Delay-Coordinate Maps and the Spectra of Koopman Operators

Suddhasattwa Das[1] · **Dimitrios Giannakis[1]**

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 rotations on tori. Here, we describe a method through 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 point spectrum subspace of the Koopman operator. This enables efficient approximation of Koopman eigenfunctions in systems with pure point or mixed spectra. We illustrate our results with applications to product dynamical systems with mixed spectra.

Keywords Koopman operators · Delay-coordinate maps · Point spectrum · Koopman eigenfunctions · Kernel methods · Galerkin approximation

Mathematics Subject Classification 37A10 · 37E99 · 37G30

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

✉ Suddhasattwa Das
  dass@cims.nyu.edu

  Dimitrios Giannakis
  dimitris@cims.nyu.edu

[1] Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

Niño Southern Oscillation (the latter, evolving on interannual timescales) both have strong associated signals in sea surface temperature [63]. In such 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 [15,23], 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 [37], 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. However, 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 [22,47,48], 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 or the Perron–Frobenius (transfer) operators, 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 [21,22,22,27], harmonic averaging [20,47,48], iterative methods [51,53], dictionary/basis representations [30,34,38, 59,64], delay-coordinate embeddings [3,14,30,34], and spectral-moment estimation [39].

Compared to observables identified by spectral analysis of kernel integral operators that do not depend on the dynamics (e.g., covariance [4,36] or heat operators [6,10,16], 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 measure-preserving, 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 from 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 [30], or systems observed through measurement functions lying in finite-dimensional invariant subspaces [3].

The main contribution of our work 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 is based on a

combination of ideas from delay-coordinate maps of dynamical systems [52], kernel integral operators for machine learning [6,8,10,16,62], and Galerkin approximation techniques for variational eigenvalue problems [5]. 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. It will be shown 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. We will demonstrate our results with applications to low-dimensional, mixed-spectrum systems, with the structure of a product of a circle rotation and a mixing system.

2 Assumptions and Statement of Main Results

A common underlying assumption in the statistical modeling of dynamical systems is ergodicity. This assumption encapsulates the working principle that the global statistical properties (with respect to an invariant measure μ) 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. Moreover, ergodicity implies that $L^2(\mu)$ inner products between observables can be approximated by 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 a metric space, equipped with its Borel σ-algebra. $\Phi^t : M \to M$, $t \in \mathbb{R}$, is a continuous flow on M with an ergodic, invariant, Borel probability measure μ with a compact support X not equal to a single point. $F : M \to \mathbb{R}^d$ 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 Δt is a fixed sampling interval such that the map $\Phi^{\Delta t}$ is ergodic for the invariant measure μ.

The Koopman operator Central to all our following discussions will be the concept of the Koopman operator. Koopman operators [15,23,49] 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 defined as

$$(U^t f) : x \mapsto f\left(\Phi^t(x)\right), \quad \text{for } \mu\text{-a.e. } x \in X.$$

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 [47]. The problems of mode decomposition and non-parametric prediction can both be stated in terms of the Koopman operator [30]. 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 $\omega \in \mathbb{R}$:

$$U^t z = \exp(i\omega t)z. \tag{1}$$

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 $\omega_1, \ldots, \omega_l$ [30,34,47]. In this representation, the z_j can be thought of as "coordinates" corresponding to distinct periodic processes operating at the timescales $2\pi/\omega_j$. 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 [47]. Data-driven techniques for computing Koopman eigenvalues, eigenfunctions, and modes that have been explored in the past include methods based on generalized Laplace analysis [47,48], dynamic mode decomposition (DMD) [51,53,54,59], extended DMD (EDMD) [38,64], Hankel matrix analysis [3,14,59], spectral moment estimation [39], and data-driven Galerkin methods [30,31,34]. The latter approach, as well as the related work in [12], additionally address the problem of nonparametric prediction of observables and probability densities.

Let $\mathcal{D}$ be the closed subspace of $L^2(X, \mu)$ spanned by the eigenfunctions of U^t, and $\mathcal{D}^\perp$ its orthogonal complement. As is well known [35], and will be discussed in more detail in Sect. 3, the subspaces $\mathcal{D}$ and $\mathcal{D}^\perp$ represent the quasiperiodic and weak-mixing (chaotic) components of the dynamics, respectively. Moreover, they are both invariant under U^t for every time $t \in \mathbb{R}$, thus inducing an invariant splitting [47]

$$L^2(X, \mu) = \mathcal{D} \oplus \mathcal{D}^\perp. \tag{2}$$

Systems for which $\mathcal{D}$ contains non-constant functions and $\mathcal{D}^\perp$ is non-zero are called mixed-spectrum systems.

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 [6,16,55]. 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 $K : L^2(X, \mu) \to L^2(X, \mu)$ such that

$$Kf(x) := \int_X k(x, y) f(y) \, d\mu(y). \tag{3}$$

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

Assumption 2 The kernel $k : M \times M \to \mathbb{R}$ is (i) symmetric and continuous; (ii) strictly positive-valued.

Overview of approach We will address the eigenvalue problem for U^t by solving an eigenvalue problem for a kernel integral operator P_Q indexed by $Q \in \mathbb{N}$, which is accessible from data, and in the limit of $Q \to \infty$ 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. The operators P_Q have Markov kernels $p_Q : M \times M \to \mathbb{R}$

(i.e., $p_Q \geq 0$ and $\int_X p_Q(x, \cdot) \, d\mu = 1$, for μ-a.e. $x \in M$), whose construction begins from a family of distance-like functions $d_Q : M \times M \rightarrow \mathbb{R}$, defined by

$$d_Q^2(x, y) = \frac{1}{Q} \sum_{q=0}^{Q-1} \left\| F(\Phi^{q \, \Delta t}(x)) - F(\Phi^{q \, \Delta t}(y)) \right\|^2. \tag{4}$$

Here, Q is a positive integer parameter, and $\|\cdot\|$ the canonical 2-norm on $\mathbb{R}^d$. Intuitively, $d_Q(x, y)$ assigns a distance-like quantity between points x and y equal to the root-mean square distance between Q consecutive "snapshots" of the observable F, measured along dynamical trajectories starting from x and y. In other words, d_Q corresponds to a distance between data in delay-coordinate space with Q delays. Several of our results will depend on the asymptotic behavior of d_Q as $Q \rightarrow \infty$, which we will study in detail.

Composing d_Q with a continuous shape function $h : \mathbb{R} \rightarrow \mathbb{R}$, leads to a kernel $k_Q : M \times M \rightarrow \mathbb{R}$, $k_Q = h \circ d_Q$, assigning a pairwise measure of similarity between points in M. In this paper, we will nominally work with Gaussian shape functions, $h(s) = e^{-s^2/\epsilon}$, parameterized by a bandwidth parameter $\epsilon > 0$, so that

$$k_Q(x, y) = e^{-d_Q^2(x,y)/\epsilon}. \tag{5}$$

Such kernels satisfy Assumption 2. They are popular in manifold learning applications [6, 10,16] due to their localizing behavior as $\epsilon \rightarrow 0$ and their ability to approximate heat kernels. However, our results also hold for many other kernel choices; e.g., [28]. Having constructed k_Q, the kernel p_Q associated with the integral operator P_Q is obtained via a *Markov normalization* procedure [10,16], described in Sect. 4.3. With these definitions, we are ready to state our main results.

Theorem 1 *Under Assumption 1, there exists a real, self-adjoint, ergodic, compact Markov operator $P : L^2(X, \mu) \rightarrow L^2(X, \mu)$, 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 $p_Q : M \times M \rightarrow \mathbb{R}$ 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 $p : M \times M \rightarrow \mathbb{R}$ 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 $1 \leq p < \infty$.*

The operator-norm convergence of the compact operators P_Q to P leads to the following spectral convergence result (e.g., Sect. 7 in [5] and [2]).

Corollary 2 (spectral convergence) *Under the assumptions of Theorem 1, the following hold:*

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

(ii) *Let Π be any projector to the eigenspace W_λ of P at eigenvalue λ. Let also Π_Q be any projector to the union of the eigenspaces of P_Q corresponding to the eigenvalues in $\mathrm{spec}(P_Q) \cap S$. Then, as $Q \rightarrow \infty$, Π_Q converges strongly to Π. Moreover, the gap (distance) between W_λ and $\mathrm{ran} \, \Pi_Q$, defined as in [5], converges to zero.*

Theorem 3 below is a continuation of Theorem 1, and can be used to conclude some useful properties of the operator P.

Theorem 3 *Let Φ^t be a measurable flow on a compact metric space X supporting an invariant ergodic probability measure μ, 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) *τ lies in the tensor product subspace $\mathcal{D} \otimes \mathcal{D}$, and is invariant under the flow $U^t \times U^t$.*
(ii) *$\mathcal{D}$ and $\mathcal{D}^\perp$ are invariant under T. Moreover, $\overline{\text{ran}\, T}$ is a subspace of $\mathcal{D}$, $\mathcal{D}^\perp$ is a subspace of $\ker T$, and both $\text{ran}\, T$ and $\ker T$ are invariant under U^t.*

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

(iii) *There exists a measurable map $\pi : X \rightarrow \mathbb{T}^D$ for some $D \in \mathbb{N}$, whose components consist of joint eigenfunctions of T and U^t, such that π factors Φ^t into a rotation on the torus by a vector $\boldsymbol{\omega} \in \mathbb{R}^D$, i.e., $\pi(\Phi^t(x)) = \pi(x) + \boldsymbol{\omega} t \bmod 2\pi$ for μ-a.e. $x \in X$.*
(iv) *If the point spectrum of U^t has a set of m generating eigenfrequencies, then there is an integer $D \leq m$, and a symmetric kernel $\hat{\tau} \in L^2(\mathbb{T}^D \times \mathbb{T}^D, \text{Leb})$ on the D-torus, such that $\tau(x, y) = \hat{\tau}(\pi(x), \pi(y))$ for $\mu \times \mu$-a.e. $(x, y) \in X \times X$.*

The concept of generating eigenfrequencies will be described in Sect. 3. Note that Theorems 1 and 3 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. In what follows, $F_\mathcal{D} : M \rightarrow \mathbb{R}^d$ will be the map given by orthogonally projecting each of the d components of F onto the quasiperiodic subspace $\mathcal{D}$ from (2).

Assumption 3 All Koopman eigenfunctions, as well as the quasiperiodic component of the observation map $F_\mathcal{D}$, are continuous.

Although we explicitly assume that $F_\mathcal{D}$ is continuous, we are not aware of a counterexample where the observation map F is continuous (in accordance with Assumption 1), the Koopman eigenfunctions z are continuous, but $F_\mathcal{D}$ is not continuous. In particular, the examples that we study in Sect. 8 are Cartesian products of two dynamical systems for which $\mathcal{D}^\perp$ and $\mathcal{D}$, respectively, are trivial. We formally state in Corollary 28(ii) why such systems satisfy Assumption 3. On the other hand, smooth dynamical systems on smooth manifolds with discontinuous Koopman eigenfunctions (and in fact, pure point spectra) are known to exist, in both discrete- [1] and continuous-time settings [19]. This indicates that the continuity requirement on Koopman eigenfunctions in Assumption 3 is complementary to the assumed continuity of the dynamical flow in Assumption 1. The following theorem establishes a number of properties of P under these additional continuity assumptions.

Theorem 4 *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 μ-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 μ-a.e. to Pf.*

Remark The class of integral operators P_Q studied in this work has previously been used for dimension reduction and mode decomposition of high-dimensional time series (e.g., [11,32,33,56]). In these works, a phenomenon called in [11] "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 3 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 ω. Even though in this work we focus on the class of Markov operators P_Q, analogous results also hold for other classes of integral operators for data analysis that employ delays, including the covariance operators used in singular spectrum analysis (SSA) [13,50,61] and the related Hankel matrix analysis [3,14,59]. Collectively, these results establish a connection between two major branches of data analysis techniques for dynamical systems, namely those based on Koopman operators, and those based on kernel integral operators.

Theorems 1–4 are proved in Sect. 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 [30]. Theorem 1 generalizes this result to non-smooth state spaces and Koopman operators with mixed spectra. The spectral convergence of kernel integral operators was also studied in [62], but in the setting of continuous kernels. In contrast, here we consider an L^2 limit of a family of continuous kernels, which may not (and generally, will not) be continuous. The convergence properties of this family are related to the ergodic properties of the underlying dynamical system, and this link with the dynamics is a new feature.

With these results, the eigenvalues and eigenfunctions of P_Q consistently approximate those of P, and the latter can be used to construct orthonormal bases of Koopman eigenspaces. The availability of such bases is useful in many applications, including approximation techniques for the eigenvalues and eigenfunctions of U^t or its generator (defined in Sect. 3 ahead). One such technique will be presented in Sect. 6, utilizing the eigenvalues and eigenfunctions of P to perform diffusion regularization of the generator, and then solving the eigenvalue problem for the generator via a Petrov–Galerkin method. Note that the Markov property of P is not trivial; for instance, it does not hold for covariance kernels. The commutativity between U^t and P, in conjunction with the Markov property, lead to well posedness of these schemes despite the presence of a continuous spectrum of the generator.

Physical measures A point $x \in M$ is said to be in the basin of the measure μ with respect to the discrete-time map $\Phi^{\Delta t}$ if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\Phi^{n\,\Delta t}(x)) = \int_X f(y)\,d\mu(y), \qquad \forall f \in C^0(M). \tag{6}$$

The basin $\mathcal{B}_\mu$ of an invariant ergodic measure μ always includes μ-a.e. point in the support of μ (in this case, X), and is a forward-invariant set. An important property that we need the invariant measure μ to have is that it is physical [65]. Moreover, we will require that the dynamics has a suitable absorbing ball property. These assumptions can be summarized as follows:

Assumption 4 The set $\mathcal{B}_\mu$ of points satisfying (6) has positive Lebesgue measure, i.e., the measure μ is physical. Moreover, there exists a subset $\mathcal{V} \subseteq \mathcal{B}_\mu$, 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.

Examples where Assumption 4 is satisfied include: (i) ergodic flows on compact manifolds with Lebesgue absolutely continuous, fully supported, invariant measures, in which case $\mathcal{U} = \mathcal{V} = \overline{\mathcal{B}_\mu} = M = X$; (ii) certain classes of dissipative flows on potentially noncompact

manifolds [e.g., the Lorenz 63 (L63) system on $M = \mathbb{R}^3$ [43] studied in Sect. 8 ahead]; and (iii) certain classes of dissipative partial differential equations possessing inertial manifolds and physical measures [42,44].

