $\mu$-a.e.), but would be $C^r$ if the eigenfunctions have a $C^r$ representation.

(iii) Normalized eigenfunctions with $\|z\| = 1$ have $|z(x)| = 1$ for $\mu$-a.e. $x \in X$, by (1). As a result, the map $z$ can now be viewed as a projection onto a circle in a measure-theoretic sense, $z(x) \in S^1$ for $\mu$-a.e. $x \in X$.

Remark 7. In [44], Anosov and Katok construct discrete-time ergodic maps which are measure-theoretically isomorphic to systems with only discrete spectrum but with no continuous Koopman eigenfunctions. However, to our knowledge, no continuous-time $C^0$ ergodic flow with discontinuous Koopman eigenfunctions has been established.

Eigenfunctions form a group. Another important property of Koopman eigenfunctions for ergodic dynamical systems is that they form a group under
multiplication. That is, the product of two eigenfunctions of \( U^t \) is again an eigenfunction, because of the following relation:

\[
U^t z_i = \exp(it\omega_i)z_i, \quad i \in \{1, 2\},
\]

\[
\implies U^t(z_1 z_2) = (U^t z_1)(U^t z_2) = \exp(it(\omega_1 + \omega_2))z_1 z_2.
\]

Moreover, an analogous relation holds for the eigenfunctions and eigenvalues of \( V \). The following lemma is about products of an element of \( \mathcal{D} \) with an element of \( \mathcal{D}^\perp \). The proof is left to the reader.

**Lemma 8.** Let \( \Phi^t \) be an ergodic flow on a probability space \( (X, \mu) \) such that \( U^t \) has a mixed spectrum. Then for every \( f \in \mathcal{D}^\perp \) and \( g \in \mathcal{D} \) for which \( fg \in L^2(X, \mu) \), \( fg \) lies in \( \mathcal{D}^\perp \).

The eigenvalues of \( V \) are closed under integer linear combinations and are generated by a finite set of rationally independent eigenvalues \( i\omega_1, \ldots, i\omega_m \). That is, every eigenvalue of \( V \) is simple, and has the form \( \omega_{\vec{a}} = \sum_{j=1}^{m} a_j \omega_j \) for some \( (a_1, \ldots, a_m) \in \mathbb{Z}^m \). Moreover, the corresponding eigenfunction is given by

\[
z_{\vec{a}} = \prod_{j=1}^{m} z_{a_1}^{a_1} \cdots z_{a_m}^{a_m},
\]

where \( z_j \) is the eigenfunction at eigenvalue \( i\omega_j \). The following is a generalization of Property 2 of Koopman eigenfunctions listed above.

**Proposition 9.** Given an arbitrary collection \( \{z_{\vec{a}_1}, z_{\vec{a}_2}, \ldots, z_{\vec{a}_l}\} \) of \( l \) Koopman eigenfunctions, there exists a map \( \pi : X \mapsto \mathbb{C}^l \) with

\[
\pi(x) = (z_{\vec{a}_1}(x), \ldots, z_{\vec{a}_l}(x)), \quad \text{for } \mu\text{-a.e. } x \in X,
\]

such that:

(i) The image \( \pi(X) \) is a torus of dimension \( D \leq \min\{m, l\} \), with \( D = l \) if \( \omega_{\vec{a}_1}, \ldots, \omega_{\vec{a}_l} \) are rationally independent.

(ii) The flow \( (\Phi^t, \mu) \) on \( X \) is semi-conjugate to an ergodic rotation \( (\Omega^t, \text{Leb}) \) on \( \mathbb{T}^D \) (i.e., \( \pi \circ \Phi^t = \Omega^t \circ \pi, \mu\text{-a.e.} \)) associated with a frequency vector whose components are a subset of \( \{\omega_{\vec{a}_1}, \ldots, \omega_{\vec{a}_l}\} \).

(iii) Every Koopman eigenfunction \( z \) whose corresponding eigenfrequency is a linear combination of the \( \omega_{\vec{a}_1}, \ldots, \omega_{\vec{a}_l} \) satisfies \( z(x) = \zeta(\pi(x)) \) for \( \mu\text{-a.e. } x \in X \), where \( \zeta \in C^\infty(\mathbb{T}^D) \) is a smooth Koopman eigenfunction of the ergodic rotation on the \( D \)-torus corresponding to the same eigenfrequency.

**Remark 10.** If \( m > 1 \), the set of eigenvalues \( \{i\omega_{\vec{a}}\}_{\vec{a} \in \mathbb{Z}^m} \) is dense on the imaginary axis. This property adversely affects the stability of numerical approximations of Koopman eigenvalues and eigenfunctions even in systems with pure point spectrum, necessitating the use of regularization [15]. We will return to this point in Section 6.
Lemma 11 ([45], Section 2.3). Let $\Delta t > 0$ be as in Assumption 1. Then, the orthogonal projection $\pi_\omega f$ of $f$ onto the eigenspace of $U^{\Delta t}$ corresponding to the eigenvalue $e^{i\omega \Delta t}$ of $U^{\Delta t}$ is given by

$$\pi_\omega f = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} e^{-i\omega n \Delta t} U^n f.$$ 

Moreover, $\pi_\omega \equiv 0$ if $\omega$ is not in the point spectrum. Otherwise, $U^{\Delta t}(\pi_\omega f) = e^{i\omega \Delta t} \pi_\omega f$.

Mixing and weak mixing. An observable $f \in L^2(X, \mu)$ is said to be mixing if for all $g \in L^2(X, \mu)$, $\lim_{t \to \infty} \langle g, U^t f \rangle = 0$; it is said to be weak mixing if $\lim_{N \to \infty} N^{-1} \sum_{n=0}^{N-1} |\langle g, U^n f \rangle| = 0$. The flow $\Phi^t$ is said to be (weak) mixing if $f$ is (weak) mixing for all $f \in L^2(X, \mu)$. It is known that every $f \in D^\perp$ is weak mixing (see, e.g., Mixing Theorem, p. 45 in [46]), whereas no observable in $D$ is weak mixing. Thus, the component $D$, often called the quasiperiodic subspace, shows no decay of correlation, unlike its complement $D^\perp$, which represents the chaotic component of the dynamics. In addition, weak mixing observables in $D^\perp$ and observables in $D$ have a useful pointwise decorrelation property:

Lemma 12. Let $f \in D^\perp$ and $g \in D$. Then, for $\mu$-a.e. $x, y \in X$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} g^*(\Phi^n \Delta t(x)) f(\Phi^n \Delta t(y)) = 0.$$ 

Proof. Without loss of generality, we may assume that $g$ is an eigenfunction of $U^{\Delta t}$ with eigenvalue $e^{i\omega \Delta t}$. Then,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} g^*(\Phi^n \Delta t(x)) f(\Phi^n \Delta t(y))$$

$$= g^*(x) \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} e^{-i\omega n \Delta t} f(\Phi^n \Delta t(y)),$$

which is equal to $g^*(x) \pi_\omega f(y)$ by Lemma 11. The latter is equal to zero since $f \in D^\perp$.

4. Kernel integral operators from delay-coordinate mapped data

4.1. Choice of kernel

Consider a kernel integral operator of the class (3) associated with an $L^2$ kernel $k : M \times M \to \mathbb{R}$. Then, the following properties hold (e.g., [47, 48]):

(i) $K$ is a Hilbert-Schmidt, and therefore compact, operator on $L^2(X, \mu)$, with operator norm bounded by $\|k\|_{L^2(X \times X)}$.

(ii) If $k$ is symmetric, then $K$ is self-adjoint.

(iii) If $k$ is $C^0$, then $Kf$ is also $C^0$ for every $f \in L^2(X, \mu)$.
(iv) If $X$ is a $C^r$ manifold and $k$ is $C^r$, then $Kf$ is also $C^r$ for every $f \in L^2(X, \mu)$.

A common approach in machine learning and harmonic analysis [26–28] is to work with kernels of the form

$$k(x, y) = h(d(x, y))$$

where $h$ is a continuous shape function on $\mathbb{R}$, and $d : M \times M \mapsto \mathbb{R}_0$ is a metric or pseudo-metric on $M$, usually defined as a pullback of a distance-like function in data space. As a concrete example, in what follows we employ a Gaussian shape function $h_\epsilon(y) = e^{-y^2/\epsilon}$ parameterized by a bandwidth parameter $\epsilon > 0$; such functions are popular in manifold learning and other related geometrical data analysis techniques due to their localizing behavior as $\epsilon \to 0$. We also consider the family of distance-like functions

$$d_Q^2(x, y) = \frac{1}{Q} \sum_{n=0}^{Q-1} \| F(\Phi^n \Delta t(x)) - F(\Phi^n \Delta t(y)) \|^2,$$  \hspace{1cm} (7)

where $F$ is the observation map, and $\| \cdot \|$ the canonical 2-norm on $\mathbb{R}^d$. Note in particular that $d_Q$ corresponds to a pseudo-distance on $M$ induced from delay-coordinate mapped data with $Q$ delays. That is, $d_Q$ is symmetric, non-negative, and satisfies the triangle inequality, but depending on the properties of $F$ and the number of delays it may vanish on distinct points. Using the Gaussian shape function $h_\epsilon$ and the distance-like functions in (7), we arrive at the family of kernels

$$k_Q(x, y) = e^{-d_Q^2(x,y)/\epsilon},$$  \hspace{1cm} (8)

and the associated integral operators $K_Q$ from (3). Note that $K_Q$ satisfies all four properties of kernels listed above, and in addition satisfies Assumption 2.

Before proceeding, we state a property of $K_Q$ valid for the kernels $K_Q$ in (8), and more generally the class of kernels satisfying Assumption 2. These properties will be important in the construction of the Markov-normalized operators in Theorems 1 and 2. In what follows $1_S$ will denote the constant function equal to 1 on a set $S$.

**Lemma 13.** For any $Q \in \mathbb{N}$, the functions $\rho_Q = K_Q 1_X$, and $\sigma_Q = K_Q (1/\rho_Q)$ are continuous and positive. Moreover, restricted on $X$, they are bounded away from zero.

**Proof.** The claims follow directly from the facts that $X$ is compact and $k_Q |_{X \times X}$ is bounded away from zero. \hfill \Box

**Remark 14.** Intuitively, $\rho_Q$ can be thought of as a “sampling density” on $X$. For instance, if $X$ were a manifold embedded in $\mathbb{R}^Q$ by a delay-coordinate map constructed from $F$, then up to an $\epsilon$-dependent scaling, $\rho_Q$ would approximate the density of the invariant measure $\mu$ relative to the volume measure associated with that embedding.

**Remark 15.** In a number of applications such as statistical learning on manifolds [26–28, 30], one-parameter families of integral operators such as $K_{\epsilon,Q}$ and $P_{\epsilon,Q}$ are studied in the limit $\epsilon \to 0$, where under certain conditions they can be used to approximate
generators of Markov semigroups; one of the primary examples being the Laplace-Beltrami operator on Riemannian manifolds. Here, the fact that the state space $X$ may not (and in general, will not) be smooth precludes us from taking such limits unconditionally. However, according to Theorem 2(ii), passing first to the limit $Q \to \infty$ allows us to view $K_\epsilon$ and $P_\epsilon$ as operators on functions on a smooth manifold, namely a $D$-dimensional torus, and study the small-$\epsilon$ behavior of these operators in that setting.

4.2. Asymptotic behavior in the infinite-delay limit

To study the behavior of $K_Q$ in the limit of infinitely many delays, $Q \to \infty$, we first consider the properties of the pseudometric $d_Q$ in the same limit. The latter can be studied in turn through a useful (nonlinear) map $\Psi : C^0(X) \mapsto L^\infty(X \times X, \mu \times \mu)$, which maps a given observation function $F$ into a (pseudo)metric on $X$, namely,

$$\Psi(F)(x,y) := \frac{1}{Q} \sum_{q=0}^{Q-1} \| F(\Phi^q(x)) - F(\Phi^q(y)) \|^2. \quad (9)$$

In what follows, $d_X : X \times X \to \mathbb{R}$ will denote the metric $X$ inherits from $M$.

**Theorem 16.** Let Assumption 1 hold, and $F = F_D + F_{D\perp}$ be the $L^2$ decomposition of $F$ from (2). Then, $\Psi(F)$ in (9) is well-defined as a function in $L^\infty(X \times X, \mu \times \mu)$, and $\Psi_Q(F)$ converges to $\Psi(F)$ in $L^p(X \times X, \mu \times \mu)$ norm for $1 \leq p < \infty$. Moreover:

(i) For every $t \in \mathbb{R}$ and $\mu$-a.e. $x, y \in X$, $\Psi(F)(\Phi^t(x), \Phi^t(y)) = \Psi(F)(x,y)$.

(ii) For $\mu$-a.e. $x, y \in X$, $\Psi(F)(x,y) = \Psi(F_{D\perp})(x,y) + \Psi(F_D)(x,y)$.

(iii) $\Psi(F_{D\perp})$ is a constant almost everywhere and equals $2\|F_{D\perp}\|_{L^2}^2$. Therefore,

$$\Psi(F) = \Psi(F_D) + 2\|F_{D\perp}\|_{L^2}^2. \quad (10)$$

In particular, $\Psi(F) \in D \times D$.

If, moreover, Assumption 3 holds:

(iv) $\Psi(F_D) \in C^0(X \times X)$ and $\Psi_Q(F_D)$ converges to $\Psi(F_D)$ uniformly on $X \times X$.

(v) $\Psi(F)$ is uniformly continuous on a full-measure, dense subset of $X \times X$.

(vi) $\Psi(F)$ has a unique continuous extension $\bar{\Psi}(F) \in C^0(X \times X)$, and $\Psi_Q(F)$ converges to $\bar{\Psi}(F)$ $\mu$-almost uniformly.

**Proof.** To prove well-definition of $\Psi$, note that $\Psi(F)$ exists $\mu$-a.e. since it is the pointwise limit of the Birkhoff averages $\Psi_Q(F)$ of the continuous function $d_1 : (x,y) \mapsto \| F(x) - F(y) \|$ with respect to the product flow $\Phi^t \times \Phi^t$ on $X \times X$. By compactness of $X \times X$, each of the functions $\Psi_Q(F)$ is bounded above by $\|d_1\|_{C^0(X \times X)}$. Therefore, $\Psi(F)$ lies in $L^\infty(X \times X, \mu \times \mu)$, and thus in $L^p(X \times X, \mu \times \mu)$, $1 \leq p < \infty$, since $\mu \times \mu$ is a probability measure. The $\Psi_Q(F) \to \Psi(F)$ convergence in $L^p(X \times X, \mu \times \mu)$, $1 \leq p < \infty$, then follows from the $L^p$ Von Neumann ergodic theorem.
By the invariance of the infinite Birkhoff averages, $\Psi(F)$ is invariant under the flow $\Phi^t \times \Phi^\Delta t$. Thus $\Psi(F)$ must lie in the kernel of $V \otimes V$ and thus is invariant under the flow $\Phi_t \times \Phi^t$ for all $t \in \mathbb{R}$, proving Claim (i).

To prove Claim (ii), let $x_q$ and $y_q$ denote $\Phi^q \Delta t(x)$ and $\Phi^q \Delta t(y)$ respectively. Let $G_D : X \times X \to \mathbb{R}^d := (x, y) \mapsto F_D(x_q) - F_D(y_q)$, and similarly define $G_{\perp D} : X \times X \to \mathbb{R}^d$. Expanding the right-hand side of (9) gives,

$$\Psi(F)(x, y) = \lim_{Q \to \infty} \frac{1}{Q} \sum_{q=0}^{Q-1} \left( \|G_D(x_q, y_q)\|^2 + \|G_{\perp D}(x_q, y_q)\|^2 \right)$$

$$- 2 \lim_{Q \to \infty} \frac{1}{Q} \sum_{q=0}^{Q-1} G_D(x_q, y_q) \cdot G_{\perp D}(x_q, y_q),$$

and the first two terms in the equation above are $\Psi(F_D)(x, y)$ and $\Psi(F_{\perp D})(x, y)$ respectively. Therefore, to prove Claim (ii), it suffices to prove that the third term vanishes. This is equivalent to showing that for $\mu$-a.e. $x, y \in X$,

$$\lim_{Q \to \infty} \frac{1}{Q} \sum_{q=0}^{Q-1} (F_{\perp D}(x_q) - F_{\perp D}(y_q)) \cdot (F_D(x_q) - F_D(y_q)) = 0,$$

which follows from Lemma 12. This completes the proof of Claim (ii).

To prove Claim (iii), let $x_n$ and $y_n$ denote $\Phi^\Delta t(x)$ and $\Phi^\Delta t(y)$, respectively. Then, (9) can be rewritten for $F_{\perp D}$ as

$$\langle \Psi F_D \rangle(x, y) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} |F_{\perp D}(x_n)|^2 + \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} |F_{\perp D}(y_n)|^2$$

$$+ 2 \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} F_{\perp D}(x_n) F_{\perp D}(y_n).$$

The first two terms converge to the constant $\|F_{\perp D}\|_F^2$. It is therefore sufficient to show that the last term vanishes. Indeed, since the function $J : (x, y) \mapsto F_{\perp D}(x) F_{\perp D}(y)$ lies in the continuous spectrum subspace of the product-system $(X \times X, \Phi^t \times \Phi^t, \mu \times \mu)$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} F_{\perp D}(x_n) F_{\perp D}(y_n) = \langle J, 1_{X \times X} \rangle = 0.$$

The proof of Claim (iv) requires the following important observations. First, $F_D$ is a continuous map. This follows from the fact that $\{a_k\}_{k \in \mathbb{N}} \subset l^1$ and for each $k \in \mathbb{N}$, $|z_k|$ is constant and equals 1. The details are left to the reader. Second, the family $\{U^t F_D : t \in \mathbb{R}\}$ is equicontinuous. To see why, note that $U^t z_k = e^{i \omega_k t} z_k$, and therefore,

$$|U^t F_D(x) - U^t F_D(y)| \leq \left| \sum_{k=1}^{K} a_k e^{i \omega_k t} (z_k(x) - z_k(y)) \right| + 2 \sum_{k=K+1}^{\infty} |a_k|.$$
Since \( \{a_k\} \in \ell^1 \), given any \( \epsilon > 0 \), \( K \) can be chosen large enough so that the second sum in the right-hand side is less than \( \epsilon \). Then, the first sum is a finite sum of uniformly continuous functions, so there is a \( \delta > 0 \) such that if \( d(x, y) < \delta \), then this sum is less than \( \epsilon \). Thus, if \( d(x, y) < \delta \), then \(|U^t F_D(x) - U^t F_D(y)| < 2\epsilon\) which establishes equicontinuity of \( \{U^t F_D : t \in \mathbb{R}\} \). Observe now that the family \( \{(x, y) \mapsto |U^t F_D(x) - U^t F_D(y)|^2 : t \in \mathbb{R}\} \) is equicontinuous too. As a result, \( \Psi(F_D) \) is continuous by a classic result of Krengel ([49], Theorem 2.6).

Turning to Claim (v), it follows directly from Claims (iii) and (iv) that there exists a full-measure subset \( S \subseteq X \times X \) on which \( k_\infty \) is uniformly continuous. Suppose that \( S \) were not dense in \( X \times X \). Then, there would exist an open set \( B \subset X \times X \) disjoint from \( S \), and with positive measure (since \( X \times X \) is the support of \( \mu \times \mu \), and every open subset of the support of a Borel measure has positive measure), which would in turn imply that \( (\mu \times \mu)(S) < 1 \), leading to a contradiction. Therefore, \( S \) is a full-measure, dense subset of \( X \times X \), completing the proof of the claim.

Finally, the existence of \( \bar{\Psi}(F) \) in Claim (vi) follows from the fact that \( \Psi(F) \) is uniformly continuous on the dense subset \( S \) of the compact metric space \( X \times X \), and the almost uniform convergence of \( \Psi_Q(F) \) to \( \Psi(F) \) is a consequence of Egorov’s theorem.

Theorem 16 establishes that the function \( d_\infty : D(d_\infty) \mapsto \mathbb{R} \), such that

\[
d_\infty(x, y) := \lim_{Q \to \infty} d_Q(x, y); \quad (x, y) \in D(d_\infty) \subseteq X \times X
\]

is well-defined as a function in \( L^p(X \times X, \mu \times \mu), 1 \leq p \leq \infty \), with \( \sup d_\infty \leq \|d_1\|_{C^0(X \times X)} \). It can also be verified that \( d_\infty \) satisfies the triangle inequality and is non-negative. However, depending on the properties of the dynamical system and observation map, it may be a degenerate metric as \( d_\infty(x, y) \) may vanish for some \( x \neq y \), even if \( d_Q(x, y) \) is non-vanishing. In fact, it is easy to check that if \( y \) lies in the stable manifold of \( x \), then \( d_\infty(x, y) = 0 \). Analogously to the finite-delay case, we employ \( d_\infty \) to define a corresponding kernel \( k_\infty : M \times M \mapsto \mathbb{R} \), for some \( \epsilon > 0 \) (cf. (8)).

\[
k_\infty(x, y) = h_\epsilon(d_\infty(x, y)) = e^{-d_\infty^2(x, y)/\epsilon}, \quad (x, y) \in D(d_\infty), \quad (11)
\]

and \( k_\infty(x, y) = 0 \) otherwise. We also let \( K \) be the kernel integral operator from (3) associated with \( k_\infty \).

Proposition 17 shows that the operator \( K \) depends only on the “discrete component” of \( F \), and is a direct consequence of Theorem 16 and (10).

**Proposition 17.** Let \( (X, \Phi^t, \mu) \) and \( F \) be as in Theorem 1. Then, the integral operator \( K_\epsilon \) is a constant scaling operator iff its kernel \( k_\infty \) is a constant \( \mu \cdot a.e. \), which occurs iff \( F_D \) is a constant.

In general, \( k_\infty \) may not be continuous. Nevertheless, it has a number of other useful properties, which follow directly from Theorem 16 in conjunction with the boundedness and continuity of the Gaussian shape function.
Lemma 18. Under Assumption 1, the following hold:

(i) $k_{\infty}$ is the $L^p(X, \mu)$-norm limit, $1 \leq p < \infty$, of the sequence of continuous kernels $k_1, k_2, \ldots$.

(ii) $k_{\infty}$ is invariant under $U^t \times U^t$ for all $t \in \mathbb{R}$.

(iii) $k_{\infty}$ and $1/k_{\infty}$ both lie in $L^\infty(X \times X, \mu \times \mu)$, with

$$\|k_{\infty}\|_{L^\infty(X \times X)} \leq 1, \quad \|1/k_{\infty}\|_{L^\infty(X \times X)} \leq \exp(\|d_1\|_{C^0(X \times X)}/\epsilon).$$

Moreover, if Assumption 3 additionally holds:

(iv) $k_{\infty}$ is uniformly continuous on a dense, full-measure subset of $X \times X$.

(v) $k_{\infty}$ has a unique continuous representative $\bar{k}_{\infty} \in C^0(X \times X)$, and as $Q \to \infty$, $K_Q$ converges to $k_{\infty}$ almost uniformly.

The stronger regularity properties of $k_{\infty}$ under Assumption 3 have the following important implications on the behavior of the corresponding integral operator.

Lemma 19. Under Assumptions 1 and 3, the kernel integral operator $K_\epsilon$ associated with $k_{\infty}$ has the following properties:

(i) For every $f \in L^2(X, \mu)$, $Kf$ has a unique continuous representative.

(ii) For every $f \in C^0(X)$, $Kf$ is continuous.

(iii) $\|K\| \leq \|k_{\infty}\|_{L^\infty(X \times X)}$ in either $L^2$ or $C^0$ operator norm.

(iv) As an operator on $C^0(X)$, $K$ is compact.

(v) For every $f \in C^0(X)$, $K_Q f$ is a sequence of continuous functions converging $\mu$-a.e. to $Kf$.

Proof. (i) Since $k_{\infty}$ is uniformly continuous on a set $S \subseteq X \times X$ of full $\mu \times \mu$ measure, there exists a set $X' \subseteq X$ of $\mu$-measure 1, such that for every $x \in X'$, $k_{\infty}(x, \cdot)$ is continuous $\mu$-a.e. on $X$. Moreover, proceeding analogously to the proof of Theorem 16(v), it can be shown that $X'$ is dense in $X$. Let now $f \in L^2(X, \mu)$, $\|f\|_{L^2} = 1$. Then, for every $x_1, x_2 \in X'$,

$$|Kf(x_1) - Kf(x_2)| = \left| \int_{X'} [k_{\infty}(x_1, y) - k_{\infty}(x_2, y)] f(y) d\mu(y) \right| \leq \|k_{\infty}(x_1, \cdot) - k_{\infty}(x_2, \cdot)\|_{L^2} \|f\|_{L^2} \leq \|k_{\infty}(x_1, \cdot) - k_{\infty}(x_2, \cdot)\|_{L^\infty}. \quad (12)$$

lem:KInf Since $k_{\infty}$ is uniformly continuous on $S$, for every $\epsilon > 0$, there exists $\delta > 0$ such that if $d_\mu(x_1, x_2) < \delta$, $\|k_{\infty}(x_1, \cdot) - k_{\infty}(x_2, \cdot)\|_{L^\infty} < \epsilon$. Thus, for all such $x_1$ and $x_2$, we have $|Kf(x_1) - Kf(x_2)| < \epsilon$, which implies that $Kf$, restricted to $X'$, is uniformly continuous. As a result, since $X'$ is dense in the compact metric space $X$, $Kf|_{X'}$ has a unique continuous extension $g \in C^0(X)$. Moreover, since $X'$ has full measure, $g$ lies in the same $L^2$ equivalence class as $Kf$, proving the claim.
Proof. Since $k_\infty$ is uniformly continuous on a dense set of full measure, for any $f \in C^0(X)$, the function $g : X \times X \mapsto \mathbb{C}$ with $g(x, y) = k_\infty(x, y) f(y)$ has a unique continuous representative $\bar{g} \in C^0(X \times X)$. Therefore, for every $x \in X$, the function $k_\infty(x, \cdot) f$ is $\mu$-a.e. equal to $\bar{g}(x, \cdot)$ by $\mu$-a.e. continuity of $k_\infty(x, \cdot)$, and

$$Kf(x) = \int_X k_\infty(x, y) f(y) \, d\mu(y) = \int_X \bar{g}_\infty(x, y) f(y) \, d\mu(y).$$

It then follows that $Kf$ is continuous by continuity of integrals of $X$-sections of continuous functions on $X \times X$.

(iii) To verify the claim on the $L^2$ and $C^0$ operator norms, observe that for every $f \in L^2(X, \mu)$ and $x \in X'$, where $X'$ is as in the proof of Claim (i),

$$|Kf(x)| \leq \left| \int_{X'} k_\infty(x, y) f(y) \, d\mu(y) \right| \leq \|k_\infty(x, \cdot)\|_{L^2} \|f\|_{L^2} \leq \|k_\infty(x, \cdot)\|_{L^\infty} \|f\|_{L^2} \leq \|K_\infty\|_{L^\infty(X \times X)} \|f\|_{L^2},$$

and therefore

$$\|K_\infty f\|_{L^\infty} \leq \|k_\infty\|_{L^\infty(X \times X)} \|f\|_{L^2}. \quad (13)$$

The bound on the $L^2$ operator norm follows by setting $\|f\|_{L^2} = 1$ in (13), together with the fact that $\|K_\infty f\|_{L^2} \leq \|K_\infty\|_{L^\infty}$. The bound on the $C^0$ operator norm follows from (13) with $f \in C^0(X)$, in conjunction with the facts that $\|f\|_{L^2} \leq \|f\|_{C^0}$ and $\|K_\infty\|_{L^\infty} = \|Kf\|_{C^0}$.

(iv) Since, by the Arzelà-Ascoli theorem, every equicontinuous sequence of functions on a compact metric space has a limit point, it suffices to show that for every sequence $f_n \in C^0(X)$ with $\|f_n\|_{C^0} \leq 1$, the sequence $g_n = K f_n$ has a limit point with respect to $C^0$ norm. Let $\bar{k}_\infty \in C^0(X \times X)$ be the unique continuous representative of $k_\infty$. Then for any $x_1, x_2 \in X$, we have

$$|g_n(x_1) - g_n(x_2)| \leq \|\bar{k}_\infty(x_1, \cdot) - \bar{k}_\infty(x_2, \cdot)\|_{C^0},$$

and by uniform continuity of $\bar{k}_\infty$, for any $\epsilon > 0$, there exists $\delta > 0$, independent of $n$, such that, for every $x_1, x_2 \in X$ with $d(x_1, x_2) < \delta$, $|g_n(x_1) - g_n(x_2)| < \epsilon$. This establishes equicontinuity of $g_n$, and thus compactness of $K$ on $C^0(X)$.