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 data-driven operators $P_{Q,N}$, acting on the finite-dimensional Hilbert space $L^2(\mathcal{U}, \mu_N)$ associated with the sampling probability measure $\mu_N = \sum_{n=0}^{N-1} \delta_{x_n}/N$. These operators are constructed from time-ordered measurements $F(x_0), \ldots, F(x_{N-1})$ of the observable F analogously to (3)–(5), replacing throughout integrals with respect to the invariant measure μ by integrals with respect to the sampling measure μ_N. Moreover, because P_Q and $P_{Q,N}$ act on different Hilbert spaces, we will approach the problem of comparing their eigenvalues and eigenfunctions through integral operators $P_Q'' : C^0(\mathcal{U}) \to C^0(\mathcal{U})$ and $P_{Q,N}'' : C^0(\mathcal{U}) \to C^0(\mathcal{U})$, defined analogously to P_Q and $P_{Q,N}$, respectively, but acting on the same Banach space of continuous functions on $\mathcal{U}$. A complete description of these constructions will be made in Sect. 7.

Theorem 5 *Let Assumptions 1–4 hold. Then, for any initial point $x_0 \in 300\ \mathcal{V}$:*

(i) *Every eigenfunction of P_Q ($P_{Q,N}$) at nonzero eigenvalue extends to a continuous eigenfunction of P_Q'' ($P_{Q,N}''$), corresponding to the same eigenvalue.*

(ii) *As $N \to \infty$, $P_{Q,N}''$ converges in spectrum to P_Q'' in the sense of Corollary 2.*

Theorem 5 will be proved in Sect. 7. There is some similarity between our methods and papers on spectral convergence of kernel algorithms, e.g., [7,58,62], but our assumptions distinguishes Theorem 5 from previously studied cases. In particular, we do not assume an i.i.d. sequence of observed quantities, or that the sampled sequence $(x_n)_{n=0}^{N-1}$ lies on the support X of the invariant measure (as assumed in [7,62]). Finally, X need not have a manifold structure (as assumed in [7,58] and other manifold learning algorithms).

Figure 1 shows numerical eigenfunctions of $P_{Q,N}$ obtained from data generated by two mixed-spectrum dynamical systems, described in (39) and (40), respectively. In both examples, we start with a C^∞ vector field $\mathbf{V}$ on a smooth manifold M. In the first example, $M = X = \mathbb{T}^4$, so $\mathcal{U} = X = M$; in the second example, $M = \mathbb{R}^3 \times S^1$ and $X = X_{\mathrm{Lor}} \times S^1 \subset M$, where X_{Lor} is the Lorenz 63 attractor embedded in $\mathbb{R}^3$. Eigenfunctions of the operator $P_{Q,N}$ are then computed using a large number of delays, $Q = 2000$.

Using the eigenvalues and eigenfunctions of $P_{Q,N}$, we will also construct data-driven Galerkin schemes for the eigenvalue problem of the generator, which are structurally identical to its counterparts formulated in terms of the eigenvalues and eigenfunctions of P. Because we do not assume a priori knowledge of the vector field of the dynamics and/or closed-form expressions for the eigenfunctions of $P_{Q,N}$, these schemes will estimate the action of the generator on eigenfunctions through finite-difference approximations at the sampling interval Δt. In effect, Δt will play the role of an additional asymptotic approximation parameter, such that the data-driven solutions converge in a suitable joint limit of vanishing sampling interval ($\Delta t \to 0$), large data ($N \to \infty$), infinitely many delays ($Q \to \infty$), and infinite Galerkin approximation space dimension. This convergence result, along with minimal regularity requirements on the dynamical flow and the kernel, will be stated in a precise manner in Proposition 26 and Assumption 6, respectively. Note that, intuitively, our data-driven Galerkin framework for the generator V requires Δt as an additional approximation parameter over methods that approximate the Koopman subgroup generated by $U^{\Delta t}$ at a fixed time step Δt,
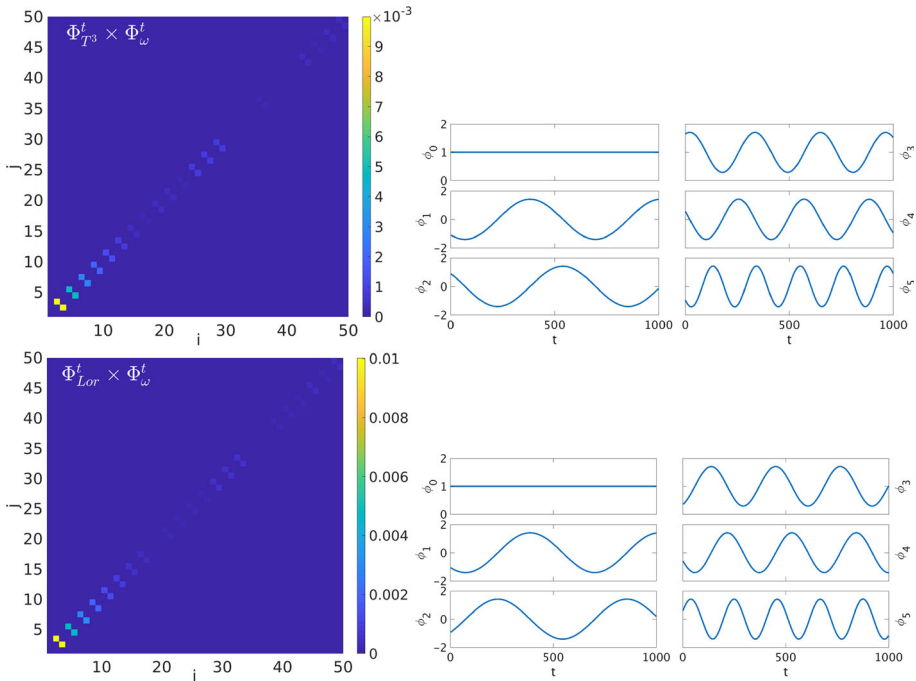
Fig. 1 Representative eigenfunctions of $P_{Q,N}$ and the associated matrix representation of the generator V from (7) for the torus-based flow $\Phi_{\mathbb{T}^3}^t \times \Phi_\omega^t$ in (39) (top panels) and the L63-based flow $\Phi_{\text{Lor}}^t \times \Phi_\omega^t$ in (40) (bottom panels). The eigenfunctions ϕ_i have been computed using a large number of delays, $Q = 2000$, and plotted as a time series along an orbit. These time series are near-sinusoidal, with frequencies close to integer multiples of the rotation frequency ω. Moreover, each frequency has multiplicity 2, and the corresponding time series are phase-shifted by $\pi/2$. The left-hand panels show the absolute values $|V_{ij}| = |\langle \phi_i, V\phi_j \rangle|$ of the matrix representation of the generator in the $\{\phi_i\}$ basis. Note that $V_{ji} \approx -V_{ij}$, which is consistent with the fact that V is a skew-adjoint operator. Since the first eigenfunction ϕ_0 of $P_{Q,N}$ is the constant function and $V\phi_0 = 0$, the first column and row only have zero entries. Together, the 2×2 block-diagonal form of the matrix representations of V and the structure of the eigenfunction time series indicate that each of the pairs (ϕ_1, ϕ_2), (ϕ_3, ϕ_4), ... spans an eigenspace of V, which is consistent with Theorems 1, 3–5 and Corollary 2

since V encodes the information of the entire Koopman group, parameterized by the real time parameter t.

Outline of the paper In Sect. 3, we review some important concepts from the spectral theory of dynamical systems. In Sect. 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, and 4. Next, we prove these theorems and Corollary 2 in Sect. 5. In Sect. 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 Sect. 7, we introduce the data-driven realization $P_{Q,N}$ 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 Sect. 8, the methods are applied to two mixed-spectrum flows, followed by a discussion of the results.

3 Overview of Spectral Methods for Dynamical Systems

In this section, we review some concepts from the spectral theory of dynamical systems and establish basic 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 Φ^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 [57], 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

$$Vf := \lim_{t \to 0} \frac{1}{t} \left(U^t f - f \right), \quad f \in D(V). \tag{7}$$

The operators U^t and V share the same eigenfunctions; in particular, $z \in D(V)$ and $\omega \in \mathbb{R}$ satisfy $U^t z = e^{i\omega t} z$ for every t iff

$$Vz = i\omega z.$$

In light of (7) 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 $\mathbf{V}$ on a C^1 manifold M, then under appropriate regularity conditions (for example, $\mathbf{V}$ is locally Lipschitz continuous and satisfies suitable growth bounds at infinity), this vector field induces a C^1 flow $\Phi^t : M \to M$ defined for all $t \in \mathbb{R}$. Suppose that there is a compact invariant set $X \subseteq M$ supporting an ergodic invariant measure μ. This set X is not necessarily a submanifold, and may not even have any differentiability properties. Nevertheless, (X, Φ^t, μ) 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 $\mathbf{V}$, the latter viewed as an operator $\mathbf{V} : C^1(M) \to C^0(M)$. For example, in quasiperiodic systems, $X = M = \mathbb{T}^m$, $\mathbf{V}$ generates a rotation, and μ is equivalent to the Lebesgue volume measure. On the other hand, for the Lorenz attractor (see (38)), $M = \mathbb{R}^3$, $\mathbf{V}$ is smooth and dissipative, X is a compact subset with non-integer fractal dimension [46], and μ is supported on X.

Eigenfunctions as factor maps We state the following properties of a Koopman eigenfunction z of a measure-preserving, ergodic dynamical system.

1. If z corresponds to a nonzero eigenfrequency ω, then it has zero mean with respect to the invariant measure μ. This can be concisely expressed as $\langle 1, z \rangle = 0$.
2. The flow Φ^t is semi-conjugate to the irrational rotation by ωt 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 μ-a.e.), but would be C^r if the eigenfunctions have a C^r representation.
3. Normalized eigenfunctions with $\|z\| = 1$ have $|z(x)| = 1$ for μ-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, i.e., $z(x) \in S^1$ for μ-a.e. $x \in X$.

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 fact that products of Koopman eigenfunctions are Koopman eigenfunctions leads to the following result about products of elements of $\mathcal{D}$ with elements of $\mathcal{D}^{\perp}$.

Lemma 6 *Let Φ^t be a measure-preserving, ergodic flow on a probability space (X, μ) such that U^t has a mixed spectrum. Then, for every $f \in \mathcal{D}$ and $g \in \mathcal{D}^{\perp}$ for which $fg \in L^2(X, \mu)$, fg lies in $\mathcal{D}^{\perp}$.*

The eigenvalues of V are closed under integer linear combinations. A finite set of eigenvalues $i\omega_1, \ldots, i\omega_m$ will be called a generating set if they are rationally independent, and every eigenvalue of V is of the form $i\omega_{\mathbf{a}} = i\sum_{j=1}^{m} a_j \omega_j$ for some $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{Z}^m$. In such a case, the corresponding eigenfunction is given by

$$z_{\mathbf{a}} = \prod_{j=1}^{m} z_1^{a_1} \cdots z_m^{a_m}, \tag{8}$$

where z_j is the eigenfunction at eigenvalue $i\omega_j$. By virtue of (8) the evolution of every observable $f \in \mathcal{D}$ under U^t has the closed-form expression

$$U^t f = \sum_{\mathbf{a} \in \mathbb{Z}^m} \hat{f}_{\mathbf{a}} e^{i\omega_{\mathbf{a}} t} z_{\mathbf{a}}, \quad \hat{f}_{\mathbf{a}} = \langle z_{\mathbf{a}}, f \rangle, \tag{9}$$

which can be evaluated given knowledge of finitely many generating eigenfunctions and eigenfrequencies. The following is a generalization of Property 2 of Koopman eigenfunctions listed above.

Proposition 7 *Given an arbitrary collection $\{z_{\mathbf{a}_1}, z_{\mathbf{a}_2}, \ldots, z_{\mathbf{a}_l}\}$ of l Koopman eigenfunctions, there exists a map $\pi : X \to \mathbb{C}^l$ with*

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

such that:

(i) *The image $\pi(X)$ is a torus of dimension $D \leq \min\{m, l\}$, where m is the number of generating frequencies. If $\omega_{\mathbf{a}_1}, \ldots, \omega_{\mathbf{a}_l}$ are rationally independent, then $D = l$.*
(ii) *The flow (Φ^t, μ) on X is semi-conjugate to an ergodic rotation (Ω^t, Leb) on $\mathbb{T}^D$ (i.e., $\pi \circ \Phi^t = \Omega^t \circ \pi$, μ-a.e.) associated with a frequency vector whose components are a subset of $\{\omega_{\mathbf{a}_1}, \ldots, \omega_{\mathbf{a}_l}\}$.*
(iii) *Every Koopman eigenfunction z whose corresponding eigenfrequency is a linear combination of the $\omega_{\mathbf{a}_1}, \ldots, \omega_{\mathbf{a}_l}$ satisfies $z(x) = \zeta(\pi(x))$ for μ-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 If $m > 1$, the set of eigenvalues $\{i\omega_{\mathbf{a}}\}_{\mathbf{a} \in \mathbb{Z}^m}$ 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 [30]. We will return to this point in Sect. 6.

Lemma 8 ([47], Sect. 2.3) *Let $\Delta t > 0$ be as in Assumption 1. Then, the orthogonal projection $\pi_\omega f$ of an observable $f \in L^2(X, \mu)$ onto the eigenspace of $U^{\Delta t}$ corresponding to the eigenvalue $e^{i\omega \Delta t}$ of $U^{\Delta t}$ is given by*

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

Moreover, $\pi_\omega \equiv 0$ if $i\omega$ is not an eigenvalue of the generator. 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 $\lim_{t \to \infty} t^{-1} \int_0^t |\langle g, U^s f \rangle| \, ds = 0$. The latter, is equivalent to the requirement that for Lebesgue almost every $\Delta t \in \mathbb{R}$, $\lim_{N \to \infty} N^{-1} \sum_{n=0}^{N-1} |\langle g, U^{n \Delta t} f \rangle| = 0$. The flow Φ^t is said to be (weak-) mixing if every $f \in L^2(X, \mu)$ is (weak-) mixing. It is known that every $f \in \mathcal{D}^\perp$ is weak-mixing (see, e.g., Mixing Theorem, p. 45 in [35]), whereas no observable in $\mathcal{D}$ is weak-mixing. Thus, the component $\mathcal{D}$, often called the quasiperiodic subspace, shows no decay of correlation, unlike its complement $\mathcal{D}^\perp$, which represents the chaotic component of the dynamics. In addition, weak-mixing observables in $\mathcal{D}^\perp$ and observables in $\mathcal{D}$ have a useful pointwise decorrelation property:

Lemma 9 *Let $f \in \mathcal{D}^\perp$ and $g \in \mathcal{D}$. Then, for μ-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,

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

which is equal to $g^*(x)\pi_\omega f(y)$ by Lemma 8. The latter is equal to zero since $f \in \mathcal{D}^\perp$. $\quad\square$

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, under the assumed compactness of X, the following properties hold [e.g., [25,26]]:

1. K is a Hilbert-Schmidt, and therefore compact, operator on $L^2(X, \mu)$, with operator norm bounded by $\|k\|_{L^2(X \times X)}$.
2. If k is symmetric, then K is self-adjoint.
3. If k is C^0, then Kf is also C^0 for every $f \in L^2(X, \mu)$.
4. If M is a C^r manifold and k is C^r, then Kf is also C^r for every $f \in L^2(X, \mu)$.