(v) The continuity of $K_Q f$ and $K f$ follows from Claim (ii). The $\mu$-a.e. convergence follows from Lemma 18(v). \ \qed

We end this section with two important corollaries of Theorem 16 and Lemmas 18, 19, which are central to both Theorems 1 and 2.

**Corollary 20.** The operators $U^t$ and $K$ commute.

**Proof.** Since $\mu$ is an invariant measure, for every $x \in X$ and $t \in \mathbb{R}$ we have

$$Kf(x) = \int_X k_\infty(x, y) f(y) \, d\mu(y) = \int_X k_\infty(x, \Phi^t(y)) f(\Phi^t(y)) \, d\mu(y).$$
It therefore follows from Lemma 18(ii) that

\[ Kf(x) = \int_X k_\infty(\Phi^{-t}(x), y)f(\Phi_t(y)) \, d\mu(y) = U^{t*}KU^t f(x), \]

and the claim of the corollary follows. \(\square\) \(\square\)

**Corollary 21.** Under Assumption 1, the function \(\rho = K1_X\) is \(\mu\)-a.e. equal to a constant bounded away from zero (i.e., \(1/\rho\) lies in \(L^\infty(X, \mu)\)). Further, if Assumption 3 holds, then \(\rho|_X\) and \(1/\rho|_X\) are continuous.

**Proof.** Corollary 20 and the fact that \(U^t 1_X = 1_X\) imply that \(U^t \rho = \rho\), and it then follows by ergodicity that \(\rho\) is constant \(\mu\)-a.e. That \(\|1/\rho\|_{L^\infty}\) is finite follows from Lemma 18(iii). Finally, the continuity of \(\rho\) under Assumption 3 is a direct consequence of Lemma 19. \(\square\) \(\square\)

### 4.3. Markov normalization

Next, we construct the Markov operators \(P_Q\) and \(P\) appearing in Theorems 1 and 2 by normalization of \(K_Q\) and \(K\). Here, we use a procedure introduced in the diffusion maps algorithm [27] and further developed in [28], although there are also other normalizations with the same asymptotic behavior. Specifically, using the normalizing functions \(\rho_Q\) and \(\sigma_Q\) from Lemma 13 and \(\rho\) from Corollary 21, we introduce the kernels \(p_Q : M \times M \mapsto \mathbb{R}\) and \(p : M \times M \mapsto \mathbb{R}\), given by

\[ p_Q(x, y) = \frac{K_Q(x, y)}{\sigma_Q(x)\rho_Q(y)}, \quad p(x, y) = \begin{cases} k_\infty(x, y)/\rho(x), & \rho(x) > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{14} \]

respectively.

By Lemma 13, \(p_Q\) satisfies the boundedness and continuity properties in Assumption 2. On the other hand, \(p\) is neither guaranteed to be continuous nor bounded on arbitrary compact sets, but it nevertheless follows from Lemma 18 and Corollary 21 that both \(p\) and \(1/p\) lie in \(L^\infty(X \times X)\). Based on these facts, we can therefore define the kernel integral operators \(P_Q : L^2(X, \mu) \mapsto L^2(X, \mu)\) and \(P : L^2(X, \mu) \mapsto L^2(X, \mu)\) from (3) associated with the kernels \(p_Q\) and \(p\), respectively, and these operators are both Hilbert-Schmidt (see Section 4.1). Note that \(p\) and \(P\) have analogous properties to those stated for \(k_\infty\) and \(K\) in Lemmas 18, 19 and Corollary 20. In particular, \(p\) is invariant under \(U^t \times U^t\), and \(P\) commutes with \(U^t\).

The operators \(P_Q\) and \(P\) can also be obtained directly from \(K_Q\) and \(K\), respectively, through the sequence of operations

\[ \bar{K}_Q f := K_Q \left( \frac{f}{K_Q 1_X} \right), \quad P_Q f = \frac{\bar{K}_Q f}{\bar{K}_Q 1_X}, \quad P f = \frac{K f}{K 1_X}. \tag{15} \]

In [28], the steps leading to \(\bar{K}_Q\) from \(K_Q\) and to \(P_Q\) from \(\bar{K}_Q\) are called right and left normalization, respectively. In the case of \(P\), the effects of right normalization cancel...
since $K^1_X$ is $\mu$-a.e. constant by Corollary 21, so it is sufficient to construct this operator directly via left normalization of $K$.

As is evident from (15), $P_Q$ and $P$ are both Markov operators preserving constant functions. Moreover, for all $x \in M$ we have $\int_X p_Q(x, \cdot) \, d\mu = 1$, and for all $x \in X$, $\int_X p(x, \cdot) \, d\mu = 1$, i.e., both $p_Q$ and $p$ are transition probability kernels. In particular, since $X$ is compact and $p_Q$ and $p$ are essentially bounded below, $P_Q$ and $P$ are both ergodic Markov operators; that is, their eigenspaces at eigenvalue 1 are one-dimensional.

The Markov kernel $p$ is $\mu$-a.e. symmetric by symmetry of $k_\infty$ and the fact that $\rho$ is $\mu$-a.e. constant. As a result, $P$ is self-adjoint, its eigenvalues admit the ordering $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots$, and there exists a real orthonormal basis $L^2(X, \mu)$ consisting of corresponding eigenfunctions, $\phi_k$, with $\phi_0$ being constant. On the other hand, because $P$ is not symmetric, the operator $P_Q$ is not self-adjoint, but is nevertheless related to a self-adjoint operator via a similarity transformation by a bounded multiplication operator with a bounded inverse. To verify this, define

$$
\hat{\sigma}_Q = \sigma_Q/\rho_Q, \quad \hat{\sigma}_Q = \sqrt{\sigma_{\epsilon, Q} \rho_Q},
$$

where $\rho_Q$ and $\sigma_Q$ are as in Lemma 13. Let also $D_Q$ be the multiplication operator which multiplies by $\hat{\sigma}_Q$, and $\hat{P}_Q$ the kernel integral operator with kernel $\hat{p}_Q : M \times M \mapsto \mathbb{R}$

$$
\hat{p}_Q(x, y) = \frac{k_Q(x, y)}{\hat{\sigma}_Q(x) \hat{\sigma}_Q(y)}.
$$

Observe now that $\hat{P}_Q$ is a symmetric operator, and $P$ is related to $\hat{P}_Q$ via the similarity transformation

$$
\hat{P}_Q = D_Q^{1/2} P_Q D_Q^{-1/2};
$$

that is, for every $f \in L^2(X, \mu)$,

$$
D_Q^{1/2} P_Q D_Q^{-1/2} f(x) = \int_X \frac{\sigma_Q(x)}{\rho_Q(x) \sigma_Q(x) \rho_Q(y)} k_Q(x, y) \sqrt{\rho_Q(y)} \sigma_Q(y) d\mu(y) = \int_X \frac{k_Q(x, y)}{\sigma_Q(x) \sigma_Q(y)} f(y) d\mu(y) = \hat{P}_Q f(x).
$$

The following are useful properties of $\hat{P}_Q$ that follow from its relation to $P_Q$.

(i) $\hat{P}_Q$ has the same discrete spectrum as $P_Q$, consisting of eigenvalues $\lambda_{k,Q}$ with $1 = \lambda_{0,Q} > \lambda_{1,Q} \geq \lambda_{2,Q} \geq \cdots$.

(ii) Let $\phi_{k,Q}$ denote the eigenfunctions of $\hat{P}_Q$ corresponding to the nonzero eigenvalues $\lambda_{k,Q}$. These form an orthonormal basis for the closed subspace $\text{ran } \hat{P}_Q = (\ker \hat{P}_Q)^\perp$.

Moreover, the $\phi_{k,Q}$ can be chosen to be real-valued.

(iii) The eigenfunction $\phi_{0,Q}$ of $\hat{P}_Q$ is equal up to proportionality constant to $\rho_Q \hat{\sigma}_Q^{1/2}$.

**Remark 22.** The class of integral operators studied in this work has previously been used for dimension reduction and mode decomposition of high-dimensional time series.
Delay-coordinate maps and the spectra of Koopman operators (e.g., [50–53]). In these works, a phenomenon called in [52] “timescale separation” was observed; namely, it was observed that at increasingly large $Q$ the eigenfunctions of $P_Q$ capture increasingly distinct timescales of a multiscale input signal. Theorems 1 and 2 provide an interpretation of this observation from the point of view of spectral properties of Koopman operators; in particular, from the fact that $P_Q$ has, in the limit $Q \to \infty$, common eigenfunctions with $U_t$ and the latter capture distinct timescales associated with the eigenfrequencies $\omega$. Analogous results should also hold for other classes of compact operators for data analysis that employ delays, including the covariance operators used in singular spectrum analysis [54–56] and the related Hankel matrix analysis [10, 13, 14].

Remark 23. In applications, it may be the case that $\rho_Q$ and $1/\rho_Q$ take a large range of values. In such situations, it may be warranted to replace (8) by a variable-bandwidth kernel of the form $K_Q(x, y) = \exp \left( -\frac{d_Q^2(x, y)}{r_Q(x) r_Q(y)} \right)$, with a bandwidth function $r_Q$ introduced so as to control the decay of the kernel away from the diagonal, $x = y$. Various types of bandwidth functions have been proposed in the literature, including functions based on neighborhood distances [57, 58], state space velocities [51, 59], and local density estimates [30]. While we do not study variable bandwidth techniques in this work, our approach should be applicable in that setting too, so long as Corollary 21 holds.

5. Proof of Theorems 1–3 and Corollary 4

Proof of Theorem 1. That $P$ and $U^t$ commute follows from the invariance of $p$ under $U^t \times U^t$ and an analogous calculation to that in the proof of Corollary 20. Next, as $Q \to \infty$, $p_Q$ converges to $p$ in any $L^p(X \times X, \mu \times \mu)$ norm with $1 \leq p < \infty$ by the analogous result to Lemma 18(i) that holds for these kernels (see Section 4.3). In particular, that $p_Q$ converges to $p$ in $L^2(X \times X, \mu \times \mu)$ norm implies that $P_Q$ converges to $P$ in $L^2(X, \mu)$ operator norm, since $P_Q - P$ is Hilbert-Schmidt and thus bounded in operator norm by $\|p_Q - p\|_{L^2(X \times X)}$. \qed

Proof of Theorem 2. We first establish that $\tau$ is a.e. invariant under $\Phi^t \times \Phi^t$. Since the integral operator $T$ commutes with $U^t$, for $\mu$-a.e. $x \in X$,

$$\int_X \tau(\Phi^t(x), \Phi^t(y')) f(\Phi^t(y')) \, d\mu(y') = \int_X \tau(\Phi^t(x), y) f(y) \, d\mu(y)$$

$$= U^t T f(x) = T(U^t f)(x) = \int_X \tau(x, y') f(\Phi^t(y')) \, d\mu(y'),$$

where the second equality was obtained by the change of variables $y = \Phi^t(y')$, and utilizes the invariance of the measure $\mu$ under $\Phi^t$. The only way the terms at the two ends of the equation can be equal for $\mu$-a.e. $x \in X$ is if $\tau(\Phi^t(x), \Phi^t(y')) = \tau(x, y')$ $\mu$-a.e.

Next, observe that, by (2), the space $L^2(X \times X, \mu \times \mu)$ splits as the $U^t \times U^t$-invariant orthogonal sum of $\mathcal{D} \otimes \mathcal{D}$, $\mathcal{D} \perp \mathcal{D} \perp$, $\mathcal{D} \perp \mathcal{D}$, and $\mathcal{D} \otimes \mathcal{D} \perp$. Since $\tau$ is an $L^2$ kernel, it has
orthogonal projections onto each of these subspaces, all of which are $U^t \otimes U^t$-invariant by the invariance of $\tau$ just established. By symmetry of $\tau$, the projections onto $D^\perp \otimes D$ and $D \otimes D^\perp$ vanish. Moreover, the projection $\tau_{D^\perp \otimes D^\perp} \in D^\perp \otimes D^\perp$ is orthogonal to constant functions, and it follows by the Birkhoff ergodic theorem that for $\mu \times \mu$-a.e. $x, y \in X \times X$,

$$0 = \langle 1_{X \times X}, \tau_{D^\perp \otimes D^\perp} \rangle$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \tau_{D^\perp \otimes D^\perp}(\Phi_n^t(x), \Phi_n^t(y))$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \tau_{D^\perp \otimes D^\perp}(x, y)$$

$$= \tau_{D^\perp \otimes D^\perp}(x, y).$$

This completes the proof of Claim (i). The statements in Claim (ii) that $D^\perp \subset \ker(T)$ and that $D$ and $D^\perp$ are invariant under $T$ are direct consequences of Claim (i).

The remaining two claims in the theorem, which requires that both $D$ and $\text{ran} T$ contain non-constant functions, can be proved by means of the following, slightly stronger, result.

**Proposition 24.** For any nonzero eigenvalue $\lambda$ of $T$, the corresponding eigenspace $W_\lambda$ is invariant under the action of the Koopman generator $V$, and $V|_{W_\lambda}$ is diagonalizable. Moreover, the constant function $1_X$ is an eigenfunction of $T$. If $W_\lambda$ does not contain $1_X$, its dimension is an even number.

**Proof.** Since $T$ is compact, every nonzero eigenvalue $\lambda$ has finite multiplicity and its corresponding eigenspace $W_\lambda$ has finite dimension, $l = \dim W_\lambda$. Since $U^t$ commutes with $T$, $U^t$ and hence $V$ leave $W_\lambda$ invariant. Similarly, since the constant function is an eigenfunction of $V$, it is an eigenfunction of $T$.

Let $\lambda_0$ be the eigenvalue of $T$ corresponding to the constant eigenfunction, and $\lambda \neq \lambda_0$ be any other eigenvalue of $T$. Then, $V|_{W_\lambda}$ is a skew-symmetric operator on a finite-dimensional space, and thus can be diagonalized with respect to a basis of simultaneous eigenfunctions of $T$ and $V$. Fix any element $\zeta$ of this basis. By our choice of $\lambda$, $\zeta$ is a non-constant eigenfunction of $V$, hence $\langle \zeta, 1 \rangle = 0$. Therefore, by ergodicity of $(\Phi^t, \mu)$, $V\zeta = i\omega \zeta$ for some $\omega \neq 0$. This implies that $\zeta$ has non-zero real and imaginary parts. Hence, the conjugate $\zeta^*$ is linearly independent from $\zeta$ and corresponds to eigenvalue $-i\omega$ of $V$. However, since $T$ is a real operator, $\zeta^*$ lies in $W_\lambda$. We therefore conclude that $W_\lambda$ can be split into disjoint 2-dimensional spaces spanned by the conjugate pair of eigenfunctions $\zeta$ and $\zeta^*$. Therefore $\dim W_\lambda$ is an even number.

By Proposition 24, $U^t$ and $T$ have joint eigenfunctions, each of which factors the dynamics into a rotation on the circle in accordance with (1). According to
Delay-coordinate maps and the spectra of Koopman operators

Proposition 9, any collection of \(D\) such eigenfunctions factors the dynamics into a rotation on \(\mathbb{T}^D\). This proves Claim (iii).

To prove Claim (iv), we use (6) to expand the kernel as
\[
\tau = \sum_{\vec{a},\vec{b} \in \mathbb{Z}^m} \tilde{\tau}_{\vec{a} \vec{b}} z_{\vec{a}} \otimes z_{\vec{b}}.
\]
In this expansion, there is a minimal number \(D \leq m\) of generating eigenfunctions \(z_j\) from (6), arranged without loss of generality as \(z_1, \ldots, z_D\), such that the expansion coefficients \(\tilde{\tau}_{\vec{a} \vec{b}}\) corresponding to \(\vec{a} = (a_1, \ldots, a_m)\) and \(\vec{b} = (b_1, \ldots, b_m)\) with nonzero \(a_{D+1}, \ldots, a_m\) and \(b_{D+1}, \ldots, b_m\), respectively, vanish (in other words, the kernel \(\tau\) does not project onto the subspaces generated by \(z_{D+1}, \ldots, z_m\) and their powers). By Proposition 9, the Koopman eigenfunctions corresponding to non-vanishing \(\tilde{\tau}_{\vec{a} \vec{b}}\) can be expressed as \(z_{\vec{a}} = \zeta_{\vec{a}} \circ \pi\), where the \(\zeta_{\vec{a}}\) are smooth Koopman eigenfunctions on \(\mathbb{T}^D\) associated with an ergodic rotation. Thus, denoting the index set for the nonzero \(\tilde{\tau}_{\vec{a} \vec{b}}\) by coefficients by \(I \in \mathbb{Z}^m \times \mathbb{Z}^m\), we have
\[
\tau(x, y) = \hat{\tau}(\pi(x), \pi(y)) \text{ for } \mu \times \mu\text{-a.e., } (x, y) \in X \times X,
\]
where \(\hat{\tau}\) is the \(L^2\) kernel on \(\mathbb{T}^D\) given by
\[
\hat{\tau} = \sum_{\vec{a}, \vec{b} \in I} \tilde{\tau}_{\vec{a} \vec{b}} \zeta_{\vec{a}} \otimes \zeta_{\vec{b}}.
\]
This completes the proof of Claim (v) and of Theorem 2.

\(\square\)

Proof of Theorem 3. That \(p\) is uniformly continuous on a full-measure, dense subset of \(X \times X\) follows from the analogous result to Lemma 18(iv), which holds for \(p\) (see Section 4.3). Claims (i)-(iv) of the theorem follow analogously to Lemma 19.

\(\square\)

Rates of convergence in the continuous case. As an auxiliary result, we state a lemma that establishes rates of convergence with respect to the number of delays \(Q\) of the kernel integral operators studied in this work.

Lemma 25 (Convergence of commutators). Under the assumptions of Theorem 3, the following operators converge in \(C^0(X)\) operator norm to 0 as \(Q \to \infty\), with rates given below:

\(\begin{align*}
(i) & \quad \|U^{\Delta t} K_Q - K_Q U^{\Delta t}\| = O((Q\epsilon)^{-1}), \\
(ii) & \quad \|U^{\Delta t} \tilde{K}_Q - \tilde{K}_Q U^{\Delta t}\| = O((Q\epsilon)^{-1}), \\
(iii) & \quad \|U^{\Delta t} P_Q - P_Q U^{\Delta t}\| = O((Q\epsilon)^{-1}).
\end{align*}\)

Proof. Let \(\bar{F}_{Q,\Delta t}(x, y) := \|F(x) - F(y)\| - \|F(\Phi^Q\Delta t x) - F(\Phi^Q\Delta t y)\|\), and notice that by continuity of \(F\) and compactness of \(X\) this quantity is bounded on \(X \times X\). Thus, since \(d_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y)) = d_Q(x, y) + Q^{-1} \bar{F}_{Q,\Delta t}(x, y)\), we have
\[
k_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y)) = k_Q(x, y) \exp \left( (Q\epsilon)^{-1} \bar{F}_{Q,\Delta t}(x, y) \right)
\]
\[
= k_Q(x, y) \left( 1 + O((Q\epsilon)^{-1}) \right).
\]
Therefore, for every $f \in L^2(X, \mu)$ and $x \in X$ we have

\[
U^t K_Q f(x) = \int_X k_Q(\Phi^t(x), y) f(y) \, d\mu(y) \\
= \int_X k_Q(\Phi^t(x), \Phi^t(y)) f(\Phi^t(y)) \, d\mu(y) \\
= [1 + O((Q\epsilon)^{-1})] K_Q U^t f(x),
\]

and note that we have used the fact that $\mu$ is an invariant measure in the second-to-last line. Next, since $K_Q$ is a kernel integral operator with a continuous kernel, it follows from the Cauchy-Schwarz inequality that

\[
\|K_Q f\|_{C^0} \leq \|K_Q\|_{C^0} \|f\|_{L^2}.
\]

Substituting this result in the right-hand side of (18) and taking the supremum over $x \in X$ yields

\[
\|\left(U^t K_Q - K_Q U^t\right)f\|_{C^0} = O((Q\epsilon)^{-1}) \|f\|_{L^2}.
\]

Claim (i) then follows from the fact that $\|\cdot\|_{L^2} \leq \|\cdot\|_{C^0}$. Claims (ii) and (iii) can be proved in a similar manner.

**Proof of Corollary 4** Claim (i) is a consequence of the uniform convergence $P_Q \to P$ and the fact that these operators are both compact on $L^2(X, \mu)$. These conditions also imply the strong convergence $\Pi_Q \to \Pi$ of the projectors in Claim (ii). It remains to show that the subspaces $W_Q = \text{ran } \Pi_Q$ converges to the eigenspace $W_l = \text{ran } \Pi$ of $P$, in the sense of a quantity called gap. We will prove this in a more general context of a Banach space. Given two closed subspaces $W$ and $W'$ of a Banach space, this quantity is defined as [see [32], Section 7]

\[
\delta(W, W') = \min \left\{ \tilde{\delta}(W, W'), \tilde{\delta}(W', W) \right\}, \\
\tilde{\delta}(W, W') = \min_{v \in W, \|v\|=1} d(v, W'), \\
d(v, W') = \inf_{w \in W'} \|v - w\|. 
\]

The following lemma establishes that strong convergence of projection operators onto subspaces of constant finite dimension implies convergence of those subspaces in the sense of gap.

**Lemma 26.** Let $B$ be a Banach space, and $\pi_N : B \mapsto B$ a sequence of projection operators onto subspaces $W_N$ of constant finite dimension $k$. Let also $\pi_N$ converge strongly to a projection $\pi : B \mapsto B$ onto a subspace $W$ (which is necessarily $k$-dimensional). Then:

(i) $\pi_N \circ \pi$ converges to $\pi$ in operator norm.

(ii) The subspaces $W_N$ converge to $W$, i.e., $\delta(W, W_N)$ converges to 0.

**Proof.** Let $W'$ be the kernel of $\pi$. Then $B = W \oplus W'$ and $W' \subset \ker(\pi_N \circ \pi)$. Thus, to prove Claim (i), it is enough to prove that $\pi_N \circ \pi$ converges to $\pi$ in the norm on the subspace $W$. This however follows from the fact that pointwise convergent operators on finite-dimensional spaces are also norm-convergent.
To prove Claim (ii), it will be first shown that $\tilde{\delta}(W, W_N) \to 0$. Let $u \in W$, $\|u\|=1$. Then $\pi_N u \to \pi u = u$, so $d(u, W_N) \leq d(u, \pi_N u)$, and therefore for every $u \in W$, $d(u, W_N) \to 0$. Let $S^1(W)$ denote the set of unit-norm vectors in $W$. Since $W$ is finite dimensional, $\sup_{u \in S^1(W)} d(u, W_N) \to 0$, so $\tilde{\delta}(W, W_N) \to 0$.

It now remains to be shown that $\tilde{\delta}(W_N, W) \to 0$. Let $\{e_i\}_{i=1}^k$ be a basis of unit vectors for $W$. Then, by Claim (i), $e_i^{(N)} := \pi_N e_i$ forms a basis for $W_N$ for large-enough $N$. Let $S_N$ denote the convex hull of $\{\pm e_i^{(N)}\}_{i=1}^k$. The important observation is that there exists $R > 0$ such that for any $y$ in $S_N$, $\|y\|_\infty < R$. Therefore, since we have already established that $\Pi_Q \to \Pi$ strongly, Lemma 26 applies, and it follows that $\tilde{\delta}(W, W_Q) \to 0$. This completes the proof of Claim (ii) and the lemma.

Returning to the proof of Corollary 4, it follows from the compactness of $P_Q$ and $P$ and Claim (i) that the $\Pi_Q$ have finite-dimensional ranges, $W_Q$, and there exists a stage $Q_0$ such that for all $Q > Q_0$, $\dim W_Q = \alpha$. Therefore, since we have already established that $\Pi_Q \to \Pi$ strongly, Lemma 26 applies, and it follows that $\tilde{\delta}(W, W) \to 0$. This completes the proof of Corollary 4.

6. Galerkin approximation of Koopman eigenvalue problems

In this section, we formulate a Galerkin method for the eigenvalue problem of the Koopman generator $V$ in the eigenbasis of $P$, under the implicit assumption that the latter operator is available to us from $P_Q$ after having taken a large number of delays $Q$. The task of finding the eigenvalues of $V$ has two challenges, namely, (i) $V$ is an unbounded operator defined on a proper subspace $D(V) \subset L^2(X, \mu)$ which is not known a priori; (ii) the spectrum of $V$ could be dense in $i\mathbb{R}$ (even for a pure point spectrum system such an ergodic rotation on $\mathbb{T}^D$ with $D \geq 2$; e.g., [15], Remark 8), in which case, solving for its eigenvalues is a numerically ill-posed problem. Following [12, 15], we will address these issues by employing a Galerkin scheme for the eigenvalue problem of $V$ with a small amount of diffusion added for regularization. Our approach has the following steps.

**Step 1. Sobolev spaces.** We first construct subspaces of $L^2$ in which we search for eigenfunctions. These spaces will be shown to be dense in $H$, defined as the closed subspace of $\text{ran} P$ orthogonal to constant functions (that is, $H$ only consists of zero-mean functions). Note that since the eigenfunctions $\phi_k$ of $P$ are mutually orthogonal, and $\phi_0$ is constant, $H$ is also the $L^2$ closure of the span of $\{\phi_1, \phi_2, \ldots\}$. For any $p \geq 0$,
we define
\[
H^p = \left\{ \sum_{k=1}^{\infty} c_k \phi_k \in H : \sum_{k=1}^{\infty} |c_k|^2 |\eta_k|^p < \infty \right\}, \quad \eta_k = (\lambda_k^{-1} - 1)/\epsilon. \tag{20}
\]

The spaces \(H^p\) are analogous to the usual Sobolev spaces associated with self-adjoint, positive-semidefinite, unbounded operators with compact resolvents and discrete spectra (here, \(\{\eta_k\}_{k \in \mathbb{N}}\)). In particular, when \((X, g)\) is a smooth Riemannian manifold with a metric tensor \(g\) satisfying \(\text{vol}_g = \mu\), and \((\eta_k, \phi_k)\) are the eigenvalues and orthonormal eigenfunctions of the corresponding Laplace-Beltrami operator, then \(H^p\) becomes the canonical Sobolev space \(H^p(X, g)\), restricted to be orthogonal to constant functions. \(H^p\) from (20) is a Hilbert space with the inner product
\[
\langle f, g \rangle_{H^p} := \sum_{q=0}^{p} \sum_{k=1}^{\infty} c_k^* d_k |\eta_k|^q,
\]
where \(f = \sum_{k=1}^{\infty} c_k \phi_k\) and \(g = \sum_{k=1}^{\infty} d_k \phi_k\). Moreover, \(\{\phi_1^{(p)}, \phi_2^{(p)}, \ldots\}\) with \(\phi_k^{(p)} = \phi_k/\|\phi_k\|_p\), \(\|\phi_k\|_{p}^2 = \sum_{q=0}^{p} \lambda_k^q\), forms an orthonormal basis of \(H^p\).

**Proposition 27.** For every \(p > 0\), the space \(H^p\) is dense in \(H\) and moreover, the inclusion map \(H^p \hookrightarrow H\), and thus \(H^p \hookrightarrow L^2(X, \mu)\), is compact.

**Proof.** To see that \(H^p\) is dense, note that \(H^p\) includes all finite linear combinations of the \(\phi_k\). Since the \(\phi_k\) are an orthonormal basis of \(H\), these finite linear combinations are dense in \(H\). Next, the embedding of \(H^p\) in \(D\) can be represented by a diagonal operator \(G : H^p \hookrightarrow H\) such that \(G_{kk} := \langle \phi_k, G\phi_k^{(p)} \rangle = (\eta_k)^{-p/2}\). This operator is compact iff \(G_{kk}\) converges to 0 as \(k \to \infty\). This is true by (22) below. The compactness of the inclusion \(H^p \hookrightarrow L^2(X, \mu)\) follows immediately. \(\square\)

**Remark 28.** In (21), we have tacitly assumed that the \(\eta_k\) are all positive, or, equivalently, that \(P\) is a positive operator. In the event that \(\eta_k < 0\) for \(k > k^*\) (with \(k^*\) the index of the smallest positive eigenvalue), the definition of \(\eta_k\) can be modified for \(k > k^*\) to the positive, increasing sequence \(\eta_k(\cdot/k^*)^2\). It should be noted that this modification is mainly formal since for any finite spectral truncation dimension \(m\) and \(\epsilon\) small enough, the first \(m\) eigenvalues of \(P\) (and its data-driven approximation in Section 7 ahead) are positive.