As stated in Sect. 2, we will work with kernels of the form

$$k_Q(x, y) = h(d_Q(x, y)), \tag{10}$$

where h is a continuous shape function on $\mathbb{R}$, and $d_Q : M \times M \to \mathbb{R}_0$ is the distance-like function on M from (4), parameterized by the number of delays Q. Kernels of this class are sometimes referred to as stationary kernels [28], as they only depend on distances between data points. For example, in (5), we used a Gaussian shape function, which is popular in manifold learning and other related geometrical data analysis techniques. Note that 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. That is, d_Q is a pseudo-distance on M, induced from delay-coordinate mapped data with Q delays.

The kernels in (10) satisfy Assumption 2(i), and the associated kernel integral operators K_Q have all four properties listed above. In addition, if h is strictly positive, k_Q satisfies Assumption 2(ii). The behavior of integral operators associated with other classes of kernels, e.g., the covariance operators employed in SSA and Hankel matrix analysis induced by inner products in data space, can be studied via similar techniques to those presented below. However, it should be kept in mind that the Markov normalization procedure described in Sect. 4.3 (which will be important for the well-posedness of the Galerkin schemes in Sects. 6 and 7) requires that the kernel be sign-definite. Another consideration to keep in mind is that the ability to approximate Koopman eigenfunctions with our techniques depends on the "richness" of the range of K_Q. As can be readily verified, the operator K_Q constructed from covariance kernels in d-dimensional data space (as in Assumption 1) has at most a dQ-dimensional range, whereas the corresponding operators associated with Gaussian kernels, as well as other non-polynomial kernels, have typically infinite-dimensional range for any Q. Our approach should also be applicable with little modification to families of kernels of the form

$$\tilde{k}_Q(x, y) = \frac{1}{Q} \sum_{q=0}^{Q-1} h(d_1(\Phi^{q\,\Delta t}(x), \Phi^{q\,\Delta t}(y))),$$

where averaging takes place after application of the shape function. Lemma 10 below states some useful properties of K_Q associated with strictly positive kernels. In what follows, 1_S will denote the constant function equal to 1 on a set S.

Lemma 10 *Under Assumptions 1 and 2(ii), for any $Q \in \mathbb{N}$, the functions $\rho_Q = K_Q 1_X$, and $\sigma_Q = K_Q \left(1/\rho_Q\right)$ are continuous and positive. Moreover, restricted on X, they are bounded away from zero.*

Proof The claims follow directly by compactness of X and the fact that $k_Q|_{X \times X}$ is a continuous function, bounded away from zero. $\qquad\square$

Intuitively, ρ_Q associated with the Gaussian kernel in (10) can be thought of as a "sampling density" on X. For instance, if X were a manifold embedded in $\mathbb{R}^{Qd}$ by a delay-coordinate map constructed from F, then up to an ϵ-dependent scaling, ρ_Q would approximate the density of the invariant measure μ relative to the volume measure associated with that embedding.

Remark In a number of applications, such as statistical learning on manifolds [6–8,10,16,58], one-parameter families of integral operators such as K_Q and P_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 3(ii), passing first to the limit $Q \to \infty$ allows one to view K and P as operators on functions on a smooth manifold, namely a D-dimensional torus, and study the small-ϵ 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 $\Psi : C^0(X) \to L^\infty(X \times X, \mu \times \mu)$, which maps a given observation function F into a (pseudo)metric on X, namely,

$$
\begin{aligned}
\Psi(F)(x, y) &:= \lim_{Q \to \infty} \Psi_Q(F)(x, y), \\
\Psi_Q(F)(x, y) &:= \frac{1}{Q} \sum_{q=0}^{Q-1} \left\| F(\Phi^{q \, \Delta t}(x)) - F(\Phi^{q \, \Delta t}(y)) \right\|^2 .
\end{aligned}
\tag{11}
$$

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

Theorem 11 *Let Assumption 1 hold, and $F = F_\mathcal{D} + F_{\mathcal{D}^\perp}$ be the L^2 decomposition of F from (2). Then, $\Psi(F)$ in (11) 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 $1 \le p < \infty$. Moreover:*

(i) *For every $t \in \mathbb{R}$ and μ-a.e. $x, y \in X$, $\Psi(F)(\Phi^t(x), \Phi^t(y)) = \Psi(F)(x, y)$.*
(ii) *For μ-a.e. $x, y \in X$, $\Psi(F)(x, y) = \Psi(F_{\mathcal{D}^\perp})(x, y) + \Psi(F_\mathcal{D})(x, y)$.*
(iii) *$\Psi(F_{\mathcal{D}^\perp})$ is a constant almost everywhere and equals $2\|F_{\mathcal{D}^\perp}\|_{L^2}^2$. Therefore,*

$$
\Psi(F) = \Psi(F_\mathcal{D}) + 2\|F_{\mathcal{D}^\perp}\|_{L^2}^2,
\tag{12}
$$

and $\Psi(F)$ lies in $\mathcal{D} \times \mathcal{D}$.

If, moreover, Assumption 3 holds:

(iv) *$\Psi(F_\mathcal{D}) \in C^0(X \times X)$ and $\Psi_Q(F_\mathcal{D})$ converges to $\Psi(F_\mathcal{D})$ 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 $\bar{\Psi}(F)$ μ-almost uniformly.*

Proof To prove well-definition of Ψ, note that $\Psi(F)$ exists μ-a.e. since it is the pointwise limit of the Birkhoff averages $\Psi_Q(F)$ of the continuous function $d_1 : (x, y) \to \|F(x) - F(y)\|$ 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 $L^p(X \times X, \mu \times \mu)$, $1 \le p < \infty$, since $\mu \times \mu$ is a probability measure. The $\Psi_Q(F) \to \Psi(F)$ convergence in $L^p(X \times X, \mu \times \mu)$, $1 \le p < \infty$, 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 $\Phi^{\Delta t} \times \Phi^{\Delta t}$. It can then be shown that $\Psi(F)$ must lie in the kernel of $V \otimes I + I \otimes V$, the generator for $U^t \otimes U^t$, 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 $G_{\mathcal{D}} : X \times X \to \mathbb{R}^d := (x, y) \mapsto F_{\mathcal{D}}(x_q) - F_{\mathcal{D}}(y_q)$, and similarly define $G_{\mathcal{D}^\perp} : X \times X \to \mathbb{R}^d$. Expanding the right-hand side of (11) gives,

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

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

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

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

which follows from Lemma 9. This completes the proof of Claim (ii).

To prove Claim (iii), let x_n and y_n denote $\Phi^{n\,\Delta t}(x)$ and $\Phi^{n\,\Delta t}(y)$, respectively. Then, (11) can be rewritten for $F_{\mathcal{D}^\perp}$ as

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

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

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

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

Since $F_{\mathcal{D}}$ is continuous, $\Psi(F_{\mathcal{D}})$ is continuous by a classic result of Krengel ([40], Theorem 1.2.7). This proves Claim (iv).

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_∞ 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 $\bar{\Psi}(F)$ is a consequence of Egorov's theorem. □

Remark Although the measure $\mu \times \mu$ is invariant under $\Phi^t \times \Phi^t$, it is not ergodic. In fact, it is ergodic iff (Φ^t, μ) is mixing (equivalently, U^t has purely continuous spectrum and a

simple eigenvalue at 1), in which case the metric d_∞ would be constant almost everywhere, in accordance with (12).

Theorem 11 establishes that the function $d_\infty : D(d_\infty) \to \mathbb{R}$, 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 $L^p(X \times X, \mu \times \mu)$, $1 \le p \le \infty$, with $\sup d_\infty \le \|d_1\|_{C^0(X \times X)}$. It can also be verified that d_∞ 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 $x \ne y$, 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 in (10), we employ d_∞ and the shape function h to define a corresponding kernel $k_\infty : M \times M \to \mathbb{R}$, where

$$k_\infty(x, y) = h(d_\infty(x, y)), \quad (x, y) \in D(d_\infty), \tag{13}$$

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

Proposition 12 shows that the operator K depends only on the quasiperiodic component of F, and is a direct consequence of Theorem 11 and (12).

Proposition 12 *Let (X, Φ^t, μ) and F be as in Theorem 1. Then, the integral operator K is a constant scaling operator iff its kernel k_∞ is a constant μ-a.e., which occurs iff $F_\mathcal{D}$ is a constant.*

In general, k_∞ may not be continuous. Nevertheless, it has a number of other useful properties, which follow directly from Theorem 11 in conjunction with the boundedness and continuity of the Gaussian shape function.

Lemma 13 *Under Assumption 1, the following hold:*

(i) *k_∞ is the $L^p(X, \mu)$-norm limit, $1 \le p < \infty$, of the sequence of continuous kernels $k_1, k_2, \ldots$.*
(ii) *k_∞ is invariant under $U^t \times U^t$ for all $t \in \mathbb{R}$.*
(iii) *k_∞ lies in $L^\infty(X \times X, \mu \times \mu)$, and under Assumption 2(ii), $1/k_\infty$ also lies in that space.*

Moreover, if Assumption 3 additionally holds:

(iv) *k_∞ is uniformly continuous on a dense, full-measure subset of $X \times X$.*
(v) *k_∞ has a unique continuous representative $\bar{k}_\infty \in C^0(X \times X)$, and as $Q \to \infty$, k_Q converges to k_∞ almost uniformly.*

The stronger regularity properties of k_∞ under Assumption 3 have the following important implications on the behavior of the corresponding integral operator.

Lemma 14 *Under Assumptions 1 and 3, the kernel integral operator K associated with k_∞ 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) *$\|K\| \le \|k_\infty\|_{L^\infty(X \times X)}$ 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 μ-a.e. to Kf.*

Proof (i) Since k_∞ is uniformly continuous on a set $S \subseteq X \times X$ of full $\mu \times \mu$ measure, there exists a full μ-measure set $X' \subseteq X$, such that for every $x \in X'$, $k_\infty(x, \cdot)$ is continuous μ-a.e. on X. Moreover, proceeding analogously to the proof of Theorem 11(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 $x_1, x_2 \in X'$,

$$
\begin{aligned}
|Kf(x_1) - Kf(x_2)| &= \left| \int_{X'} [k_\infty(x_1, y) - k_\infty(x_2, y)] f(y) d\mu(y) \right| \\
&\leq \|k_\infty(x_1, \cdot) - k_\infty(x_2, \cdot)\|_{L^2} \|f\|_{L^2} \\
&\leq \|k_\infty(x_1, \cdot) - k_\infty(x_2, \cdot)\|_{L^\infty}.
\end{aligned} \tag{14}
$$

Since k_∞ is uniformly continuous on S, for every $\epsilon > 0$, there exists $\delta > 0$ such that if $d_X(x_1, x_2) < \delta$, $\|k_\infty(x_1, \cdot) - k_\infty(x_2, \cdot)\|_{L^\infty} < \epsilon$. 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.

(ii) Since k_∞ is uniformly continuous on a dense set of full measure, for any $f \in C^0(X)$, the function $g : X \times X \to \mathbb{C}$ 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 μ-a.e. equal to $\bar{g}(x, \cdot)$ by μ-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),

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

and therefore

$$
\|Kf\|_{L^\infty} \leq \|k_\infty\|_{L^\infty(X \times X)} \|f\|_{L^2}. \tag{15}
$$

The bound on the L^2 operator norm follows by setting $\|f\|_{L^2} = 1$ in (15), together with the fact that $\|Kf\|_{L^2} \leq \|Kf\|_{L^\infty}$. The bound on the C^0 operator norm follows from (15) with $f \in C^0(X)$, in conjunction with the facts that $\|f\|_{L^2} \leq \|f\|_{C^0}$ and $\|Kf\|_{L^\infty} = \|Kf\|_{C^0}$.

(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 $g_n = Kf_n$ is equicontinuous. Let $\bar{k}_\infty \in C^0(X \times X)$ be the unique continuous representative of k_∞. For every $x_1, x_2 \in X$, 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 μ-a.e. convergence follows from Lemma 13(v).   □

We end this section with two important corollaries of Theorem 11 and Lemmas 13, 14, which are central to both Theorems 1 and 3.

Corollary 15 *The operators U^t and K commute.*

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

$$K f(x) = \int_X k_\infty(x, y) f(y) \, d\mu(y) = \int_X k_\infty(x, \Phi^t(y)) f(\Phi^t(y)) \, d\mu(y).$$

It therefore follows from Lemma 13(ii) that

$$K f(x) = \int_X k_\infty(\Phi^{-t}(x), y) f(\Phi^t(y)) \, d\mu(y) = U^{t*} K U^t f(x),$$

and the claim of the corollary follows.   □

Corollary 16 *Under Assumptions 1 and 2(ii), the function $\rho = K 1_X$ is μ-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 15 and the fact that $U^t 1_X = 1_X$ imply that $U^t \rho = \rho$, and it then follows by ergodicity that ρ is constant μ-a.e. That $\|1/\rho\|_{L^\infty}$ is finite follows from Lemma 13(iii). Finally, the continuity of ρ under Assumption 3 is a direct consequence of Lemma 14.   □

4.3 Markov Normalization

Next, we construct the Markov operators P_Q and P appearing in Theorems 1 and 3 by normalization of K_Q and K. Throughout this section, we consider that Assumptions 1 and 2 hold. Under these assumptions, we employ a normalization procedure introduced in the diffusion maps algorithm [16] and further developed in [10], although there are also other approaches with the same asymptotic behavior. Specifically, using the normalizing functions ρ_Q and σ_Q from Lemma 10 and ρ from Corollary 16, we introduce the kernels $p_Q : M \times M \to \mathbb{R}$ and $p : M \times M \to \mathbb{R}$, given by

$$p_Q(x, y) = \frac{k_Q(x, y)}{\sigma_Q(x) \rho_Q(y)}, \quad p(x, y) = \begin{cases} k_\infty(x, y)/\rho(x), & \rho(x) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

respectively. By Lemma 10, 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 13 and Corollary 16 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 $P_Q : L^2(X, \mu) \to L^2(X, \mu)$ and $P : L^2(X, \mu) \to L^2(X, \mu)$ from (3) associated with the kernels p_Q and p, respectively, and these operators are both Hilbert-Schmidt (see Sect. 4.1). Note that p and P have analogous properties to those stated for k_∞ and K in Lemmas 13, 14 and Corollary 15. In particular, p is invariant under $U^t \times U^t$, and P commute 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

$$\tilde{K}_Q f := K_Q \left(\frac{f}{K_Q 1_X} \right), \quad P_Q f = \frac{\tilde{K}_Q f}{\tilde{K}_Q 1_X}, \quad P f = \frac{K f}{K 1_X}. \tag{17}$$

In [10], the steps leading to $\tilde{K}_Q$ from K_Q and to P_Q from $\tilde{K}_Q$ are called right and left normalization, respectively. In the case of P, the effects of right normalization cancel since $K1_X$ is μ-a.e. constant by Corollary 16, so it is sufficient to construct this operator directly via left normalization of K.