**Step 2. Regularized generator.** For every \(\theta > 0\), we define the unbounded operators 
\[
\Delta : D(\Delta) \hookrightarrow H \quad \text{and} \quad L_\theta : D(L_\theta) \hookrightarrow H,
\]
where
\[
\Delta := f \mapsto \sum_{k=1}^{\infty} \eta_k \langle \phi_k, f \rangle \phi_k, \quad L_\theta := V - \theta \Delta. \tag{21}
\]

As we will see in Step 3 below, the role of the diffusion term \(\theta \Delta\) is to penalize the eigenfunctions of \(V\) with large eigenvalues of a Dirichlet energy functional. Theorem 29 below identifies a domain in which the operators in (21) are continuous, and establishes that the eigensolutions of \(L_\theta\) converge to eigensolutions of \(V\) as \(\theta \to 0\).
Theorem 29. Viewed as operators from $H^2$ to $H$, the generator $V$, as well as the operators $L_\theta$ and $\Delta$ from (21), are bounded. In particular, we can set $D(\Delta) = D(L_\theta) = H^2$. Moreover, all of $V$, $\Delta$, and $L_\theta$ are block diagonalizable with $2 \times 2$ blocks in the eigenbasis $\{\phi_1, \phi_2, \ldots\}$ of $P$, and therefore commute with each other on the eigenspaces of $P$ corresponding to nonzero eigenvalues. Finally, the eigenvalues of $L_\theta$ converge to the eigenvalues of $V|_{H^2}$ as $\theta \to 0^+$. 

Proof. By Theorem 1 and (21), the operators $V$, $\Delta$, and $L_\theta$ are diagonal with respect to the orthonormal set $\{\phi_1, \phi_2, \ldots\}$ and annihilate functions outside the span of this set. Therefore, they commute with each other.

By Proposition 24, the eigenfunctions $\phi_k$ can be chosen so that $V\phi_{2k-1} = -\omega_k \phi_{2k}$ and $V\phi_{2k} = i\omega_k \phi_{2k-1}$ for some eigenfrequency $\omega_k$. From this it follows that $V$ commutes with $\Delta$ and $L_\theta$. To see that $V$ is a bounded operator on $H^2$, first observe that $\omega_k = O(k)$, which follows from the fact that the eigenvalues of $V$ are integer linear combinations of $m$ rationally independent frequencies (see Section 3). By Theorem 3, the kernel $p$ associated with $P$ is $L^2$ integrable, and thus by a result of Ferreira and Menegatto on integral operators ([48], Corollary 2.5), $\lambda_k = o(k^{-1})$. Combining these estimates, we obtain

$$k = o(\eta_k), \quad \omega_k = o(\eta_k), \quad \eta_k^{-1} = o(k^{-1}),$$

(22)

and therefore deduce that there exists a constant $C > 0$ such that

$$\omega_k \leq C \eta_k; \quad \forall k \in \mathbb{N}$$

(23)

Hence, for $f = \sum_{k=1}^{\infty} c_k \phi_k \in H^2$,

$$\|Vf\|^2 = \left\| \sum_{k=1}^{\infty} c_k V\phi_k \right\|^2 = \left\| \sum_{k=1}^{\infty} ic_k \omega_k \phi_k \right\|^2 \leq C^2 \sum_{k=1}^{\infty} |c_k|^2 |\eta_k|^2 \leq C^2 \|f\|^2_{H^2},$$

proving that $V$ is a bounded operator on $H^2$. The same reasoning applies for $L_\theta$ and $\Delta$.

Finally, the eigenvalue of $L_\theta$ corresponding to $\phi_k$ is $i \omega_k - \theta \eta_k$, which converges to $i \omega_k$ as $\theta \to 0^+$. This completes the proof of Theorem 29. 

Remark 30. Theorem 29 establishes that $H^2$ is a domain on which $V$ is a bounded operator, but if $X$ had a smooth manifold structure, it is possible to show that the standard $H^1$ Sobolev space associated with a Riemannian metric on $X$ is also a suitable domain. In this work, $X$ has no smooth structure, and we can state Theorem 29 above only for $V|_{H^2}$. In separate calculations, we have observed that an analog of the weak eigenvalue problem for $L_\theta$ formulated in $H^1 \times H^1$ actually performs well numerically.

Step 3. Galerkin method. By virtue of Theorem 29, the eigenvalues of $L_\theta$ can be considered to be approximations of the eigenvalues of $V$. We will take the Galerkin approach in finding the eigenvalues of $L_\theta$ by solving for $z \in H^2$ and $\gamma \in \mathbb{C}$ in the following variational (weak) eigenvalue problem:
**Definition 31** (Regularized Koopman eigenvalue problem). Find $\lambda \in \mathbb{C}$ and $z \in H^2$ such that for all $f \in H$,

$$A(f, z) = \langle f, z \rangle,$$

where $A : H \times H^2 \mapsto \mathbb{C}$ is the sesquilinear form defined by

$$A(g, f) = \langle g, L_{\theta} f \rangle = \langle g, V f \rangle - \theta E(g, f), \quad E(g, f) = \langle g, \Delta f \rangle.$$

In the above, the form $E : H \times H^2 \mapsto \mathbb{C}$ induces a Dirichlet energy functional $E(f) = E(f, f)$, $f \in H^2$, providing a measure of roughness of functions in $H^2$. In particular, if $X$ were a smooth Riemannian manifold, and the $\eta_k$ were set to Laplace-Beltrami eigenvalues, we would have $E(f) = \int_X \| \text{grad} f \|^2 d\mu$. While the lack of smoothness of $X$ in our setting precludes us from defining $E$ by means of a gradient operator, its definition in terms of the $\eta_k$ from (20) still provides a meaningful measure of roughness of functions. For instance, it follows from results in spectral graph theory that the variance of estimates $\eta_k^{(N)}$ of the $\eta_k$ computed from finite data sets (e.g., as described in Section 7 ahead) increases with $k$ [31, 58], which is consistent with the intuitive expectation that rough (highly oscillatory) functions require larger numbers of samples for accurate approximations.

Following [12, 15], we will order all solutions $(\gamma_k, z_k)$ of the problem in Definition 31 in order of increasing Dirichlet energy $E(z_k)$. Since $A(f, f) = -\theta E(f, f)$ by skew-symmetry of $V$, we can compute the Dirichlet energy of eigenfunction $z_k$ directly from the corresponding eigenvalue, viz. $E(z_k) = -\text{Re} \gamma_k / \theta$. Similarly, we have $\omega_k = \text{Im} \gamma_k$.

By (23), there exist constants $C_1, C_2 > 0$ such that

$$C_2 \leq \left| \frac{i \omega_k - \theta \eta_k}{\eta_k} \right| \leq C_1, \forall k \in \mathbb{N}. \quad (24)$$

To justify the well-posedness of the eigenvalue problem in Definition 31, we will state three important properties of $A$, namely,

$$|A(u, v)| \leq C_1 \|u\|_H \|v\|_{H^2}, \forall u \in H, \forall v \in H^2, \quad (25)$$

$$\sup_{\|f\|_H = 1} |A(f, v)| \geq C_2 \|v\|_{H^2}^2, \forall v \in H^2, \quad (26)$$

$$\sup_{\|g\|_{H^2} = 1} |A(u, g)| \geq C_2 \|u\|^2_H, \forall u \in H. \quad (27)$$

We now give brief proofs of these results. In the following, $v = \sum_{k=1}^{\infty} d_k \phi_k$ and $u = \sum_{k=1}^{\infty} c_k \phi_k$ will be arbitrary functions in $H^2$ and $H^0$, respectively. First, note that,

$$|A(u, v)| = \left| \sum_{k=1}^{\infty} (i \omega_k - \theta \eta_k) c_k^* d_k \right| \leq \sum_{k=1}^{\infty} |i \omega_k - \theta \eta_k| |c_k^* d_k|.$$
By the Cauchy-Schwartz inequality on \( l^2 \) and (24),
\[
|A(u, v)| \leq C_1 \sum_{k=1}^{\infty} |\eta_k| c_k^* d_k \leq C_1 \|u\|_H \|v\|_{H^2},
\]
proving (25). To prove (26), let \( f = \sum_{k=1}^{\infty} a_k \phi_k \in H \). Then, the left-hand side of that equation becomes \( \sum_{k=1}^{\infty} (i \omega_k/\eta_k - \theta) a_k^* d_k \). Let \( R_k := i \omega_k/\eta_k - \theta \), where \( |R_k| \geq C_2 \) by (24). By the Cauchy-Schwartz inequality, under the constraint \( \sum_{k=1}^{\infty} |a_k|^2 = 1 \), the sum \( |\sum_{k=1}^{\infty} a_k^* d_k| \) attains the maximum value of \( \sum_{k=1}^{\infty} |\eta_k|^2 |d_k|^2 \). Therefore,
\[
\sup_{f \in H} |A(f, v)| = \sup_{\sum_{k=1}^{\infty} |a_k|^2 = 1} \left| \sum_{k=1}^{\infty} a_k^* d_k R_k \eta \right| \geq C_2 \sum_{k=1}^{\infty} |\eta_k d_k|^2 = C_2 \|v\|_{H^2}^2.
\]
This proves (26). The proof of (27) is similar to that of (26), with \( f \) replaced by a trial function \( g = \sum_{k=1}^{\infty} b_k \phi_k \in H^2 \) and the constraint \( \|g\|_{H^2} = \sum_{k=1}^{\infty} |b_k|^2 = 1 \). A direct consequence of (26) and (27) is,
\[
\inf_{v \in H^2} \sup_{\|v\|_{H^2} = 1} |A(u, v)| \geq C_2, \quad \inf_{u \in H} \sup_{\|u\|_{H^2} = 1} |A(u, v)| \geq C_2.
\]
Equations (25), (26), (28), and the compact embedding of \( H^2 \) in \( H \) by Proposition 27 together guarantee that the eigenvalues of \( A \) restricted to the finite-dimensional subspaces of \( H \times H^2 \) spanned by the leading \( m \) eigenfunctions \( \phi_1, \ldots, \phi_m \) converge, as \( m \to \infty \), to the weak eigenvalues of \( L_\theta \). See [32], Section 8, for an exposition on this classic result. The resulting finite-dimensional Galerkin approximations of the weak eigenvalue problem for \( L_\theta \) can be summarized as follows:

**Definition 32** (Koopman eigenvalue problem, Galerkin approximation). Set \( \tilde{H}_m = \text{span}\{\phi_1, \ldots, \phi_m\} \) and \( \tilde{H}_m^2 = \text{span}\{\phi_1^{(2)}, \ldots, \phi_m^{(2)}\} \), \( m \geq 1 \). Then, find \( \lambda \in \mathbb{C} \) and \( z \in \tilde{H}_m^2 \) such that for all \( f \in \tilde{H}_m \),
\[
A(f, z) = \langle f, z \rangle,
\]
where the sesquilinear form \( A : H \times H^2 \to \mathbb{C} \) is as in Definition 31.

This problem is equivalent to solving a matrix generalized eigenvalue problem
\[
A \tilde{c} = \lambda B \tilde{c},
\]
where \( A \) and \( B \) are \( m \times m \) matrices with elements
\[
A_{ij} = A(\phi_i, \phi_j^{(2)}) = V_{ij}/\eta_j - \theta \Delta_{ij}, \quad V_{ij} = \langle \phi_i, V \phi_j \rangle, \quad \Delta_{ij} = \delta_{ij},
\]
\[
B_{ij} = \langle \phi_i, \phi_j^{(2)} \rangle = \eta_i^{-1} \delta_{ij},
\]
respectively, and \( \tilde{c} = (c_1, \ldots, c_m)^T \) is a column vector in \( \mathbb{C}^m \) containing the expansion coefficients of the solution \( z \) in the \( \{\phi_k^{(2)}\} \) basis of \( \tilde{H}_m^2 \), viz. \( z = \sum_{k=1}^{m} c_k \phi_k^{(2)} \). This concludes the description of our Galerkin approximation of the eigenvalue problem for \( L_\theta \) and therefore for \( V \).
7. Data-driven approximation

In this section, we discuss the numeric procedures used to approximate the integral operators described in Sections 4, 5, and implement the Galerkin method of Section 6 using a finite, time-ordered dataset of observations \((F(x_n))_{n=0}^{N-1}\). In addition, we will prove Theorem 5. Throughout this section, we will assume that Assumptions 1–3 hold. In particular, by Assumption 4, we can assume without loss of generality that the underlying trajectory \((x_n)_{n=0}^{N-1}\) starts at a point \(x_0\) in the compact set \(U\) (for, if \(x_0\) were to lie in \(V \setminus U\), the trajectory would enter \(U\) after finitely many steps, and its portion lying in \(V \setminus U\) would not affect the asymptotic behavior of our schemes as \(N \to \infty\)).

Besides this assumption, the trajectory \((x_n)_{n=0}^{N-1}\) is assumed to be unknown, and note that it need not lie on \(X\).

For the purposes of the analysis that follows, it will be important to distinguish between operators that act on \(L^2\) and \(C^0\) spaces. Specifically, to every kernel \(k : M \times M \to \mathbb{R}\) satisfying Assumption 2, we will assign a bounded operator \(K' : L^2(X, \mu) \to C^0(U)\), acting on \(f \in L^2(X, \mu)\) via the same integral formula as in (3), but with the image \(K'f\) understood as an everywhere-defined, continuous function on \(U\). With this definition, the operator \(K : L^2(X, \mu) \to L^2(X, \mu)\) acting on \(L^2\) equivalence classes can be expressed as \(K'' = \iota \circ K'\), where \(\iota : C^0(U) \to L^2(X, \mu)\) is the canonical \(L^2\) inclusion map on \(C^0(U)\), and we can also define an analog \(K'' : C^0(U) \to C^0(U)\) acting on continuous functions via \(K'' = K' \circ \iota\). It can be verified using the Arzelà-Ascoli theorem that \(K''\) is compact.

**Data-driven Hilbert spaces.** Let \(\mu_N := N^{-1} \sum_{n=0}^{N-1} \delta_{x_n}\) be the sampling probability measure associated with the finite trajectory \((x_n)_{n=0}^{N-1}\). The compact set \(U\) from Assumption 4 always contains the support of \(\mu_N\). Moreover, since \(x_0\) lies in the basin of the physical measure \(\mu\), as \(N \to \infty\), \(\mu_N\) converges weakly to \(\mu\), in the sense that

\[
\lim_{N \to \infty} \int_U f \, d\mu_N = \int_X f \, d\mu, \quad \forall f \in C^0(U).
\] (31)

Our data-driven analog of the space \(L^2(X, \mu)\) will be \(L^2(U, \mu_N)\); the set of equivalence classes of complex-valued functions on \(M\) which are square-summable and have common values at the sampled states \(x_n\). Note that \(L^2(U, \mu_N) \cong \mathbb{C}^N\), and therefore every element \(f \in L^2(U, \mu_N)\) can be represented in the canonical basis of \(\mathbb{C}^N\) as an \(N\)-vector \(\vec{f} = (f(x_0), \ldots, f(x_{N-1}))\). In fact, \(L^2(U, \mu_N)\) is the image of \(C^0(U)\) under the restriction map \(\pi_N : C^0(U) \to L^2(U, \mu_N)\), where \(\pi_N f = (f(x_0), \ldots, f(x_{N-1}))\). Moreover, given any \(f, g \in L^2(U, \mu_N)\), we have \((f, g)_{L^2(U, \mu_N)} = \vec{f} \cdot \vec{g}/N\), where \(\cdot\) denotes the canonical inner product on \(\mathbb{C}^N\).

**Kernel integral operators.** In the data-driven setting, given a kernel \(k : M \times M \to \mathbb{R}\) satisfying Assumption 2, we define a kernel integral operator \(K_N : L^2(U, \mu_N) \to C^0(U)\)
by (cf. (3))

\[ K_N' f(x) = \int_\mathcal{U} k(x, y) f(y) \, d\mu_N(y) = \frac{1}{N} \sum_{n=0}^{N-1} k(x, x_n) f(x_n), \]

and we also set \( K_N : L^2(\mathcal{U}, \mu_N) \mapsto L^2(\mathcal{U}, \mu_N) \) and \( K''_N : C^0(\mathcal{U}) \mapsto C^0(\mathcal{U}) \) with \( K_N = \pi_N \circ K_N' \) and \( K''_N = K_N' \circ \pi_N \). Note that \( K_N \) can be represented by an \( N \times N \) matrix \( K \) with elements \( K_{ij} = k(x_i, x_j) \). In this representation, the function \( g = K_N f, f \in L^2(\mathcal{U}, \mu_N) \), is represented by \( \tilde{g} = \tilde{K} \tilde{f} \).

When \( k = k_Q \), then one can similarly define operators \( K'_{Q,N} : L^2(\mathcal{U}, \mu_N) \mapsto C^0(\mathcal{U}) \), \( K_{Q,N} : L^2(\mathcal{U}, \mu_N) \mapsto L^2(\mathcal{U}, \mu_N) \), and \( K''_{Q,N} : C^0(\mathcal{U}) \mapsto C^0(\mathcal{U}) \). This family of operators has the analogous properties to those stated for \( K_Q \) in Lemma 13; namely, the functions \( \rho_{Q,N} = K'_{Q,N} 1_{\mathcal{U}} \) and \( \sigma_{Q,N} = K''_{Q,N} (1/\rho_{Q,N}) \) are both continuous, positive, and bounded away from zero on \( \mathcal{U} \). Therefore, one can define a kernel \( p_{Q,N} : M \times M \mapsto \mathbb{R} \) by

\[ \rho_{Q,N} = K'_{Q,N} 1_{\mathcal{U}}, \quad \sigma_{Q,N} = K''_{Q,N} (1/\rho_{Q,N}), \]

\[ p_{Q,N}(x, y) = \frac{k_{Q,N}(x, y)}{\sigma_{Q,N}(x) \rho_{Q,N}(y)}. \]

The kernel \( p_{Q,N} \) has the Markov property, i.e., \( \int_\mathcal{U} p_{Q,N}(x, \cdot) \, d\mu_N = 1 \) for every \( x \in M \). Associated to \( p_{Q,N} \) are the Markov operators \( P'_{Q,N} : L^2(\mathcal{U}, \mu_N) \mapsto C^0(\mathcal{U}) \), \( P_{Q,N} : L^2(\mathcal{U}, \mu_N) \mapsto L^2(\mathcal{U}, \mu_N) \) and \( P''_{Q,N} : C^0(\mathcal{U}) \mapsto C^0(\mathcal{U}) \). Moreover, \( P_{Q,N} \) is related to the self-adjoint operator \( \hat{P}_{Q,N} : L^2(\mathcal{U}, \mu_N) \mapsto L^2(\mathcal{U}, \mu_N) \) with kernel \( \hat{p}_{Q,N} : M \times M \mapsto \mathbb{R} \),

\[ \hat{p}_{Q,N}(x, y) = \frac{k_{Q}(x, y)}{\hat{\sigma}_{Q,N}(x) \hat{\sigma}_{Q,N}(y)}, \quad \hat{\sigma}_{Q,N} = \sigma_{Q,N}/\rho_{Q,N}, \]

via a similarity transformation analogous to (17). From the kernel \( \hat{p}_{Q,N} \) one can construct the operators \( \hat{P}_{Q,N} \) and \( \hat{P}'_{Q,N} \) and \( \hat{P}''_{Q,N} \) as above.

**Data-driven basis.** We will use the eigenvectors \( \phi_{k,Q,N} \) of \( \hat{P}_{Q,N} \) as an orthonormal basis of \( L^2(\mathcal{U}, \mu_N) \), and employ the corresponding eigenvalues, \( \lambda_{k,Q,N} \), to define data-driven analogs

\[ \eta_{k,Q,N} = (\lambda_{k,Q,N}^{-1} - 1)/\epsilon \]

of the \( \eta_k \) in (20). This eigenvalue problem is equivalent to a matrix eigenvalue problem for the \( N \times N \) symmetric matrix \( \hat{P} = [\hat{p}_{Q,N}(x_i, x_j)] \) representing \( \hat{P}_{Q,N} \). Details on the numerical solution of this problem can be found in [15, 16]. Since the kernels \( K_Q \) constructed in (8) have exponential decay, \( \hat{P} \) can be well approximated by a sparse matrix, allowing scalability of our techniques to large \( N \). In what follows, we will abbreviate \( \phi_{k,Q,N}, \lambda_{k,Q,N}, \) and \( \eta_{k,Q,N} \) by \( \hat{\phi}_k, \hat{\lambda}_k, \) and \( \hat{\eta}_k \), respectively.

To establish convergence of our schemes in the limit of large data, \( N \to \infty \), we would like to establish a correspondence between the eigenvalues and eigenvectors of \( \hat{P}_{Q,N} \) accessible from data and those of \( \hat{P}_Q \), but because these operators act on the different spaces, a direct comparison of their eigenvectors is not possible. Therefore, we
will first establish a correspondence between the eigenvalues and eigenvectors of \( \hat{P}_{Q,N} \) (\( \hat{P}_Q \)) and those of \( \hat{P}'_{Q,N} \) (\( \hat{P}'_Q \)), and show that \( \hat{P}'_{Q,N} \) spectrally converges to \( \hat{P}'_Q \). The latter problem is meaningful since both \( \hat{P}'_{Q,N} \) and \( \hat{P}'_Q \) act on \( C^0(\mathcal{U}) \).

**Lemma 33.** The following correspondence between the spectra of operators holds:

(i) \( \hat{\lambda}_k \) is a nonzero eigenvalue of \( \hat{P}_{Q,N} \) iff it is a nonzero eigenvalue of \( \hat{P}'_{Q,N} \). Moreover, if \( \hat{\phi}_k \in L^2(\mathcal{U},\mu_N) \) is an eigenfunction of \( \hat{P}_{Q,N} \) corresponding to \( \hat{\lambda}_k \), then \( \hat{\phi}_k = \hat{\lambda}_k^{-1} \hat{P}'_{Q,N} \hat{\phi}_k \in C^0(\mathcal{U}) \) is an eigenfunction of \( \hat{P}'_{Q,N} \) corresponding to the same eigenvalue.

(ii) \( \hat{\lambda}_{k,Q} \) is a nonzero eigenvalue of \( \hat{P}_Q \) iff it is a nonzero eigenvalue of \( \hat{P}'_Q \). Moreover, if \( \hat{\phi}_{k,Q} \in L^2(X,\mu) \) is an eigenfunction of \( \hat{P}_Q \) corresponding to \( \hat{\lambda}_{k,Q} \), then \( \hat{\phi}_{k,Q} = \hat{\lambda}_{k,Q}^{-1} \hat{P}'_Q \hat{\phi}_{k,Q} \in C^0(\mathcal{U}) \) is an eigenfunction of \( \hat{P}'_Q \) corresponding to the same eigenvalue.

The proof of Lemma 33 is left to the reader. Next, we establish spectral convergence of \( \hat{P}'_{Q,N} \) to \( \hat{P}_Q \). For that, we will need the following notion of convergence of operators.

**Compact convergence.** A sequence of operators \( A_n \) on a Banach space \( B \) is said to be compactly convergent to an operator \( A \) if \( A_n \to A \) pointwise, and for every bounded sequence of vectors \( (f_n)_{n \in \mathbb{N}} \), \( f_n \in B \), the sequence \( ((A - A_n)f_n)_{n \in \mathbb{N}} \) has compact closure. The following proposition states that the data-driven operators \( \hat{P}_{Q,N} \) converge compactly, and as result in spectrum; for a proof, see [31], Proposition 11, and [60], Theorem 2.4.1.

**Proposition 34.** Let Assumptions 1 and hold. Given a trajectory \( (x_n)_{n \in \mathbb{N}} \) starting in \( \mathcal{B} \), the corresponding sequence of operators \( \hat{P}'_{Q,N} \) constructed from the observations \( (F(x_n))_{n=0}^{N-1} \) converges compactly as \( N \to \infty \) to \( \hat{P}'_Q \). As a result, the sequence \( \hat{P}_{Q,N} \) converges spectrally, in the sense of Corollary 4, to \( \hat{P}_Q \). In particular, since the nonzero spectrum of a compact operator only consists of isolated eigenvalues, the convergence holds for all nonzero eigenvalues of \( \hat{P}'_{Q,N} \) and the corresponding eigenspaces.

The proof of Theorem 5 then follows by similar arguments:

**Proof of Theorem 5.** The claims of the theorem follow from analogous results to Lemma 33 and Proposition 34 for the operators \( \hat{P}_{Q,N} \), \( \hat{P}'_{Q,N} \), \( \hat{P}'_Q \) and \( \hat{P}_Q \), \( \hat{P}'_Q \), \( \hat{P}'_Q \). □

Together, Lemma 33 and Proposition 34 imply that every eigenpair \( (\hat{\lambda}_{k,Q}, \hat{\phi}_{k,Q}) \) of \( \hat{P}_Q \) can be consistently approximated by a sequence of eigenpairs \( (\hat{\lambda}_k, \hat{\phi}_k) \) of \( \hat{P}_{Q,N} \). Moreover, Corollary 4, as \( Q \to \infty \), \( (\hat{\lambda}_{k,Q}, \hat{\phi}_{k,Q}) \) approximates in turn eigenair \( (\hat{\lambda}_k, \hat{\phi}_k) \) of \( \hat{P}_Q \); that is,

\[
\lim_{Q \to \infty} \hat{\lambda}_k = \hat{\lambda}_k, \quad \lim_{Q \to \infty} \hat{\lambda}_k^{-1} \hat{P}'_{Q,N} \hat{\phi}_k = \hat{\phi}_k, \tag{33}
\]

where the second limit is taken with respect to the \( L^2(X,\mu) \) norm. Since, as can be seen in (30), the Galerkin scheme in Section 6 can be entirely formulated using the \( \hat{\lambda}_k \) and the matrix elements \( \langle \hat{\phi}_i, V \hat{\phi}_j^{(2)} \rangle \) of the generator, (33) indicates in turn that we can construct a consistent data-driven Galerkin scheme if we can consistently compute...
approximate generator matrix elements using the data-driven eigenfunctions $\hat{\phi}_k$. To that end, we will employ finite-difference approximations, as described below.

**Finite-difference approximation.** Recall that the action $Vf$ of the generator on a function $f \in D(V)$ is defined via the limit in (5). This suggests that for data sampled discretely in time (at the sampling interval $\Delta t$), we can approximate $Vf$ by an $r$-th order finite-difference approximation [12, 15, 59]. For example, for $r = 2$, the following is a second-order, central approximation scheme for $V$:

$$V_{\Delta t} f := \frac{1}{2\Delta t} (U_{\Delta t} f - U^{-\Delta t} f) \quad (34)$$

In the finite-sample case, we approximate $V_{\Delta t}$ by a corresponding $r$-th order finite-difference operator $V_{\Delta t,N}$:

$$L^2(U, \mu_N) \mapsto L^2(U, \mu_N).$$

For example, in the case of (34), $V_{\Delta t,N}$ becomes

$$V_{\Delta t,N} f(x_n) = \frac{f(x_{n+1}) - f(x_{n-1})}{2\Delta t}, \quad n \in \{2, \ldots, N-2\},$$

and $V_{\Delta t,N} f(x_0) = V_{\Delta t,N} f(x_{N-1}) = 0$. To ensure that the approximations $V_{\Delta t,N} f$ converge to the true function $V f$ for a class of functions of sufficient regularity, the following smoothness conditions are sufficient:

**Assumption 5.** $U$ is a $C^{r+1}$ compact manifold, and $\Phi^t|_U$ is generated by a $C^r$ vector field $\vec{V}$. Moreover, $F|_U \in C^{r+1}(U; \mathbb{R}^d)$, and the kernel shape function $h : \mathbb{R} \rightarrow \mathbb{R}$ is $C^{r+1}$. $V_{\Delta t}$ is an $r$-th order finite difference scheme.

**Remark 35.** If $r = 1$, then it is sufficient for $\vec{V}$ to be Lipschitz.

**Proposition 36.** Let Assumptions 1, 2, and 5 hold. Then for every $i, j \in \mathbb{N}$:

(i) The eigenfunctions $\hat{\phi}_k$ and $\phi_{k,Q}$ from Lemma 33 lie in $C^r(U)$. Moreover, as $\Delta t \rightarrow 0$,

$$V_{\Delta t} \phi_{i,Q} = V\phi_{i,Q} + \|\phi_{i,Q}\|_{C^{r+1}(U)} O(\Delta t)^r,$$

where the estimate holds uniformly on $U$.

(ii) $\lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \langle \hat{\phi}_i, V_{\Delta t,N} \phi_j \rangle_{L^2(U, \mu_N)} = \langle \phi_i, V \phi_j \rangle$.