As is evident from (17), 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 μ-a.e. $x \in M$, $\int_X p(x, \cdot) \, d\mu = 1$, 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 μ-a.e. symmetric by symmetry of k_∞ and the fact that ρ is μ-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 of $L^2(X, \mu)$ consisting of corresponding eigenfunctions, ϕ_j, with ϕ_0 being constant. On the other hand, because p_Q 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

$$\tilde{\sigma}_Q = \sigma_Q/\rho_Q, \quad \hat{\sigma}_Q = \sqrt{\sigma_Q \rho_Q},$$

where ρ_Q and σ_Q are as in Lemma 10. Let also D_Q be the multiplication operator which multiplies by $\tilde{\sigma}_Q$, and $\hat{P}_Q$ the kernel integral operator with kernel $\hat{p}_Q : M \times M \to \mathbb{R}$,

$$\hat{p}_Q(x, y) = \frac{k_Q(x, y)}{\hat{\sigma}_Q(x)\hat{\sigma}_Q(y)}. \tag{18}$$

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

$$\hat{P}_Q = D_Q^{1/2} P_Q D_Q^{-1/2}; \tag{19}$$

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

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

$$= \int_X \frac{k_Q(x, y)}{\hat{\sigma}_Q(x)\hat{\sigma}_Q(y)} f(y) \, d\mu(y) = \hat{P}_Q f(x).$$

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

1. $\hat{P}_Q$ has the same discrete spectrum as P_Q, consisting of eigenvalues $\lambda_{j,Q}$ with $1 = \lambda_{0,Q} > \lambda_{1,Q} \geq \lambda_{2,Q} \geq \cdots$.
2. Let $\phi_{j,Q}$ denote the eigenfunctions of $\hat{P}_Q$ corresponding to the nonzero eigenvalues $\lambda_{j,Q}$. These form an orthonormal basis for the closed subspace $\overline{\operatorname{ran} \hat{P}_Q} = (\ker \hat{P}_Q)^\perp$. Moreover, the $\phi_{j,Q}$ can be chosen to be real-valued.
3. The eigenfunction $\phi_{0,Q}$ of $\hat{P}_Q$ is equal up to proportionality constant to $\rho_Q \sigma_Q^{1/2}$.

Remark In applications, it may be the case that ρ_Q and $1/\rho_Q$ take a large range of values. In such situations, it may be warranted to replace (4) by a variable-bandwidth kernel of the form $k_Q(x, y) = \exp\left(-\frac{d_Q^2(x,y)}{\epsilon r_Q(x) r_Q(y)}\right)$, 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 [9,66], state space velocities [29,33], and local density estimates [8]. While we do not study variable bandwidth techniques in this work, our approach should be applicable in that setting too, so long as Corollary 16 holds.

5 Proof of Theorems 1, 3, 4 and Corollary 2

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 15. Next, as $Q \to \infty$, p_Q converges to p in any $L^p(X \times X, \mu \times \mu)$ norm with $1 \le p < \infty$ by the analogous result to Lemma 13(i) that holds for these kernels (see Sect. 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 $\|p_Q - p\|_{L^2(X \times X)}$.
□

Proof of Theorem 3 We first establish that τ is a.e. invariant under $\Phi^t \times \Phi^t$. Since the integral operator T commutes with U^t, for every $f \in L^2(X, \mu)$ and μ-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 μ under Φ^t. The only way the terms at the two ends of the equation can be equal for every $f \in L^2(X, \mu)$ is if $\tau(\Phi^t(x), \Phi^t(y')) = \tau(x, y')$ μ-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 $\mathcal{D} \otimes \mathcal{D}$, $\mathcal{D}^\perp \otimes \mathcal{D}^\perp$, $\mathcal{D}^\perp \otimes \mathcal{D}$, and $\mathcal{D} \otimes \mathcal{D}^\perp$. Since τ 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 τ just established. By symmetry of τ, the projections onto $\mathcal{D}^\perp \otimes \mathcal{D}$ and $\mathcal{D} \otimes \mathcal{D}^\perp$ vanish. Moreover, the projection $\tau_{\mathcal{D}^\perp \otimes \mathcal{D}^\perp} \in \mathcal{D}^\perp \otimes \mathcal{D}^\perp$ is orthogonal to constant functions, and it follows by the Birkhoff ergodic theorem that for $\mu \times \mu$-a.e. $x, y \in X \times X$,

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

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

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

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

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

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

Proposition 17 *For any nonzero eigenvalue λ of T, the corresponding eigenspace W_λ is invariant under the action of the Koopman generator V, and $V|_{W_\lambda}$ is diagonalizable. More-*

over, the constant function 1_X is an eigenfunction of T. If W_λ does not contain 1_X, its dimension is an even number.

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

Let λ_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 ζ of this basis. By our choice of λ, ζ is a non-constant eigenfunction of V, hence $\langle \zeta, 1 \rangle = 0$. Therefore, by ergodicity of (Φ^t, μ), $V\zeta = i\omega\zeta$ for some $\omega \neq 0$. This implies that ζ has non-zero real and imaginary parts. Hence, the conjugate ζ^* is linearly independent from ζ and corresponds to eigenvalue $-i\omega$ of V. However, since T is a real operator, ζ^* lies in W_λ. We therefore conclude that W_λ can be split into disjoint 2-dimensional spaces spanned by the conjugate pair of eigenfunctions ζ and ζ^*. Therefore $\dim W_\lambda$ is an even number if $\lambda \neq 1$. $\qquad\square$

Corollary 18 below follows from the fact that the closure of the range of P is spanned by the ϕ_j, and by Proposition 17, all ϕ_j with nonzero corresponding eigenvalue lie in $D(V)$.

Corollary 18 *The representation of $V|_{\overline{\mathrm{ran}\, P}}$ in the basis $\{\phi_0, \phi_1, \ldots\}$ has a block-diagonal structure, consisting of even-sized blocks associated with the eigenspaces $W_{\lambda \neq 1}$, and a 1×1 block with the element 0, associated with W_1. Moreover, the range of P lies in the domain of V, and $V|_{\overline{\mathrm{ran}\, P}}$ and $P|_{\overline{\mathrm{ran}\, P}}$ are simultaneously diagonalizable.*

Returning to the proof of Theorem 3, note that 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 Proposition 7, any collection of D such eigenfunctions factors the dynamics into a rotation on $\mathbb{T}^D$. This proves Theorem 3 (iii). To prove Claim (iv), we use (8) to expand the kernel as

$$\tau = \sum_{\mathbf{a},\mathbf{b} \in \mathbb{Z}^m} \tilde{\tau}_{\mathbf{ab}} z_{\mathbf{a}} \otimes z_{\mathbf{b}},$$

where m is the number of generating eigenfrequencies. In this expansion, there is a minimal number $D \leq m$ of generating eigenfunctions z_j from (8), arranged without loss of generality as $z_1, \ldots, z_D$, such that the expansion coefficients $\tilde{\tau}_{\mathbf{ab}}$ corresponding to $\mathbf{a} = (a_1, \ldots, a_m)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ with nonzero $a_{D+1}, \ldots, a_m$ and $b_{D+1}, \ldots, b_m$, respectively, vanish (in other words, the kernel τ does does not project onto the subspaces generated by $z_{D+1}, \ldots, z_m$ and their powers). By Proposition 7, the Koopman eigenfunctions corresponding to non-vanishing $\tilde{\tau}_{\mathbf{ab}}$ can be expressed as $z_{\mathbf{a}} = \zeta_{\mathbf{a}} \circ \pi$, where the $\zeta_{\mathbf{a}}$ are smooth Koopman eigenfunctions on $\mathbb{T}^D$ associated with an ergodic rotation. Thus, denoting the index set for the nonzero $\tilde{\tau}_{\mathbf{ab}}$ coefficients by $I \in \mathbb{Z}^m \times \mathbb{Z}^m$, we have $\tau(x, y) = \hat{\tau}(\pi(x), \pi(y))$ for $\mu \times \mu$-a.e., $(x, y) \in X \times X$, where $\hat{\tau}$ is the L^2 kernel on $\mathbb{T}^D$ given by

$$\hat{\tau} = \sum_{\mathbf{a},\mathbf{b} \in I} \tilde{\tau}_{\mathbf{ab}} \zeta_{\mathbf{a}} \otimes \zeta_{\mathbf{b}}.$$

This completes the proof of Claim (v) and of Theorem 3. $\qquad\square$

Proof of Theorem 4 That p is uniformly continuous on a full-measure, dense subset of $X \times X$ follows from the analogous result to Lemma 13(iv), which holds for p (see Sect. 4.3). Claims (i)–(iv) of the theorem follow analogously to Lemma 14. $\qquad\square$

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 19 (Convergence of commutators) *Let the assumptions of Theorem 4 hold, and the shape function h from (10) be continuously differentiable. Then, the following operators converge in $C^0(X)$ operator norm to 0 as $Q \to \infty$, with rates given below:*

(i) $\left\| U^{\Delta t} K_Q - K_Q U^{\Delta t} \right\|_{C^0} = O\left(Q^{-1}\right),$

(ii) $\left\| U^{\Delta t} \tilde{K}_Q - \tilde{K}_Q U^{\Delta t} \right\|_{C^0} = O\left(Q^{-1}\right),$

(iii) $\left\| U^{\Delta t} P_Q - P_Q U^{\Delta t} \right\|_{C^0} = O\left(Q^{-1}\right).$

Proof Let $\tilde{F}_{Q,\Delta t}(x, y) := \| F(x) - F(y) \|^2 - \left\| F(\Phi^{Q\,\Delta t} x) - F(\Phi^{Q\,\Delta t}(y)) \right\|^2$, and notice that by continuity of F and compactness of X this quantity is bounded on $X \times X$. Note that (i) $d_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y)) = d_Q(x, y) + Q^{-1}\tilde{F}_{Q,\Delta t}(x, y)$; and (ii) $h(\sqrt{u^2 + \Delta u}) = h(u) + O(\Delta u)$, as $\Delta u \to 0$. Thus,

$$k_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y)) = h(d_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y))) = h\left(\sqrt{d_Q^2(x, y) + Q^{-1}\tilde{F}_{Q,\Delta t}(x, y)}\right)$$

$$= h(d_Q(x, y)) + O(Q^{-1}) = k_Q(x, y) + O(Q^{-1}),$$

where the estimate holds uniformly with respect to $x, y \in X$. Therefore, for every $f \in L^2(X, \mu)$ and $x \in X$ we have

$$U^{\Delta t} K_Q f(x) = \int_X k_Q(\Phi^{\Delta t}(x), y) f(y) \, d\mu(y)$$

$$= \int_X k_Q(\Phi^{\Delta t}(x), \Phi^{\Delta t}(y)) f(\Phi^{\Delta t}(y)) \, d\mu(y)$$

$$= \int_X \left[k_Q(x, y) + O(Q^{-1}) \right] (U^{\Delta t} f)(y) \, d\mu(y).$$

Note that we have used the fact that μ is an invariant measure in the second-to-last line. Since k_Q is continuous, 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 and taking the supremum over $x \in X$ yields

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

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. □

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., [30], Remark 8), in which case, solving for its eigenvalues is a

numerically ill-posed problem. Following [30,34], we will address these issues by employing a Galerkin scheme for the eigenvalue problem of V, with a small amount of judiciously constructed diffusion added for regularization. Throughout this section, we consider that Assumptions 1, 2, and 3 hold. Further, we assume the following.

Assumption 5 The kernels k_Q, and thus k_∞, are symmetric positive-definite. That is, (i) $k_Q(x, y) = k_Q(y, x)$, for every $x, y \in M$; (ii) for every $x_0, x_1, \ldots, x_n \in M$ and $c_0, c_1, \ldots, c_n \in \mathbb{C}$, $\sum_{i,j=0}^{n-1} c_i^* k_Q(x_i, x_j) c_j \geq 0$; and (iii) the analogous conditions hold for k_∞.

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, the latter defined as the closed subspace of $\overline{\operatorname{ran} P} \subseteq \mathcal{D}$ orthogonal to constant functions (that is, H only consists of zero-mean functions). Note that $\{\phi_j\}_{j \in J}$, where J is an index set for the nonzero eigenvalues λ_j of P, strictly less than 1, is an orthonormal basis of H. For any $p \geq 0$, we define

$$H^p = \left\{ \sum_{j \in J} c_j \phi_j \in H : \sum_{j \in J} |c_j|^2 |\eta_j|^p < \infty \right\}, \quad \eta_j = (\lambda_j^{-1} - 1)/(\lambda_1^{-1} - 1). \quad (20)$$

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, $\{\eta_j\}_{j \in J}$). In particular, when (X, g) is a smooth Riemannian manifold with a metric tensor g satisfying $\operatorname{vol}_g = \mu$, and (η_j, ϕ_j) 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 equipped with the inner product

$$\langle f, g \rangle_{H^p} := \sum_{q=0}^p \sum_{j \in J} c_j^* d_j |\eta_j|^q,$$

where $f = \sum_{j \in J} c_j \phi_j$ and $g = \sum_{j \in J} d_j \phi_j$. Moreover, $\{\phi_j^{(p)}\}_{j \in J}$ with $\phi_j^{(p)} = \phi_j / \|\phi_j\|_p$, $\|\phi_j\|_{H^p}^2 = \sum_{q=0}^p \lambda_j^q$, forms an orthonormal basis of H^p.

Proposition 20 *For every $p > 0$, the space H^p is dense in H and moreover, the inclusion map $H^p \to H$, and thus $H^p \to L^2(X, \mu)$, is compact.*

Proof To see that H^p is dense, note that it includes all finite linear combinations of the ϕ_j. Since the ϕ_j form an orthonormal basis of H, these finite linear combinations are dense in H. Next, the embedding of H^p in H can be represented by a diagonal operator $G : H^p \to H$ such that $G_{jj} := \langle \phi_j, G\phi_j^{(p)} \rangle = \eta_j^{-p/2}$. This operator is compact iff G_{jj} converges to 0 as $j \to \infty$, which is true since $\lambda_j \to 0$. The compactness of the inclusion $H^p \to L^2(X, \mu)$ follows immediately. $\square$

Step 2. Regularized generator For every $\theta > 0$, we define the unbounded operators $\Delta : D(\Delta) \to H$ and $L_\theta : D(L_\theta) \to H$, where $D(\Delta) = D(L_\theta) \subset D(V)$, and

$$\Delta := f \mapsto \sum_{k=1}^\infty \eta_j \langle \phi_j, f \rangle \phi_j, \quad L_\theta := V|_{D(\Delta)} - \theta \Delta. \quad (21)$$

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 values of a Dirichlet energy functional. Theorem 21 below identifies a domain in which the operators in (21) are continuous, and establishes that the eigensolutions of L_θ converge to eigensolutions of V as $\theta \to 0$.

Theorem 21 *Viewed as operators from H^2 to H, the generator V, as well as the operators L_θ and Δ from (21), are bounded. In particular, we can set $D(\Delta) = D(L_\theta) = H^2$. Moreover, for every eigenvalue $i\omega$ of V, whose corresponding eigenspace lies in H^2, there exists an eigenvalue η of Δ such that the smooth curve $\theta \mapsto \gamma_\theta := i\omega - \theta\eta$ consists of eigenvalues γ_θ of L_θ, converging to $i\omega$ as $\theta \to 0^+$.*

Proof First, by Corollary 18, we can consider that the basis $\{\phi_j\}_{j \in J}$ of H consists of simultaneous eigenfunctions of V and P (and thus Δ), without loss of generality. To verify that V is a bounded operator on H^2, observe that the frequencies ω_j satisfy the growth bound

$$\omega_j = \|V\phi_j\| = \|\mathbf{V}\lambda_j^{-1}P\phi_j\| \le \lambda_j^{-1}\|p'\|_{L^2(X \times X, \mu \times \mu)} \le C\eta_j, \quad \forall j \in J. \tag{22}$$

Here $C > 0$ is some constant independent of j, and p' is the kernel defined pointwise as $p'(x, y) = Vp(\cdot, y)(x)$. Hence, for $f = \sum_{j \in J} c_j\phi_j \in H^2$,

$$\|Vf\|^2 = \left\|\sum_{j \in J} c_j V\phi_j\right\|^2 = \left\|\sum_{j \in J} ic_j\omega_j\phi_j\right\|^2 \le C^2\sum_{j \in J}|c_j|^2|\eta_j|^2 \le C^2\|f\|_{H^2}^2,$$

proving that V is a bounded operator on H^2. The same reasoning applies for L_θ and Δ. Finally, to establish convergence of the eigenvalues of L_θ to those of $V|_{H^2}$, let $i\omega_j$ be the eigenvalue of V corresponding to ϕ_j. Then, by definition of L_θ and the basis $\{\phi_j\}_{j \in J}$,

$$L_\theta\phi_j = V\phi_j - \theta\Delta\phi_j = (i\omega_j - \theta\eta_j)\phi_j,$$

and the claim follows immediately. This completes the proof of Theorem 21. □

Remark Theorem 21 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 21 above only for $V|_{H^2}$. In separate calculations, we have observed that an analog of the weak eigenvalue problem for L_θ formulated in $H^1 \times H^1$ actually performs well numerically.

Step 3. Galerkin method By virtue of Theorem 21, the eigenvalues of L_θ can be considered to be approximations of the eigenvalues of V. We will take the Galerkin approach in finding the eigenvalues of L_θ by solving for $z \in H^2$ and $\gamma \in \mathbb{C}$ in the following variational (weak) eigenvalue problem:

Definition 22 (*Regularized Koopman eigenvalue problem*) Find $\gamma \in \mathbb{C}$ and $z \in H^2$ such that for all $f \in H$,

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

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

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

In the above, the form $E : H \times H^2 \to \mathbb{C}$ 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 (η_j, ϕ_j) were set to Laplace-Beltrami eigenvalues and eigenfunctions, respectively, 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 η_j 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 $\eta_j^{(N)}$ of the η_j computed from finite data sets [e.g., as described in Sect. 7 ahead] increases with j [9,58,62], which is consistent with the intuitive expectation that rough (highly oscillatory) functions require larger numbers of samples for accurate approximations.

Following [30,34], we will order all solutions (γ_j, z_j) of the problem in Definition 22 in order of increasing Dirichlet energy $E(z_j)$. Since $A(f, f) = -\theta E(f, f)$ by skew-symmetry of V, we can compute the Dirichlet energy of eigenfunction z_j directly from the corresponding eigenvalue, viz. $E(z_j) = -\text{Re } \gamma_j / \theta$. Similarly, we have $\omega_j = \text{Im } \gamma_j$. By (22), there exist constants $C_1, C_2 > 0$ such that

$$C_2 \leq \frac{|i\omega_j - \theta\eta_j|}{|\eta_j|} \leq C_1, \ \forall j \in J. \tag{23}$$

To justify the well-posedness of the eigenvalue problem in Definition 22, we will state three important properties of A, namely,

$$|A(u, v)| \leq C_1 \|u\|_H \|v\|_{H^2}, \quad \forall u \in H, \quad \forall v \in H^2, \tag{24}$$

$$\sup_{\substack{f \in H \\ \|f\|_H = 1}} |A(f, v)| \geq C_2 \|v\|_{H^2}^2, \ \forall v \in H^2, \tag{25}$$

$$\sup_{\substack{g \in H^2 \\ \|g\|_{H^2} = 1}} |A(u, g)| \geq C_2 \|u\|_H^2, \ \forall u \in H. \tag{26}$$

We now give brief proofs of these results. In the following, $v = \sum_{j \in J} d_j \phi_j$ and $u = \sum_{j \in J} c_j \phi_j$ will be arbitrary functions in H^2 and H, respectively. Moreover, as in the proof of Theorem 21, we will assume that the basis $\{\phi_j\}_{j \in J}$ consists of simultaneous eigenfunctions of V and Δ. First, note that,

$$|A(u, v)| = \left| \sum_{j \in J} (i\omega_j - \theta\eta_j) c_j^* d_j \right| \leq \sum_{j \in J} |i\omega_j - \theta\eta_j| |c_j^* d_j|,$$

and by the Cauchy–Schwartz inequality on ℓ^2 and (23),

$$|A(u, v)| \leq C_1 \sum_{j \in J} |\eta_j| |c_j^* d_j| \leq C_1 \|u\|_H \|v\|_{H^2},$$

proving (24). To prove (25), let $f = \sum_{j \in J} a_j \phi_j \in H$. Then, the left-hand side of that equation becomes $\sum_{j \in J} (i\omega_j / \eta_j - \theta) \eta_j a_j^* d_j$. Let $R_j := i\omega_j / \eta_j - \theta$, where $|R_j| \geq C_2$ by (23). By the Cauchy–Schwarz inequality, under the constraint $\sum_{j \in J} |a_j|^2 = 1$, the sum $\left| \sum_{j \in J} a_j^* \eta_j d_j \right|$ attains the maximum value of $\sum_{j \in J} |\eta_j^2 d_j|^2$. Therefore,

$$\sup_{\substack{f \in H \\ \|f\|_H = 1}} |A(f, v)| = \sup_{\sum_{j \in J} |a_j|^2 = 1} \left| \sum_{j \in J} a_j^* d_j R_j \eta_j \right| \geq C_2 \sum_{j \in J} |\eta_j d_j|^2 = C_2 \|v\|_{H^2}^2.$$

This proves (25). The proof of (26) is similar to that of (25), with f replaced by a trial function $g = \sum_{j \in J} b_j \phi_j \in H^2$ and the constraint $\|g\|_{H^2}^2 = \sum_{j \in J} |b_j|^2 = 1$. A direct consequence of (25) and (26) is,

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

Equations (24), (25), (27), and the compact embedding of H^2 in H by Proposition 20 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_θ. See [5], Sect. 8, for an exposition on this classic result. The resulting finite-dimensional Galerkin approximations of the weak eigenvalue problem for L_θ can be summarized as follows:

Definition 23 *(Koopman eigenvalue problem, Galerkin approximation)* Set $\tilde{H}_m = $ 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 $\gamma \in \mathbb{C}$ and $z \in \tilde{H}_m^2$ such that for all $f \in \tilde{H}_m$,

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

where the sesquilinear form $A : H \times H^2 \to \mathbb{C}$ is as in Definition 22.

This problem is equivalent to solving a matrix generalized eigenvalue problem

$$\boldsymbol{A}\mathbf{c} = \lambda \boldsymbol{B}\mathbf{c}, \tag{28}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are $m \times m$ matrices with elements

$$
\begin{aligned}
A_{ij} &= A(\phi_i, \phi_j^{(2)}) = \frac{V_{ij}}{\eta_j} - \theta \Delta_{ij}, \quad V_{ij} = \langle \phi_i, V \phi_j \rangle, \quad \Delta_{ij} = \delta_{ij}, \\
B_{ij} &= \langle \phi_i, \phi_j^{(2)} \rangle = \eta_i^{-1} \delta_{ij},
\end{aligned}
\tag{29}
$$

respectively, and $\mathbf{c} = (c_1, \ldots, c_m)^\top$ is a column vector in $\mathbb{C}^m$ containing the expansion coefficients of the solution z in the $\{\phi_j^{(2)}\}$ basis of $\tilde{H}_m^2$, viz. $z = \sum_{j=1}^m c_j \phi_j^{(2)}$. It is important to note that, unlike the proofs of Theorem 21 and (24)–(26), in (29) we do not require that the ϕ_j be simultaneous eigenfunctions of V and P. This concludes the description of our Galerkin approximation of the eigenvalue problem for L_θ and therefore for V.

7 Data-Driven Approximation

In this section, we discuss the numeric procedures used to approximate the integral operators described in Sects. 4, 5, and implement the Galerkin method of Sect. 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 1–4 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 $\mathcal{U}$ (for, if x_0 were to lie in $\mathcal{V} \setminus \mathcal{U}$, the trajectory would enter $\mathcal{U}$ after finitely many steps, and its portion lying in $\mathcal{V} \setminus \mathcal{U}$ 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 $K' : L^2(X, \mu) \to C^0(\mathcal{U})$, 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 $\mathcal{U}$. 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 as $K'' = \iota \circ K'$, where $\iota : C^0(\mathcal{U}) \to L^2(X, \mu)$ is the canonical L^2 inclusion map on $C^0(\mathcal{U})$, and we can also define an analog $K'' : C^0(\mathcal{U}) \to C^0(\mathcal{U})$ 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 $\mathcal{U}$ from Assumption 4 always contains the support of μ_N. Moreover, since x_0 lies in the basin of the physical measure μ, as $N \to \infty$, μ_N converges weakly to μ, in the sense that

$$\lim_{N \to \infty} \int_{\mathcal{U}} f \, d\mu_N = \int_X f \, d\mu, \quad \forall f \in C^0(\mathcal{U}). \tag{30}$$

Our data-driven analog of the space $L^2(X, \mu)$ will be $L^2(\mathcal{U}, \mu_N)$; the set of equivalence classes of complex-valued functions on $\mathcal{U}$ which are square-summable and have common values at the sampled states x_n. By Assumption 1, X is not a single point, thus $L^2(\mathcal{U}, \mu_N)$ is isomorphic to $\mathbb{C}^N$. As a result, every element $f \in L^2(\mathcal{U}, \mu_N)$ can be represented in the canonical basis of $\mathbb{C}^N$ as an N-vector $\mathbf{f} = (f(x_0), \dots, f(x_{N-1}))$. In fact, $L^2(\mathcal{U}, \mu_N)$ is the image of $C^0(\mathcal{U})$ under the restriction map $\pi_N : C^0(\mathcal{U}) \to L^2(\mathcal{U}, \mu_N)$, where $\pi_N f = (f(x_0), \dots, f(x_{N-1}))$. Moreover, given any $f, g \in L^2(\mathcal{U}, \mu_N)$, we have $\langle f, g \rangle_{L^2(\mathcal{U}, \mu_N)} = \mathbf{f} \cdot \mathbf{g}/N$, where $\cdot$ denotes the canonical inner product on $\mathbb{C}^N$.

Kernel integral operators In the data-driven setting, given a continuous kernel $k : M \times M \to \mathbb{R}$, we define a kernel integral operator $K'_N : L^2(\mathcal{U}, \mu_N) \to C^0(\mathcal{U})$ 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 $K_N : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$ and $K''_N : C^0(\mathcal{U}) \to C^0(\mathcal{U})$ 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 $\boldsymbol{K}$ with elements $K_{ij} = k(x_i, x_j)/N$. In this representation, the function $g = K_N f$, $f \in L^2(\mathcal{U}, \mu_N)$, is represented by $\mathbf{g} = \boldsymbol{K}\mathbf{f}$.

When $k = k_Q$ from (10), one can similarly define operators $K'_{Q,N} : L^2(\mathcal{U}, \mu_N) \to C^0(\mathcal{U})$, $K_{Q,N} : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$, and $K''_{Q,N} : C^0(\mathcal{U}) \to C^0(\mathcal{U})$. This family of operators has the analogous properties to those stated for K_Q in Lemma 10; 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 $p_{Q,N} : M \times M \to \mathbb{R}$ 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 $x \in M$, and therefore induces the Markov operators $P'_{Q,N} : L^2(\mathcal{U}, \mu_N) \to C^0(\mathcal{U})$, $P_{Q,N} : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$, and $P''_{Q,N} : C^0(\mathcal{U}) \to C^0(\mathcal{U})$. Moreover, $P_{Q,N}$ is related to the self-adjoint operator $\hat{P}_{Q,N} : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$ with kernel $\hat{p}_{Q,N} : M \times M \to \mathbb{R}$,

$$\hat{p}_{Q,N}(x, y) = \frac{k_Q(x, y)}{\hat{\sigma}_{Q,N}(x)\hat{\sigma}_{Q,N}(y)}, \quad \tilde{\sigma}_{Q,N} = \sigma_{Q,N}/\rho_{Q,N}, \tag{31}$$

via a similarity transformation analogous to (19). From the kernel $\hat{p}_{Q,N}$ one can construct the operators $\hat{P}_{Q,N} : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$, $\hat{P}'_{Q,N} : L^2(\mathcal{U}, \mu_N) \to C^0(\mathcal{U})$, and $\hat{P}''_{Q,N} : C^0(\mathcal{U}) \to C^0(\mathcal{U})$.

Data-driven basis We will use eigenvectors $\phi_{j,Q,N}$ of $\hat{P}_{Q,N}$ as an orthonormal basis of $L^2(\mathcal{U}, \mu_N)$, and employ the corresponding eigenvalues, $1 = \lambda_{0,Q,N} > \lambda_{1,Q,N} \geq \cdots \geq \lambda_{N-1,Q,N} \geq 0$, to define data-driven analogs

$$\eta_{j,Q,N} = (\lambda_{j,Q,N}^{-1} - 1)/(\lambda_{1,Q,N}^{-1} - 1), \quad j \in J_N, \tag{32}$$

of the η_j in (20), where $J_N = \{j : \lambda_{j,Q,N} > 0\}$. The eigenvalue problem for $\hat{P}_{Q,N}$ 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 [30,31]. Note that for kernels k_Q with exponential decay, such as the Gaussian kernels in (4), $\hat{P}$ can be well approximated by a sparse matrix, allowing scalability of our techniques to large N.