**Proof.** To prove Claim (i), note that under Assumption 5, for a finite number of delays $Q$, by (7), $k_Q$ is a $C^{r+1}$-smooth kernel. Hence, according to [48], its eigenfunctions have representatives in $C^{r+1}(U)$. Since the vector field is $C^r$, the trajectories are $C^{r+1}$-smooth and therefore, $V\phi_{i,Q}$, which is the time-derivative along the orbit, has an $r$-th order Taylor expansion. The $r$-th order finite difference scheme gives the $\|\phi_{i,Q}\|_{C^{r+1}(U)} O(\Delta t)^r$ error.

Claim (ii) is a consequence of Claim (i) in conjunction with the weak convergence of measures in (31).
Data-driven Galerkin method. Using the \( \hat{\eta}_k \) from (32), we define the data-driven normalized basis vectors \( \hat{\phi}_k^{(p)} = \phi_k/\hat{\eta}_k^{p/2} \) (cf. the \( \hat{\phi}_k \) from Step 1 in Section 6) and the associated Galerkin approximation spaces \( H_{Q,N}^p = \text{span}\{\hat{\phi}_k^{(p)}\}_{k=1}^m \). We also define the positive semidefinite, self-adjoint operator \( \Delta_{Q,N} : L^2(U, \mu_N) \mapsto L^2(U, \mu_N) \), where

\[
\Delta_{Q,N} f = \frac{1}{\epsilon} \sum_{k=0}^{N-1} \hat{\eta}_k c_k \hat{\phi}_k, \quad \hat{f} = \sum_{k=0}^{N-1} c_k \hat{\phi}_k.
\]

This operator is a data-driven analog of \( \Delta \) in (21). With these definitions and the finite-difference approximation of \( V \) described above, we pose the following data-driven analog of the Galerkin approximation in Definition 32:

**Definition 37** (Koopman eigenvalue problem, data-driven form). Find \( \lambda \in \mathbb{C} \) and \( z \in H_{Q,N,m}^2 \) such that for all \( f \in H_{Q,N,m}^0 \),

\[
A_{\Delta t,Q,N}(f, z) = \langle f, z \rangle_{L^2(U, \mu_N)},
\]

where \( A_{\Delta t,Q,N} : L^2(U, \mu_N) \times L^2(U, \mu_N) \mapsto \mathbb{C} \) is the sesquilinear form defined as

\[
A_{\Delta t,Q,N}(f, z) = \langle f, V_{\Delta t,N} z \rangle_{L^2(U, \mu_N)} - \theta \langle f, \Delta_{Q,N} z \rangle_{L^2(U, \mu_N)}.
\]

Numerically, this is equivalent to solving a matrix generalized eigenvalue problem analogous to that in (29), viz.

\[
A\tilde{c} = \lambda B\tilde{c},
\]

where \( A \) and \( B \) are \( m \times m \) matrices with elements

\[
A_{ij} = A_{\Delta t,Q,N}(\hat{\phi}_i, \hat{\phi}_j^{(2)}) = \frac{V_{ij}}{\hat{\eta}_j} - \theta \delta_{ij},
\]

\[
V_{ij} = \langle \hat{\phi}_i, V_{\Delta t,N} \hat{\phi}_j \rangle_{L^2(U, \mu_N)}, \quad \Delta_{ij} = \delta_{ij},
\]

\[
B_{ij} = \langle \hat{\phi}_i, \hat{\phi}_j^{(2)} \rangle_{L^2(U, \mu_N)} = \hat{\eta}_i^{-1} \delta_{ij},
\]

respectively, and \( \tilde{c} = (c_1, \ldots, c_m)^T \) is a column vector in \( \mathbb{C}^m \) containing the expansion coefficients of the solution \( z = \sum_{k=1}^m c_k \hat{\phi}_k^{(2)} \) in the \( \{\hat{\phi}_k^{(2)}\} \) basis of \( H_{Q,N,m}^2 \). For any \( m \) and up to similarity transformations, the matrices \( A \) and \( B \) converge in the limits \( Q \to \infty \), after \( \Delta t \to 0 \), after \( N \to \infty \) (in that order) to the corresponding matrices in the variational eigenvalue problem in (29). We therefore conclude that the data-driven Galerkin method in Definition 37 is consistent (as \( \Delta t \to 0 \) and \( Q, N \to \infty \)) with the Galerkin method in Definition 32, which is in turn consistent (as \( m \to \infty \)) with the weak eigenvalue problem for the regularized generator \( L_\theta \) in Definition 31.

8. Results and discussion

In this section, we apply the methods described in Sections 4–7 to two ergodic dynamical systems with mixed spectrum, constructed as products of either a mixing flow on the 3-torus, or the L63 system, with circle rotations. Our objectives are to demonstrate that (i) the results of Theorem 1 and Corollaries 4 and 2 hold, that is, the eigenfunctions \( \hat{\phi}_k \) of \( P_{Q,N} \) from (15) are eigenfunctions of \( U^t \); (ii) the eigenvalues obtained using the Galerkin scheme in Definition 37 are consistent with those expected theoretically.
8.1. Two systems with mixed spectrum

The first system studied below is based on a strongly mixing flow on the 3-torus introduced by Fayad [61]. The flow, denoted by \( \Phi^t_{3T} \), is given by the solution of the ordinary differential equation (ODE) \( d(x, y, z)/dt = \vec{V}(x, y, z) \), where \((x, y, z) \in T^3\), and \( \vec{V} \) is the smooth vector field

\[
\vec{V}(x, y, z) = \vec{\nu}/\varphi(x, y, z), \quad \varphi(x, y, z) = 1 + \sum_{k=1}^{\infty} e^{-k} \frac{\Re}{k} \sum_{|l| \leq k} e^{ik(x+y)+ilz},
\]

parameterized by the constant frequency vector \( \vec{\nu} \). Hereafter, we set \( \vec{\nu} = (\sqrt{2}, \sqrt{10}, 1)^T \). Note that the orbits under \( \Phi^t_{3T} \) are the same as the orbits under the ergodic, non-mixing quasiperiodic flow with constant vector \( \vec{\nu} \), and and the (unique) ergodic invariant measure under \( \Phi^t_{3T} \) has density \( \varphi/\int_M \varphi \, d\text{Leb} \) relative to Lebesgue measure. Such flows are also called reparameterized flows as \( \vec{\nu} \) is scaled by the function \( \varphi \) at each point \((x, y, z) \in T^3\).

This system is strongly mixing with respect to its invariant measure, i.e., its generator has purely continuous spectrum [61]. To construct an associated mixed-spectrum system, we take the product \( \Phi^t_{3T} \times R^t_\omega \) with a periodic flow \( R^t_\omega \) on \( S^1 \), defined as

\[
dR^t_\omega(\alpha)/dt = \omega, \quad \omega = 1.
\]

Thus, the state space of the product system is \( M = T^3 \times S^1 = T^4 \). Note that in this example the attracting set \( X \) is smooth and coincides with the state space, \( M = X \); in particular, all states sampled experimentally lie exactly on \( X \). Moreover, the Koopman generator \( V : D(V) \to L^2(X, \mu) \) is a skew-adjoint extension of the differential operator \( \vec{V} \oplus \omega : C^\infty(X) \to L^2(X, \mu) \). Since \( R^t_\omega \) has a pure point spectrum consisting of integer multiples of \( i\omega \) and \( \Phi^t_{3T} \) has no eigenvalues, the discrete spectrum of the product system is \( \{ik\omega, k \in \mathbb{Z}\} \).

The second system that we study is based on the L63 system [39]. This system is known to have a chaotic attractor \( X_{\text{Lor}} \subset \mathbb{R}^3 \) with fractal dimension \( 2.0627160 \) [43], supporting a physical invariant measure [62] with a corresponding purely continuous spectrum of the generator [63]. The flow, denoted by \( \Phi^t_{\text{Lor}} \), is generated by a smooth vector field \( \vec{V} \in C^\infty(\mathbb{R}^3, \mathbb{R}^3) \), whose components at \((x, y, z) \in \mathbb{R}^3\) are

\[
V^{(x)} = \sigma(y-x), \quad V^{(y)} = x(\rho-z) - y, \quad V^{(z)} = xy - \beta z.
\]

Throughout, we use the standard parameter values \( \beta = 8/3, \rho = 28, \sigma = 10 \). As in the torus case, we form the product \( \Phi^t_{\text{Lor}} \times R^t_\omega \) with the rotation \( R^t_\omega \) in (36), leading to a mixed spectrum system with the same discrete spectrum \( \{ik\omega, k \in \mathbb{Z}\} \). Note that unlike the torus-based system, the attracting set \( X = X_{\text{Lor}} \times S^1 \) is a strict subset of the state space \( M = \mathbb{R}^3 \times S^1 \).

For each product system, we define a continuous map \( F : M \mapsto \mathbb{R}^3 \) coupling the degrees of freedom of the continuous-spectrum subsystem with the rotation. In the case
8.2. Experimental results

We generated numerical trajectories \( x_0, x_1, \ldots, x_{N-1} \) of the torus- and L63-based systems described above starting in each case from an arbitrary initial condition \( y \in M \). In the torus experiments, the system is always on the attractor, so the starting state \( x_0 \) in the training data was set to \( y \). In the L63 experiments, we let the system relax towards the attractor, and set \( x_0 \) to a state sampled after a long spinup time (4000 time units); that is, we formally assume that \( y \) (and therefore \( x_0 \)) lie in the basin \( \mathcal{B}_\mu \) of the physical measure associated with \( X \). In both cases, the number of samples was \( N = 50,000 \), the integration time-step was 0.01, and the number of delays was \( Q = 2000 \).

We used the \texttt{ode45} solver of Matlab to compute the trajectories and generated time series \( F(x_0), F(x_1), \ldots, F(x_{N-1}) \) by applying the observation maps in (38) and (39) to
Delay-coordinate maps and the spectra of Koopman operators

Figure 3: Eigenvalues $\hat{\lambda}_j$ of the integral operator $P_{Q,N}$ for representative values of the delay parameter $Q$ for the torus system in (38) (left) and the L63-based system in (39) (right). The blue lines correspond to either no delays or one delay, while the red curves correspond to 2000 delays. When $Q \leq 1$, the eigenvalues are seen clustering around 1. The eigenvalues cannot exceed 1 as $P_{Q,N}$ is a Markov operator. At $Q = 2000$, the eigenvalues decay more rapidly towards zero and, at least up to eigenvalue 15, have multiplicity 2 as expected from Proposition 24.

the respective states $x_n$. Portions of the observable time series from each system are displayed in Fig. 2. Note that the $x_n$ were not presented to our kernel algorithm.

We computed data-driven eigenfunctions $\hat{\phi}_k$ by solving the eigenvalue problem for the operator $P_{Q,N}$ as described in Section 7, using Matlab’s eigs iterative solver. The kernel bandwidth parameter $\epsilon$ was selected using the tuning procedure described in [30, 36, 64], which yielded $\epsilon \approx 3.6$ and $\approx 2.053$ for the torus and L63 systems, respectively. Representative eigenfunctions $\hat{\phi}_k$, plotted as time series $n \mapsto \hat{\phi}_k(x_n)$, and the corresponding eigenvalues are displayed in Figs. 1 and 3, respectively. We now describe these results in more detail.

According to Theorem 1 and Proposition 24, at large numbers of delays (here, $Q = 2000$), the eigenfunctions $\hat{\phi}_k$ of $P_{Q,N}$ should form doubly degenerate pairs, and each pair should exhibit a single frequency associated with an eigenvalue of $V$ (more precisely, $\phi_k + i\phi_{k+1}$ with $k \in \{1,3,\ldots\}$ should be an eigenfunction of $V$). Both systems studied here have exactly one rationally independent eigenvalue $i\omega = i$, so the eigenfunctions of $P_{Q,N}$ are expected to evolve at frequencies $k\omega$, $k \in \mathbb{N}$. This is evidently the case in the time series plots in Fig. 1. Also, each of the $\phi_k$ has multiplicity 2 (note that only one eigenfunction from each eigenspace is shown in Fig. 1). The left-hand panels of Fig. 1 show a matrix representation of the generator $V$ (approximated via the finite difference scheme in (34)) in the 51-dimensional data-driven subspace spanned by $\hat{\phi}_0, \ldots, \hat{\phi}_{50}$. Note that the matrix is skew-symmetric since $V$ is a skew-symmetric operator, and, as expected, consists of $2 \times 2$ diagonal blocks associated with the eigenspaces of $V$ spanned by $(\phi_1, \phi_2), (\phi_3, \phi_4), \ldots$.

Figure 4 shows the eigenvalues $\gamma_k$ of the regularized generator $L_\theta$ obtained from this basis using the Galerkin scheme in definition 37 with the diffusion regularization and spectral order parameters $\theta = 10^{-4}$ and $m = 50$, respectively. Each plot in Fig. 4 shows the first 20 eigenvalues corresponding to eigenfunctions of increasing Dirichlet
Figure 4: Galerkin approximations of the eigenvalues $\gamma_k$ of the regularized generator $L_\theta$ for the torus-based system (38) (left) and the L63-based system (39) (right). The numerical eigenvalues were obtained through the data-driven variational eigenvalue problem in definition 37 with a spectral order parameter $m = 50$. Each plot shows the first 20 eigenvalues corresponding to eigenfunctions of increasing Dirichlet energy, with the first 14 plotted in blue and the remaining 5 in red. Dashes on the imaginary axes indicate the imaginary parts of the eigenvalues. The intervals between the blue-colored dashes are to a good approximation equal to 1, in agreement with the exact Koopman eigenvalues of these systems.

energy $E(z_k)$ of the corresponding eigenfunction $z_k$ (recall that $\text{Re}(\gamma_k) = -\theta E(z_k)$). According to Section 6, the imaginary parts of the $\gamma_k$ should approximate the Koopman eigenfrequencies $j(k)\omega$, where $j$ is an integer-valued function giving the frequency of the Koopman eigenfunction with the $k$-th smallest Dirichlet energy. In Fig. 4, the $\text{Im} \gamma_k$ are indeed equal to integer multiples of $\omega = 1$ to a good approximation for the first $m$ eigenvalues (ordered in order of increasing Dirichlet energy). For indices $k$ close to $m$, the accuracy of the eigenvalues begins to deteriorate. This is due to the facts that (i) even with a “perfect” basis $\{\hat{\phi}_k\}$, eigenfunctions of higher Dirichlet energy (and stronger oscillatory behavior) require increasingly higher-order Galerkin approximation spaces; (ii) at finite sample numbers $N$, the quality of the data-driven elements $\hat{\phi}_k$ degrades at large $k$.