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, as stated in Sect. 2, 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 24 *The following correspondence between the spectra of operators holds:*

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

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

Lemma 24 is a direct consequence of the definitions of $\hat{P}_{Q,N}$ and $P''_{Q,N}$. 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.

Collective compact convergence [62] A sequence of operators A_n on a Banach space B is said to be collectively compactly convergent to an operator A if $A_n \to A$ pointwise, and there is an $N \in \mathbb{N}$ such that $\cup_{n=N}^{\infty}(A - A_n)(B_1)$ has a compact closure in B. Here, B_1 is the

unit ball in B. The following proposition states that the data-driven operators $\hat{P}_{Q,N}$ converge collectively compactly, and as result in spectrum.

Proposition 25 *Let Assumptions* 1–5 *hold. Given a trajectory* $(x_n)_{n \in \mathbb{N}}$ *starting in* $\mathcal{U}$, *the corresponding sequence of operators* $\hat{P}''_{Q,N}$ *constructed from the observations* $F(x_0), \ldots, F(x_{N-1})$ *converges collectively compactly as* $N \to \infty$ *to* $\hat{P}''_Q$. *As a result, the sequence* $\hat{P}_{Q,N}$ *converges spectrally, analogously to Corollary* 2, *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 for the operators $P''_{Q,N}$ on $C^0(\mathcal{U})$ is analogous to techniques given in [62], and the full details will be omitted. In particular, the general assumptions in [62] that the sampling measures μ_N converge weakly to μ is guaranteed by the assumption that $x_0 \in \mathcal{U}$. As a result, Proposition 13 in [62] applies, and we obtain collective compact convergence of $\hat{P}''_{Q,N}$ to $\hat{P}''_Q$. By Propositions 14 and 15 in [62], respectively, this implies *compact convergence* (a weaker notion than collectively compact convergence), which is sufficient for spectral convergence on $C^0(\mathcal{U})$. The spectral convergence of $\hat{P}_{Q,N}$ to $\hat{P}_Q$ on $L^2(X, \mu)$ then follows from Lemma 24.

***Proof of Theorem* 5**   The claims of the theorem follow from analogous results to Lemma 24 and Proposition 25 for the operators $P_{Q,N}, P'_{Q,N}, P''_{Q,N}$ and P_Q, P'_Q, P''_Q.   □

Together, Lemma 24 and Proposition 25 imply that every eigenpair $(\lambda_{j,Q}, \phi_{j,Q})$ of $\hat{P}_Q$ can be consistently approximated by a sequence of eigenpairs $(\lambda_{j,Q,N}, \phi_{j,Q,N})$ of $\hat{P}_{Q,N}$. Moreover, by Corollary 2, as $Q \to \infty$, $(\lambda_{j,Q}, \phi_{j,Q})$ approximates in turn the eigenpair (λ_j, ϕ_j) of P; that is,

$$\lim_{Q \to \infty} \lim_{N \to \infty} \lambda_{j,Q,N} = \lambda_j, \quad \lim_{Q \to \infty} \lim_{N \to \infty} \lambda_{j,Q,N}^{-1} \iota \hat{P}'_{Q,N} \phi_{j,Q,N} = \phi_j, \tag{33}$$

where the limit $Q \to \infty$ in the second equation is taken with respect to the $L^2(X, \mu)$ norm. Since, as can be seen in (29), the Galerkin scheme in Sect. 6 can be entirely formulated using the λ_j and the matrix elements $\langle \phi_i, V\phi_j^{(2)} \rangle$ 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 $\phi_{j,Q,N}$. To that end, we will employ finite-difference approximations, as described below.

Finite-difference approximation   The action Vf of the generator on a function $f \in D(V)$ is defined via the limit in (7). This suggests that for data sampled discretely at sampling interval Δt, we can approximate Vf by a finite-difference approximation [29,30,34]. For example, the following are first- and second-order, approximation schemes for V, respectively:

$$V_{\Delta t} f = \frac{1}{\Delta t} (U^{\Delta t} f - f), \quad V_{\Delta t} f = \frac{1}{2\Delta t} \left(U^{\Delta t} f - U^{-\Delta t} f \right). \tag{34}$$

In the finite-sample case, we approximate $V_{\Delta t}$ by a corresponding r-th order finite-difference operator $V_{\Delta t, N} : L^2(\mathcal{U}, \mu_N) \to L^2(\mathcal{U}, \mu_N)$. For example, in the case of the first-order scheme in (34), $V_{\Delta t, N}$ becomes

$$V_{\Delta t, N} f(x_n) = \frac{f(x_{n+1}) - f(x_n)}{\Delta t}, \quad n \in \{0, \ldots, N-2\}, \tag{35}$$

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

Assumption 6 $\mathcal{U}$ is a forward-invariant C^1 compact manifold, and $\Phi^t|_{\mathcal{U}}$ is generated by a C^0 vector field $\mathbf{V}$. Moreover, $F|_{\mathcal{U}} \in C^1(\mathcal{U}; \mathbb{R}^d)$, and the kernel shape function $h : \mathbb{R} \to \mathbb{R}$ is C^1. $V_{\Delta t}$ and $V_{\Delta t, N}$ are first-order finite-difference schemes, as in (34) and (35), respectively.

Under Assumption 6, the generator V of the Koopman group is an extension of $\mathbf{V}$, viewed as a differential operator on $C^1(\mathcal{U})$. Moreover, we can approximate $\mathbf{V}$ by finite-difference schemes $\mathbf{V}_{\Delta t} : C^0(\mathcal{U}) \to C^0(\mathcal{U})$, defined analogously to (34) with $U^{\Delta t}$ replaced by $\Phi^{\Delta t}$. We then have:

Proposition 26 *Let Assumptions* 1, 2, *and* 6 *hold. Then for every* $i, j \in \mathbb{N}$:

(i) *The eigenfunctions* $\varphi_{j,Q,N}$ *and* $\varphi_{j,Q}$ *from Lemma* 24 *lie in* $C^1(\mathcal{U})$. *Moreover,* $\mathbf{V}_{\Delta t}\varphi_{j,Q}$ *converges to* $\mathbf{V}_{\Delta t}\varphi_{j,Q}$ *as* $\Delta t \to 0$, *uniformly on* $\mathcal{U}$.
(ii) $\lim_{\Delta t \to 0} \lim_{N \to \infty} \langle \hat{\phi}_i, V_{\Delta t, N} \hat{\phi}_j \rangle_{L^2(\mathcal{U}, \mu_N)} = \langle \phi_{i,Q}, V\phi_{j,Q} \rangle$.

Proof To prove Claim (i), note that under Assumption 6, for a finite number of delays Q, by (4), $\hat{p}_Q$ is a C^1-smooth kernel. Hence, according to [25], the ranges of the integral operators $\hat{P}'_Q$ and $\hat{P}'_{Q,N}$, and thus $\varphi_{j,Q}$ and $\varphi_{j,Q,N}$ lie in $C^1(\mathcal{U})$. Since the vector field $\mathbf{V}$ is C^0, the trajectories are C^1-smooth, and therefore, $\mathbf{V}\varphi_{j,Q}$, which is the time derivative along the orbit, has a first-order Taylor expansion. Claim (ii) is a consequence of Claim (i), in conjunction with the weak convergence of measures in (30) and Lemma 24. $\qquad\square$

Remark In many cases, such as flows induced on inertial manifolds in dissipative PDEs [18], the C^1 regularity in Assumption 6 cannot be strengthened. Proposition 26 provides the basis for numerically approximating V for these cases. If $M, \mathcal{U}, \mathbf{V}, F$, and h have a higher degree of smoothness, say C^r for some $r \geq 2$, then taking $V_{\Delta t}$ to be an $(r-1)$-th order finite-difference scheme would lead to an improved, $O(\Delta t)^{r-1}$, convergence.

Data-driven Galerkin method Using the $\eta_{j,Q,N}$ from (32), we define the data-driven normalized basis vectors $\hat{\phi}_j^{(p)} = \phi_{j,Q,N}/\hat{\eta}_{j,Q,N}^{p/2}$, $j \in J_N$ (cf. the $\phi_j^{(p)}$ from Step 1 in Sect. 6), and the associated Galerkin approximation spaces $H_{N,Q,m}^p = \text{span}\{\hat{\phi}_j^{(p)}\}_{j=1}^m \subseteq L^2(\mathcal{U}, \mu_N)$, $m \leq J_N$, where we abbreviate $H_{Q,N,J_N}^p =: H_{Q,N}^p$ and $H_{Q,N}^0 =: H_{Q,N}$. We also define the positive semidefinite, self-adjoint operator $\Delta_{Q,N} : H_{Q,N} \to H_{Q,N}$, where

$$\Delta_{Q,N} f = \sum_{j \in J_N} \eta_{j,Q,N} c_j \phi_{j,Q,N}, \quad f = \sum_{j=0}^{N-1} c_j \phi_{j,Q,N}.$$

This operator is a data-driven analog of Δ 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 23:

Definition 27 (*Koopman eigenvalue problem, data-driven form*) Find $\gamma \in \mathbb{C}$ and $z \in H_{Q,N,m}^2$ such that for all $f \in H_{Q,N,m}$,

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

where $A_{\Delta t, Q, N} : H_{Q,N} \times H_{Q,N} \to \mathbb{C}$ is the sesquilinear form defined as

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

Numerically, this is equivalent to solving a matrix generalized eigenvalue problem analogous to that in (28), viz.

$$\mathbf{Ac} = \lambda \mathbf{Bc},$$

where $\mathbf{A}$ and $\mathbf{B}$ are $m \times m$ matrices with elements

$$A_{ij} = A_{\Delta t, Q, N}(\phi_{i, Q, N}, \phi_{j, Q, N}^{(2)}) = \frac{V_{ij}}{\eta_{j, Q, N}} - \theta \Delta_{ij},$$

$$V_{ij} = \langle \phi_{i, Q, N}, V_{\Delta t, N} \phi_{j, Q, N} \rangle_{L^2(\mathcal{U}, \mu_N)}, \quad \Delta_{ij} = \delta_{ij},$$

$$B_{ij} = \langle \phi_{i, Q, N}, \phi_{j, Q, N}^{(2)} \rangle_{L^2(\mathcal{U}, \mu_N)} = \eta_{i, Q, N}^{-1} \delta_{ij},$$

respectively, and $\mathbf{c} = (c_1, \ldots, c_m)^\top$ is a column vector in $\mathbb{C}^m$ containing the expansion coefficients of the solution $z = \sum_{j=1}^m c_j \hat{\phi}_{j, Q, N}^{(2)}$ in the $\{\hat{\phi}_j^{(2)}\}$ basis of $H_{Q, N, m}^2$. Analogously to the continuous case, we define a data-driven Dirichlet energy functional $E_{Q, N}$ on $H_{Q, N}^2$, given by $E_{Q, N}(f) = \langle f, \Delta_{Q, N} f \rangle_{L^2(\mathcal{U}, \mu_N)}$, and use that functional to order the computed eigenfunctions in order of increasing Dirichlet energy. Note that, unless an antisymmetrization is explicitly performed, in the data driven setting, V_{ij} will generally not be equal to $-V_{ji}$, and thus Re γ will not be equal to $-\theta E_{Q, N}(z)$ (cf. Sect. 6). Nevertheless, in practice we observe that Re $\gamma \approx -\theta E_{Q, N}(z)$, at least for the leading eigenfunctions.

For any fixed m, and up to similarity transformations, the matrices $\mathbf{A}$ and $\mathbf{B}$ 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 (28). We therefore conclude that the data-driven Galerkin method in Definition 27 is consistent (as $\Delta t \to 0$ and $Q, N \to \infty$) with the Galerkin method in Definition 23, which is in turn consistent (as $m \to \infty$) with the weak eigenvalue problem for the regularized generator L_θ in Definition 22.

8 Results and Discussion

In this section, we apply the methods described in Sects. 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 eigenspaces of $P_{Q, N}$ from (17) are eigenspaces of U^t; and (ii) the eigenvalues obtained using the Galerkin scheme in Definition 27 are consistent with those expected theoretically.

8.1 Two Systems with Mixed Spectrum

The first system studied below is based on a mixing flow on the 3-torus introduced by Fayad [24]. The flow, denoted by $\Phi_{\mathbb{T}^3}^t$, is given by the solution of the ordinary differential equation (ODE) $d(x, y, z)/dt = \mathbf{V}(x, y, z)$, where $(x, y, z) \in \mathbb{T}^3$, and $\mathbf{V}$ is the smooth vector field

$$\mathbf{V}(x, y, z) = \mathbf{v}/\varphi(x, y, z), \quad \varphi(x, y, z) = 1 + \sum_{k=1}^\infty \frac{e^{-k}}{k} \text{Re} \left[\sum_{|l| \leq k} e^{ik(x+y)+ilz} \right], \quad (36)$$

parameterized by the constant frequency vector $\mathbf{v}$. Hereafter, we set $\mathbf{v} = (\sqrt{2}, \sqrt{10}, 1)^\top$. Note that the orbits under $\Phi_{\mathbb{T}^3}^t$ are the same as that of the ergodic, non-mixing linear flow along the vector $\mathbf{v}$. $\Phi_{\mathbb{T}^3}^t$ has a unique Borel, invariant, ergodic probability measure $\mu_{\mathbb{T}^3}$ with density

$\varphi / \int_M \varphi \, d$ Leb relative to Lebesgue measure. Such flows are also called time-reparameterized flows as $\boldsymbol{v}$ is scaled by the function $\varphi(x, y, z)$ at each point $(x, y, z) \in \mathbb{T}^3$.

In [24], it is shown that this system is mixing, and thus weak-mixing, with respect to its invariant measure $\mu_{\mathbb{T}^3}$. As a result, its Koopman generator has continuous spectrum and a single eigenvalue at zero corresponding to constant eigenfunctions (see Sect. 3). To construct an associated mixed-spectrum system, we take the product $\Phi^t_{\mathbb{T}^3} \times \Phi^t_\omega$ with a periodic flow Φ^t_ω on S^1, defined as

$$d\Phi^t_\omega(\alpha)/dt = \omega, \quad \omega = 1. \tag{37}$$

Thus, the state space of the product system is $M = \mathbb{T}^3 \times S^1 = \mathbb{T}^4$. Note that in this example the invariant 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 $\mathbf{V} \oplus \boldsymbol{\omega} : C^\infty(X) \to L^2(X, \mu)$, where $\boldsymbol{\omega} : C^\infty(S^1) \to C^\infty(S^1)$ is the differential operator $f \mapsto \boldsymbol{\omega}(f) := \omega f'$. Φ^t_ω has an associated pure point spectrum, with the eigenfrequencies being integer multiples of ω. Since $\Phi^t_{\mathbb{T}^3}$ has no non-zero corresponding eigenfrequencies, 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 [43]. This system is known to have a chaotic attractor $X_{\text{Lor}} \subset \mathbb{R}^3$ with fractal dimension 2.0627160 [46], supporting a physical invariant measure μ_{Lor} [60] with a compact absorbing ball [41]. Moreover, the flow is mixing with respect to μ_{Lor} [45]. Similarly to the $\mathbb{T}^3$ system described above, the latter implies that the Koopman generator of the L63 system has only constant eigenfunctions, corresponding to eigenvalue 0. The flow, denoted by Φ^t_{Lor}, is generated by a smooth vector field $\mathbf{V} \in C^\infty(\mathbb{R}^3; \mathbb{R}^3)$, whose components at $(x, y, z) \in \mathbb{R}^3$ are

$$V_1 = \sigma(y - x), \quad V_2 = x(\rho - z) - y, \quad V_3 = xy - \beta z. \tag{38}$$

Throughout, we use the standard parameter values $\beta = 8/3$, $\rho = 28$, and $\sigma = 10$. As in the torus case, we form the product $\Phi^t_{\text{Lor}} \times \Phi^t_\omega$ with the rotation Φ^t_ω in (37), 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 invariant 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 $F : M \to \mathbb{R}^3$ coupling the degrees of freedom of the mixing subsystem with the rotation. In the case of the torus-based system, we define $F(x, y, z, \alpha) = (F_1, F_2, F_3)$, $(x, y, z) \in \mathbb{T}^3$, $\alpha \in S^1$, via additive coupling, viz.

$$F_1 = \sin\alpha + \sin x, \quad F_2 = \cos\alpha + \sin y, \quad F_3 = \sin(2\alpha) + \sin z. \tag{39}$$

In the case of the L63-based system, the coupling is nonlinear with $F(x, y, z, \alpha) = (F_1, F_2, F_3)$, $(x, y, z) \in \mathbb{R}^3$, $\alpha \in S^1$, and

$$F_1 = \sin(\alpha + x), \quad F_2 = \cos(2\alpha + y), \quad F_3 = \cos(\alpha + z). \tag{40}$$

Note that all of the examples that we study satisfy Assumption 4. For Φ^t_{Lor}, $M = \mathbb{R}^3$, $\mathcal{V} = \mathcal{V}_{\text{Lor}} = \mathcal{B}_{\mu_{\text{Lor}}}$, $\mathcal{U}$ is the absorbing ball $\mathcal{U}_{\text{Lor}}$, and $\mu = \mu_{\text{Lor}}$. For $\Phi^t_{\text{Lor}} \times \Phi^t_\omega$, $M = \mathbb{R}^3 \times S^1$, $\mathcal{V} = \mathcal{V}_{\text{Lor}} \times S^1$, $\mathcal{U} = \mathcal{U}_{\text{Lor}} \times S^1$, and $\mu = \mu_{\text{Lor}} \times \text{Leb}_{S^1}$. For $\Phi^t_{\mathbb{T}^3} \times \Phi^t_\omega$, $M = \mathcal{V} = \mathcal{U} = \mathbb{T}^3 \times S^1$, and $\mu = \mu_{\mathbb{T}^3} \times \text{Leb}_{S^1}$.
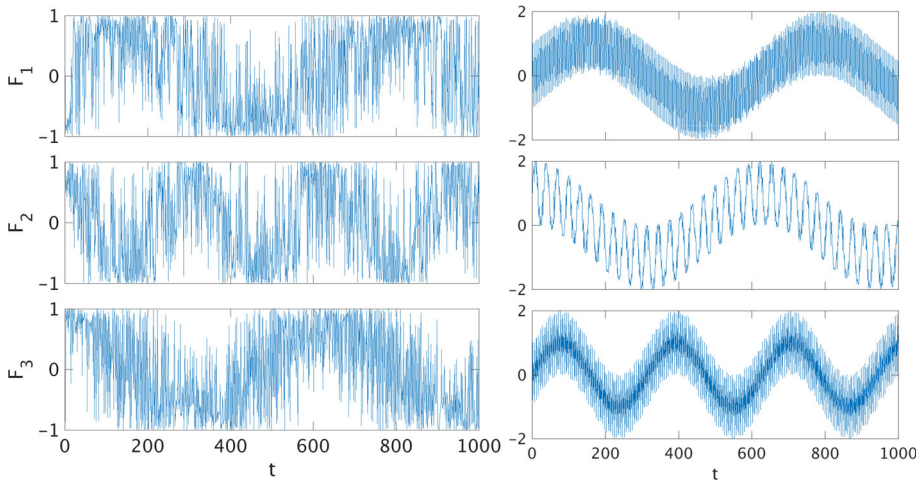
Fig. 2 Time series of the observation maps of the L63-based system (40) (left) and the torus-based system (39) (right). Each time series is a nonlinear combination of data generated from two sources, one with a purely continuous spectrum, (38) and (36), respectively, and one with a purely discrete spectrum, (37). The time series clearly exhibit complex evolution, characteristic of chaotic dynamics, and recovering from them Koopman eigenvalues and eigenfunctions is a non-trivial task

8.2 Experimental Results

We generated numerical trajectories $x_0, x_1, \ldots, x_{N-1}$ of the torus- and L63-based systems from Sect. 8.1, 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, and that x_0 has entered the forward-invariant set $\mathcal{U}$. In both cases, the number of samples was $N = 50{,}000$, the integration time-step was 0.01, and the number of delays was $Q = 2000$. Gaussian kernels k_Q from (5) were used throughout. We employed the `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 (39) and (40) to 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 eigenpairs $(\lambda_{j,Q,N}, \phi_{j,Q,N})$ by solving the eigenvalue problem for the operator $\hat{P}_{Q,N}$ from (31), using Matlab's `eigs` iterative solver. Henceforth, for ease of notation, we abbreviate $\lambda_{j,Q,N}$ and $\phi_{j,Q,N}$ by $\hat{\lambda}_j$ and $\hat{\phi}_j$, respectively. The bandwidth parameter ϵ of the Gaussian kernels was selected using the tuning procedure described in [8,12,17], which yielded $\epsilon \approx 3.6$ and ≈ 2.053 for the torus and L63-based systems, respectively. Representative eigenfunctions $\hat{\phi}_j$, plotted as time series $n \mapsto \hat{\phi}_j(x_n)$, 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 17, at large numbers of delays (here, $Q = 2000$), the eigenfunctions $\hat{\phi}_j$ of $\hat{P}_{Q,N}$ should form doubly degenerate pairs, and each pair should exhibit a single frequency associated with an eigenvalue of V. More precisely, $\hat{\phi}_j \pm i\hat{\phi}_{j+1}$
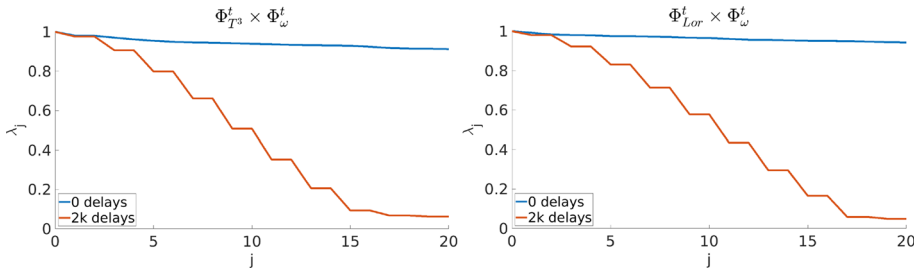
Fig. 3 Eigenvalues $\hat{\lambda}_j = \lambda_{j,Q,N}$ of the integral operator $P_{Q,N}$ for representative values of the delay parameter Q for the torus system in (39) (left) and the L63-based system in (40) (right). The blue and red lines correspond to no delays ($Q = 1$) and 2000 delays, respectively. When $Q = 1$, 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 17

with $j \in \{1, 3, \ldots\}$ should approximate 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 $j\omega$, $j \in \mathbb{N}$. This is evidently the case in the time series plots in Fig. 1. Also, each of the $\hat{\phi}_j$ 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 (35)) in the 51-dimensional data-driven subspace spanned by $\hat{\phi}_0, \ldots, \hat{\phi}_{50}$. The matrix is, to a good approximation, skew-symmetric, consistent with the fact that V is a skew-symmetric operator, and exhibits prominent 2×2 diagonal blocks associated with the eigenspaces of V approximated by $(\hat{\phi}_1, \hat{\phi}_2)$, $(\hat{\phi}_3, \hat{\phi}_4)$, ..., in agreement with Corollary 18.

Figure 4 shows the approximated eigenvalues γ_j of the regularized generator L_θ obtained from this basis using the Galerkin scheme in Definition 27 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 energy $E(z_j)$ of the corresponding eigenfunction z_j (recall that $\mathrm{Re}\,\gamma_j \approx -\theta E_{Q,N}(z_j)$). According to Sect. 6, the imaginary parts of the γ_j 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 $\mathrm{Im}\,\gamma_j$ are indeed equal to integer multiples of $\omega = 1$ to a good approximation for the first ≈ 10 eigenvalues (ordered in order of increasing Dirichlet energy). The accuracy of the eigenvalues begins to deteriorate as the index k approaches m. This is due to the facts that (i) even with a "perfect" basis $\{\hat{\phi}_j\}$, eigenfunctions of higher Dirichlet energy (and stronger oscillatory behavior) require increasingly higher-order Galerkin approximation spaces; (ii) at fixed sample number N, the quality of the data-driven elements $\hat{\phi}_j$ degrades at large j.

8.3 Discussion

The examples presented in Sects. 8.1 and 8.2 are Cartesian products of weak-mixing and pure-point spectrum flows, with their state-space variables combined through some observation map. We begin with some observations about our kernel method applied to Cartesian products.

Cartesian products Let (X, Φ_X^t, μ_X) and (Y, Φ_Y^t, μ_Y) be two ergodic flows with μ_X, μ_Y compact, and with weak-mixing and pure-point spectra, respectively. We are interested in the
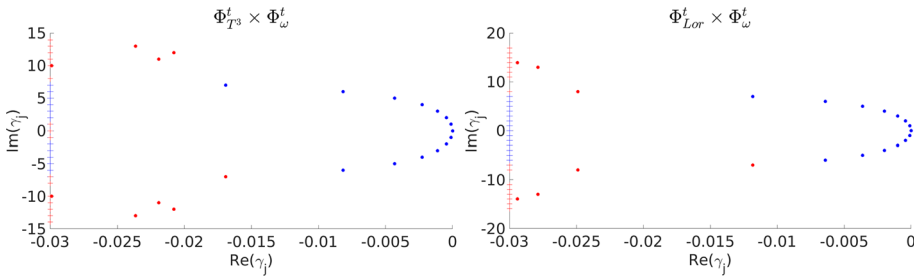
Fig. 4 Galerkin approximations γ_j of the eigenvalues of the regularized generator L_θ for the torus-based system (39) (left) and the L63-based system (40) (right). The numerical eigenvalues were obtained through the data-driven variational eigenvalue problem in Definition 27 with a spectral order parameter $m = 50$. Each plot shows the first ≈ 25 eigenvalues corresponding to eigenfunctions of increasing Dirichlet energy, with the first 14 plotted in blue and the remaining ≈ 10 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

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 $\{f \otimes g : f \in L^2(X, \mu_X), g \in L^2(Y, \mu_Y)\}$. Corollary 28 below characterizes the component $F_{\mathcal{D}}$ of the observation map F in this scenario.

Corollary 28 *Let (X, Φ_X^t, μ_X) and (Y, Φ_Y^t, μ_Y) be as described above, and $F \in L^2(X \times Y, \mu_X \times \mu_Y)$ be decomposed as the sum $F = \sum_{n=1}^{\infty} f_n \otimes g_n$. Then,*

(i) $F_{\mathcal{D}} = \sum_{n=1}^{\infty} \mathbb{E}(f_n)g_n$, *where $\mathbb{E}(f_n) = \int_X f_n \, d\mu_X$. Hence, a necessary and sufficient condition that P is not trivial is that $\mathbb{E}(f_n) \neq 0$ for at least one $n \in \mathbb{N}$.*

(ii) *If F is a continuous map, then $F_{\mathcal{D}}$ is a continuous map given pointwise by $F_{\mathcal{D}}(x, y) = \int_X F(x, y) d\mu_X(x)$.*

Claim (i) is a direct consequence of Proposition 12, 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. The proof of Claim (ii) is also direct and is omitted for brevity. It shows that if the eigenfunctions of (X, Φ_X^t, μ_X) are continuous functions, then the product system satisfies Assumption 3. Note that Corollary 28 also applies to all dynamical systems which are topologically conjugate to such product systems.

Kernels with a small number of delays An implicit assumption in the approximation of the operator P in (17) by the operator P_Q in (17) with finitely many delays Q, is that Q is large-enough for the asymptotic analysis of Lemma 19 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 $i = j$. As a result 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 (39) and (40) cluster near 1 for $Q = 1$ and decay more rapidly for $Q = 2000$.