8.3. Discussion

The examples presented in Sections 8.1 and 8.2 are Cartesian products of weak mixing and quasiperiodic flows, with their phase variables combined through some observation map. We begin with some observations about our kernel method applied to Cartesian products.

*Cartesian products.* Let $(X, \Phi_X^t, \mu_X)$ and $(Y, \Phi_Y^t, \mu_Y)$ be two ergodic flows on compact metric spaces with purely continuous and pure-point spectra, respectively. We are interested in the measure-preserving mixed spectrum dynamical system $(X \times Y, \Phi_X^t \times \Phi_Y^t, \mu_X \times \mu_Y)$. It is well known that the space $L^2(X \times Y, \mu_X \times \mu_Y)$ is densely spanned by products of the form $\{f \otimes g : f \in L^2(X, \mu_X), g \in L^2(Y, \mu_Y)\}$. Recall that the observation map $F$ is the basis of our construction of all our data-driven operators. Corollary 38
below is a direct consequence of Proposition 17, and gives an “observability” condition that must be fulfilled by the observation map $F$ in order for the methods presented here to yield non-trivial results.

**Corollary 38.** Let $(X, \Phi^t_X, \mu_X)$ and $(Y, \Phi^t_Y, \mu_Y)$ be as described above, and $F \in L^2(X \times Y, \mu_X \times \mu_Y)$ be the sum $F = \sum_{n=1}^{\infty} f_n \otimes g_n$. Then, $F_D = \sum_{n=1}^{\infty} E(f_n)g_n$, where $E(f_n) = \int_X f_n \, d\mu_X$. Hence, a necessary and sufficient condition that $P$ is not trivial is that $E(f_n) \neq 0$ for at least one $n \in \mathbb{N}$.

**Kernels with a low number of delays.** An implicit assumption in the approximation of the operator $P$ in (15) by the operator $P_Q$ in (15) with finitely many delays $Q$, is that $Q$ is large-enough for the asymptotic analysis of Lemma 25 to hold. When $Q$ is small, $d_Q$ is closer to a proper metric and therefore, the entries $K_{ij} = \exp(-d_Q(x_i, x_j)^2/\epsilon)$ of the kernel matrix $K$ decay rapidly away from the diagonal $i = j$. Then $K_{ij}$ is close to a diagonal matrix, and $P_{ij}$ is close to the identity matrix. On the other hand, for $Q$ large, $d_Q$ becomes a pseudo-metric and $P_{ij}$ is not necessarily close to a diagonal matrix. Figure 3 shows how the Koopman eigenvalues computed for the two examples from (38) and (39) cluster near 1 for $Q = 1$ and decay more rapidly for $Q = 2000$.

**Systems with purely continuous spectra.** An important assumption of our kernel-based method is that the dynamics has Koopman eigenvalues, i.e., $\mathcal{D}$ contains non-constant functions. This underlies the ability of our regularized operator $L_\theta$ in (21) to be a suitable substitute of $V$ (Theorem 29). In fact, by Theorem 2, in the limit of infinitely many delays $Q \to \infty$, if $\mathcal{D}$ only contains constant functions, then the kernels $k_Q, p_Q$ converge to 0 (in the $L^2$ sense). However, when using finitely many delays, $k_Q \neq 0$, and correspondingly $p$ obtained by normalization of $k_\infty$ is not close to 0. It is not currently understood how the operator $P_Q$ should behave for purely continuous spectrum systems (i.e., $\mathcal{D} = \text{span}\{1_X\}$) and $Q < \infty$.

Numerical results shown in Fig. 5 indicate that the finite-rank, data-driven operator $P_{Q,N}$ for the L63 system still has nonzero eigenvalues strictly less than 1, but these eigenvalues are clustered around a small value ($\lambda_k \approx 0.1$). This behavior is in agreement with Theorem 2, according to which all the eigenvalues of $P_Q$ other than 1 should converge to zero as $Q \to \infty$. Note that the matrix representation of $V$ (also shown in Fig. 5) is still skew-symmetric since $V$ is a skew-symmetric operator. Intriguingly, the matrix has a $2 \times 2$ block-diagonal form, despite $V$ having no eigenfunctions. This form of the generator matrix has some aspects in common with the recent results of Brunton et al. [13], who obtained a bi-diagonal matrix representation of the L63 generator in a data-driven basis from Hankel matrix analysis. In Fig. 5, the lack of Koopman eigenfunctions is evident from the time-series plots of the numerical eigenfunctions $\hat{\phi}_k$, which are are clearly non-periodic. Moreover, a phase space plot of $\hat{\phi}_2$ illustrates that it is a highly rough function on the Lorenz attractor. One of the goals of our future work is to investigate the behavior of the techniques presented above away from the asymptotic limit $Q \to \infty$ in the presence of a purely continuous spectrum.
Figure 5: Eigenvalues $\hat{\lambda}_j$ and eigenfunctions $\hat{\phi}_j$ of $P_{N,Q}$, and a matrix representation of the generator of the L63 system in (37) obtained with $Q = 4000$. This is a system with a purely continuous spectrum system, and according to Theorem 2, as $Q \to \infty$ all $\hat{\lambda}_j \neq 1$ converge to 0. This behavior can be seen in the bottom-right panel, where the $\hat{\lambda}_j$ not equal to 1 are seen clustered around a small value $\approx 0.1$. Moreover, the time series of the $\hat{\phi}_j$, shown in the bottom-left panel, are manifestly non-periodic since they fail to converge to Koopman eigenfunctions. As illustrated by the phase space plot of $\hat{\phi}_2$ in the top-right panel, the leading eigenfunctions have a highly rough geometrical structure on the Lorentz attractor. The top left panel shows a matrix representation of the generator $V$ with respect to the $\{\hat{\phi}_j\}$ basis. Remarkably, this matrix is very nearly bi-diagonal, yet we do not have a theoretical result justifying this behavior.

acknowledgements Dimitrios Giannakis received support from ONR YIP grant N00014-16-1-2649, NSF grant DMS-1521775, and DARPA grant HR0011-16-C-0116. Suddhasattwa Das is supported as a postdoctoral research fellow from the first grant. The authors are grateful to L S Young for her suggestions.

References

[1] C. Wang, C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement. El Niño and Southern Oscillation (ENSO): A review. In P. W. Glynn, D. P. Manzello, and I. C. Enoch, editors, Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment, volume 8 of Coral Reefs of the World, pages 85–106. Springer Netherlands, Dordrecht, 2017.

[2] M. Budisić, R. Mohr, and I. Mezić. Applied Koopmanism. Chaos, 22:047510, 2012.
Delay-coordinate maps and the spectra of Koopman operators

[3] T. Eisner, B. Farkas, M. Haase, and R. Nagel. Operator Theoretic Aspects of Ergodic Theory, volume 272 of Graduate Texts in Mathematics. Springer, 2015.
[4] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. Proc. Natl. Acad. Sci., 17(5):315–318, 1931.
[5] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. SIAM J. Numer. Anal., 36:491, 1999.
[6] I. Mezić and A. Banaszuk. Comparison of systems with complex behavior. Phys. D., 197:101–133, 2004.
[7] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn., 41:309–325, 2005.
[8] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. J. Fluid Mech., 641:115–127, 2009.
[9] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech., 656:5–28, 2010.
[10] J. H. Tu, C. W. Rowley, C. M. Livershenbg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: Theory and applications. J. Comput. Dyn., 1(2):391–421, 2014.
[11] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. J. Nonlinear Sci., 2015.
[12] D. Giannakis, J. Slawinska, and Z. Zhao. Spatiotemporal feature extraction with data-driven Koopman operators. J. Mach. Learn. Res. Proceedings, 44:103–115, 2015.
[13] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz. Chaos as an intermittently forced linear system. Nat. Commun., 8(19), 2017.
[14] H. Arbabi and I. Mezić. Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator, 2016.
[15] D. Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. Appl. Comput. Harmon. Anal., 2017. In press.
[16] D. Giannakis and S. Das. Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis, 2017. Preprint https://arxiv.org/pdf/1706.06450.pdf.
[17] M. Korda and I. Mezić. On convergence of extended dynamic mode decomposition to the koopman operator, 2017.
[18] M. Korda, M. Putinar, and I. Mezić. Data-driven spectral analysis of the Koopman operator, 2017.
[19] M. Dellnitz and G. Froyland. On the isolated spectrum of the Perron-Frobenius operator. Nonlinearity, pages 1171–1188, 2000.
[20] G. Froyland, C. González-Tokman, and A. Quas. Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. J. Comput. Dyn., 1(2):249–278, 2014.
[21] G. Froyland, G. A. Gottwald, and A. Hammerlindl. A computational method to extract macroscopic variables and their dynamics in multiscale systems. SIAM J. Appl. Dyn. Sys., 13(4):1816–1846, 2014.
[22] G. Froland and K. Padberg. Almost-invariant sets and invariant manifolds – Connecting probabilistic and geometric descriptions of coherent structures in flows. Phys. D, 238:1507–1523, 2009.
[23] S. Das and D. Giannakis. Koopman spectra in reproducing kernel Hilbert spaces, 2018.
[24] N. Aubry, R. Guyonnet, and R. Lima. Spatiotemporal analysis of complex signals: Theory and applications. J. Stat. Phys., 64:683–739, 1991.
[25] P. Holmes, J. L. Lumley, and G. Berkooz. Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press, Cambridge, 1996.
[26] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput., 15:1373–1396, 2003.
[27] R Coifman and S Lafon. Diffusion maps. Appl. Comput. Harmon. Anal., 21:530, 2006.
[28] T. Berry and T. Sauer. Local kernels and the geometric structure of data. *Appl. Comput. Harmon. Anal.*, 40:439–469, 2016.
[29] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, 65(3–4):579–616, 1991.
[30] T. Berry and J. Harlim. Variable bandwidth diffusion kernels. *J. Appl. Comput. Harmon. Anal.*, 2015.
[31] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Ann. Stat.*, 26(2):555–586, 2008.
[32] I. Babuska and J. Osborn. *Eigenvalue Problems*, volume 2. North Holland, 1991.
[33] M. Brin and G. Stuck. *Ergodic theory*. Cambridge University Press, 2002.
[34] M. G. Nadkarni. *The spectral theorem for unitary operators*. Springer Science and Business Media, 1998.
[35] P. J. Schmid and J. L. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In *Bull. Amer. Phys. Soc.*, 61st APS meeting, page 208, San Antonio, 2008.
[36] T. Berry, D. Giannakis, and J. Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E.*, 91:032915, 2015.
[37] B Scholkopf, A Smola, and K Mu. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, 10:12991319, 1998.
[38] L. S. Young. What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.
[39] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.
[40] K. Lu, Q. Wang, and L.-S. Young. Strange attractors for periodically forced parabolic equations. *Mem. Amer. Math. Soc.*, 23(1154):1–85, 1979.
[41] Z. Lian, P. Liu, and K. Lu. SRB measures for a class of partially hyperbolic attractors in Hilbert spaces. *J. Differential Equ.*, 261:1532–1603, 2016.
[42] M. H. Stone. On one-parameter unitary groups in Hilbert space. *Ann. Math.*, 33:643648, 1932.
[43] M. J. McGuinness. The fractal dimension of the Lorenz attractor. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 262:413–458, 1968.
[44] D. S. Broomhead and D. Lowe. New examples in smooth ergodic theory. Ergodic diffeomorphisms. *Trans. Mosc. Math. Soc.*, 23:1–35, 1970.
[45] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.*, 41:309–325, 2005.
[46] P Halmos. *Volume 142*. Amer. Math. Soc., 1956.
[47] J. C. Ferreira and V. A. Menegatto. Eigenvalue decay rates for positive integral operators. *Ann. Mat. Pura Appl.*, 192(6):1–17, 2013.
[48] J. C. Ferreira and V. A. Menegatto. Eigenvalues of integral operators defined by smooth positive definite kernels. *Integral Equations Operator Theory*, 64(1):61–81, 2009.
[49] U. Krengel. *Ergodic theorems*, volume 6. Walter de Gruyter, 1985.
[50] D. Giannakis and A. J. Majda. Time series reconstruction via machine learning: Revealing decadal variability and intermittency in the North Pacific sector of a coupled climate model. In *Conference on Intelligent Data Understanding 2011*, Mountain View, California, 2011.
[51] D. Giannakis and A. J. Majda. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.*, 109(7):2222–2227, 2012.
[52] T. Berry, R. Cressman, Z. Gregurić-Ferenček, and T. Sauer. Time-scale separation from diffusion-mapped delay coordinates. *SIAM J. Appl. Dyn. Sys.*, 12:618–649, 2013.
[53] D. Sławinska, J. anD. Giannakis. Indo-Pacific variability on seasonal to multidecadal time scales. Part I: Intrinsic SST modes in models and observations. *J. Climate*, 30:5265–5294, 2017.
[54] N. H. Packard et al. Geometry from a time series. *Phys. Rev. Lett.*, 45:712–716, 1980.
[55] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Phys. D*, 20(2–3):217–236, 1986.
[56] R. Vautard and M. Ghil. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Phys. D*, 35:395–424, 1989.
[57] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In Advances in Neural Information Processing Systems, volume 17, pages 1601–1608, 2004.

[58] T. Berry and T. Sauer. Consistent manifold representation for topological data analysis, 2016.

[59] D. Giannakis. Dynamics-adapted cone kernels. SIAM J. Appl. Dyn. Sys., 14(2):556–608, 2015.

[60] A. van der Vaart and J. Wellner. Empirical processes. Springer-Verlag New York, 1996.

[61] B. Fayad. Analytic mixing reparametrizations of irrational flows. Ergod. Th. & Dynam. Sys., 22:437–468, 2002.

[62] W. Tucker. The Lorenz attractor exists. C. R. Acad. Sci. Paris, Ser. I, 328:1197–1202, 1999.

[63] S. Luzzatto, I. Melbourne, and F. Paccaut. The Lorenz attractor is mixing. Comm. Math. Phys., 260(2):393–401, 2005.

[64] R. Coifman, Y. Shkolnisky, F. Sigworth, and A. Singer. Graph Laplacian tomography from unknown random projections. IEEE Trans. Image Process., 17(10):1891–1899, 2008.