Weak-mixing systems An important assumption of our approach is that the dynamics has non-zero Koopman eigenfrequencies, i.e., $\mathcal{D}$ contains non-constant functions. This underlies the ability of our regularized operator L_θ in (21) to be a suitable substitute of V (Theorem 21). In fact, by Theorem 3, in the limit of infinitely many delays, $Q \to \infty$, if $\mathcal{D}$ only contains
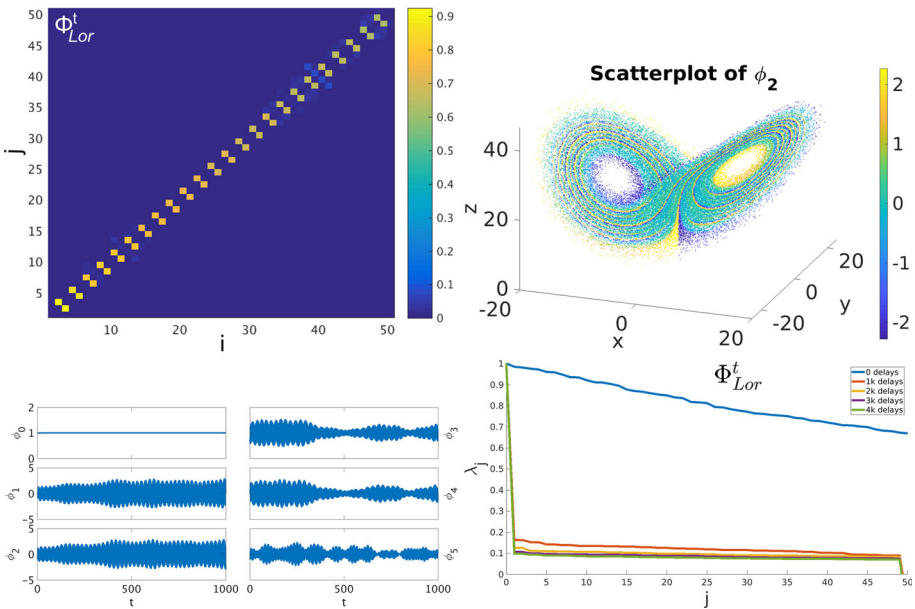
Fig. 5 Eigenvalues $\hat{\lambda}_j = \lambda_{j,Q,N}$ and eigenfunctions $\hat{\phi}_j = \phi_{j,Q,N}$ of $P_{Q,N}$, and absolute values of a matrix representation of the generator of the L63 system in (38) obtained with $Q = 4000$ delays. The generator of this system has continuous spectrum and a trivial eigenvalue at 0. As a result, according to Theorem 3, as $Q \rightarrow \infty$ 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 ≈ 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 the matrix representation $V_{i,j} = \langle \hat{\phi}_i, V_{\Delta t, N} \hat{\phi}_j \rangle_{\mu_n}$ of the approximate generator $V_{\Delta t, N}$ 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

constant functions, then the kernels k_Q and p_Q converge to constants (in the L^2 sense). However, when using finitely many delays, k_Q and p_Q are generally non-constant. It is not currently understood how the operator P_Q should behave for weak-mixing systems and $Q < \infty$.

One of the consequences of Theorem 3(ii) is that in the limit of $Q \rightarrow \infty$, the continuous spectrum subspace $\mathcal{D}^{\perp}$ is annihilated by the integral operator P_{∞}, thus rendering this operator ineffective for studying or reconstructing the weak mixing component of the dynamics. In particular, for weak-mixing systems, P_{∞} should have all but one of its eigenvalues to equal to zero. 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 ($\hat{\lambda}_j \approx 0.1$). This behavior is in agreement with Theorem 3, according to which all the eigenvalues of P_Q other than 1 should converge to zero as $Q \rightarrow \infty$. Note that the matrix representation of V (also shown in Fig. 5) is still skew-symmetric to a good approximation, since V is a skew-symmetric operator. Intriguingly, the matrix has a 2×2 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. [14], 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 $\hat{\phi}_j$, which 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.

In light of the above, the results established in this work have implications for delay-embedding techniques, as they point to a tradeoff between reconstruction of the system's state space topology in delay embedding space (favored by large numbers of delays) and the ability of operators for data analysis, such as P_Q, to adequately represent the mixing component of the dynamics. Nevertheless, the ability to consistently approximate the quasiperiodic dynamics through Koopman eigenfunctions is still useful, as it allows identification and efficient modeling [e.g., via (9)] of observables with high predictability. At the very least, the "negative" result described above provides a reference point that may aid the design of delay-embedding methodologies aiming to reconstruct the full structure of the dynamics. One of the goals of our future work is to investigate the behavior of the techniques presented here away from the asymptotic limit $Q \to \infty$ in the presence of a continuous spectrum.

References

1. Anosov, D.V., Katok, A.B.: New examples in smooth ergodic theory: ergodic diffeomorphisms. Trans. Mosc. Math. Soc. **23**, 1–35 (1970)
2. Ahues, M., Largillier, A., Limaye, B.: Spectral computations for bounded operators. Chapman and Hall/CRC, Boca Raton (2001)
3. Arbabi, H., Mezić, I.: Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. SIAM J. Appl. Dyn. Sys. **16**(4), 2096–2126 (2017). https://doi.org/10.1137/17M1125236
4. Aubry, N., Guyonnet, R., Lima, R.: Spatiotemporal analysis of complex signals: theory and applications. J. Stat. Phys. **64**, 683–739 (1991). https://doi.org/10.1007/bf01048312
5. Babuška, I., Osborn, J.: Eigenvalue Problems, Handbook of Numerical Analysis, vol. 2. North Holland, Amsterdam (1991)
6. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**, 1373–1396 (2003). https://doi.org/10.1162/089976603321780317
7. Belkin, M., Niyogi, P.: Convergence of Laplacian eigenmaps. In: Advances in Neural Information Processing Systems, pp. 129–136 (2007). http://papers.nips.cc/paper/2989-convergence-of-laplacian-eigenmaps.pdf
8. Berry, T., Harlim, J.: Variable bandwidth diffusion kernels. Appl. Comput. Harmon. Anal. (2015). https://doi.org/10.1016/j.acha.2015.01.001
9. Berry, T., Sauer, T.: Consistent manifold representation for topological data analysis (2016). https://arxiv.org/pdf/1606.02353.pdf
10. Berry, T., Sauer, T.: Local kernels and the geometric structure of data. Appl. Comput. Harmon. Anal. **40**, 439–469 (2016). https://doi.org/10.1016/j.acha.2015.03.002
11. Berry, T., Cressman, R., Gregurić-Ferenček, Z., Sauer, T.: Time-scale separation from diffusion-mapped delay coordinates. SIAM J. Appl. Dyn. Sys. **12**, 618–649 (2013). https://doi.org/10.1137/12088183x
12. Berry, T., Giannakis, D., Harlim, J.: Nonparametric forecasting of low-dimensional dynamical systems. Phys. Rev. E **91**, 032,915 (2015). https://doi.org/10.1103/PhysRevE.91.032915
13. Broomhead, D.S., King, G.P.: Extracting qualitative dynamics from experimental data. Phys. D **20**(2–3), 217–236 (1986). https://doi.org/10.1016/0167-2789(86)90031-x
14. Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N.: Chaos as an intermittently forced linear system. Nat. Commun. **8**(19) (2017). https://doi.org/10.1038/s41467-017-00030-8
15. Budišić, M., Mohr, R., Mezić, I.: Applied Koopmanism. Chaos **22**, 047,510 (2012). https://doi.org/10.1063/1.4772195
16. Coifman, R., Lafon, S.: Diffusion maps. Appl. Comput. Harmon. Anal. **21**, 5–30 (2006). https://doi.org/10.1016/j.acha.2006.04.006

17. Coifman, R., Shkolnisky, Y., Sigworth, F., Singer, A.: Graph Laplacian tomography from unknown random projections. IEEE Trans. Image Process. **17**(10), 1891–1899 (2008). https://doi.org/10.1109/tip.2008.2002305

18. Constantin, P., Foias, C., Nicolaenko, B., Témam, R.: Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations. Springer, New York (1989). https://doi.org/10.1007/978-1-4612-3506-4

19. Constantin, P., Kiselev, A., Ryzhik, L., Zlatoš, A.: Diffusion and mixing in fluid flow. Ann. Math. **168**, 643–674 (2008). https://www.jstor.org/stable/40345422

20. Das, S., Giannakis, D.: Koopman spectra in reproducing kernel Hilbert spaces (2018). https://arxiv.org/pdf/1801.07799.pdf

21. Dellnitz, M., Froyland, G., Sertl, S.: On the isolated spectrum of the Perron-Frobenius operator. Nonlinearity **13**(4), 1171–1188 (2000). https://doi.org/10.1088/0951-7715/13/4/310

22. Dellnitz, M., Junge, O.: On the approximation of complicated dynamical behavior. SIAM J. Numer. Anal. **36**, 491 (1999). https://doi.org/10.1137/S0036142996313002

23. Eisner, T., Farkas, B., Haase, M., Nagel, R.: Operator Theoretic Aspects of Ergodic Theory, Graduate Texts in Mathematics, vol. 272. Springer, Berlin (2015)

24. Fayad, B.: Analytic mixing reparametrizations of irrational flows. Ergod. Theory Dyn. Sys. **22**, 437–468 (2002). https://doi.org/10.1017/s0143385702000214

25. Ferreira, J.C., Menegatto, V.A.: Eigenvalues of integral operators defined by smooth positive definite kernels. Integral Equations Operator Theory **64**(1), 61–81 (2009). https://doi.org/10.1007/s00020-009-1680-3

26. Ferreira, J.C., Menegatto, V.A.: Eigenvalue decay rates for positive integral operators. Ann. Mat. Pura Appl. **192**(6), 1–17 (2013). https://doi.org/10.1007/s10231-012-0256-z

27. Froyland, G., González-Tokman, C., Quas, A.: Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. J. Comput. Dyn. **1**(2), 249–278 (2014). https://doi.org/10.3934/jcd.2014.1.249

28. Genton, M.C.: Classes of kernels for machine learning: a statistics perspective. J. Mach. Learn. Res. **2**, 299–312 (2001)

29. Giannakis, D.: Dynamics-adapted cone kernels. SIAM J. Appl. Dyn. Sys. **14**(2), 556–608 (2015). https://doi.org/10.1137/140954544

30. Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. Appl. Comput. Harmon. Anal. (2017). https://doi.org/10.1016/j.acha.2017.09.001

31. Giannakis, D., Das, S.: Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis (2017). https://arxiv.org/pdf/1706.06450.pdf

32. Giannakis, D., Majda, A.J.: 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)

33. Giannakis, D., Majda, A.J.: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. Proc. Natl. Acad. Sci. **109**(7), 2222–2227 (2012). https://doi.org/10.1073/pnas.1118984109

34. Giannakis, D., Slawinska, J., Zhao, Z.: Spatiotemporal feature extraction with data-driven Koopman operators. J. Mach. Learn. Res. Proc. **44**, 103–115 (2015)

35. Halmos, P.: Lectures on Ergodic Theory, vol. 142. American Mathematical Society, Providence (1956)

36. Holmes, P., Lumley, J.L., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press, Cambridge (1996)

37. Koopman, B.O.: Hamiltonian systems and transformation in Hilbert space. Proc. Natl. Acad. Sci. **17**(5), 315–318 (1931). https://doi.org/10.1073/pnas.17.5.315

38. Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. J. Nonlinear Sci. **28**(2), 687–710 (2018). https://doi.org/10.1007/s00332-017-9423-0

39. Korda, M., Putinar, M., Mezić, I.: Data-Driven Spectral Analysis of the Koopman Operator. Appl. Comput. Harmon. Anal. (2018). https://doi.org/10.1016/j.acha.2018.08.002

40. Krengel, U.: Ergodic Theorems, vol. 6. Walter de Gruyter, Berlin (1985)

41. Law, K., Shukla, A., Stuart, A.M.: Analysis of the 3DVAR filter for the partially observed Lorenz'63 model. Discret. Contin. Dyn. Syst. **34**(3), 1061–10,178 (2013). https://doi.org/10.3934/dcds.2014.34.1061

42. Lian, Z., Liu, P., Lu, K.: SRB measures for a class of partially hyperbolic attractors in Hilbert spaces. J. Differ. Equ. **261**, 1532–1603 (2016). https://doi.org/10.1016/j.jde.2016.04.006

43. Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130–141 (1963). https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

44. Lu, K., Wang, Q., Young, L.S.: Strange attractors for periodically forced parabolic equations. Mem. Am. Math. Soc. **224**(1054), 1–85 (2013). https://doi.org/10.1090/S0065-9266-2012-00669-1
45. Luzzatto, S., Melbourne, I., Paccaut, F.: The Lorenz attractor is mixing. Commun. Math. Phys. **260**(2), 393–401 (2005). https://doi.org/10.1007/s00220-005-1411-9
46. McGuinness, M.J.: The fractal dimension of the Lorenz attractor. Philos. Trans. R. Soc. Lond. Ser. A **262**, 413–458 (1968). https://doi.org/10.1098/rsta.1968.0001
47. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**, 309–325 (2005). https://doi.org/10.1007/s11071-005-2824-x
48. Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior. Physica D **197**, 101–133 (2004). https://doi.org/10.1016/j.physd.2004.06.015
49. Nadkarni, M.G.: The spectral theorem for unitary operators. Springer, Berlin (1998)
50. Packard, N.H., et al.: Geometry from a time series. Phys. Rev. Lett. **45**, 712–716 (1980). https://doi.org/10.1103/physrevlett.45.712
51. Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S.: Spectral analysis of nonlinear flows. J. Fluid Mech. **641**, 115–127 (2009). https://doi.org/10.1017/s0022112009992059
52. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. J. Stat. Phys. **65**(3–4), 579–616 (1991). https://doi.org/10.1007/bf01053745
53. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010). https://doi.org/10.1017/S0022112010001217
54. Schmid, P.J., Sesterhenn, J.L.: Dynamic mode decomposition of numerical and experimental data. In: Bulletin of American Physical Society (BAPS), 61st APS Meeting, p. 208. San Antonio (2008)
55. Scholkopf, B., Smola, A., Mu, K.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**, 1299–1319 (1998). https://doi.org/10.1162/089976698300017467
56. Slawinska, J., Giannakis, D.: Indo-Pacific variability on seasonal to multidecadal time scales. Part I: intrinsic SST modes in models and observations. J. Climate **30**, 5265–5294 (2017). https://doi.org/10.1175/JCLI-D-16-0176.1
57. Stone, M.H.: On one-parameter unitary groups in Hilbert space. Ann. Math. **33**, 643–648 (1932). https://doi.org/10.2307/1968538
58. Trillos, N., Slepčev, D.: A variational approach to the consistency of spectral clustering. Appl. Comput. Harmon. Anal. **45**(2), 239–281 (2018). https://doi.org/10.1016/j.acha.2016.09.003
59. Tu, J.H., Rowley, C.W., Lucthenburg, C.M., Brunton, S.L., Kutz, J.N.: On dynamic mode decomposition: theory and applications. J. Comput. Dyn. **1**(2), 391–421 (2014). https://doi.org/10.3934/jcd.2014.1.391
60. Tucker, W.: The Lorenz attractor exists. C. R. Acad. Sci. Paris Ser. I **328**, 1197–1202 (1999)
61. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D **35**, 395–424 (1989). https://doi.org/10.1016/0167-2789(89)90077-8
62. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. Ann. Stat. **26**(2), 555–586 (2008). https://doi.org/10.1214/009053607000000640
63. Wang, C., Deser, C., Yu, J.Y., DiNezio, P., Clement, A.: El Niño and Southern Oscillation (ENSO): a review. In: P.W. Glynn, D.P. Manzello, I.C. Enoch (eds.) Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment, *Coral Reefs of the World*, vol. 8, pp. 85–106. Springer Netherlands, Dordrecht (2017). https://doi.org/10.1007/978-94-017-7499-4_4
64. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. (2015). https://doi.org/10.1007/s00332-015-9258-5
65. Young, L.S.: What are SRB measures, and which dynamical systems have them? J. Stat. Phys. **108**, 733–754 (2002). https://doi.org/10.1023/A:1019762724717
66. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. Adv. Neural Inf. Process. Syst. **17**, 1601–1608 (2004)