

RESEARCH



Delay-coordinate maps, coherence, and approximate spectra of evolution operators

Dimitrios Giannakis* 

*Correspondence:
dimitris@cims.nyu.edu
Department of Mathematics and
Center for Atmosphere Ocean
Science, Courant Institute of
Mathematical Sciences, New
York University, 251 Mercer St,
New York, NY 10012, USA

This paper is dedicated to
Andrew Majda on the occasion
of his 70th birthday.

Abstract

The problem of data-driven identification of coherent observables of measure-preserving, ergodic dynamical systems is studied using kernel integral operator techniques. An approach is proposed whereby complex-valued observables with approximately cyclical behavior are constructed from a pair of eigenfunctions of integral operators built from delay-coordinate mapped data. It is shown that these observables are ε -approximate eigenfunctions of the Koopman evolution operator of the system, with a bound ε controlled by the length of the delay-embedding window, the evolution time, and appropriate spectral gap parameters. In particular, ε can be made arbitrarily small as the embedding window increases so long as the corresponding eigenvalues remain sufficiently isolated in the spectrum of the integral operator. It is also shown that the time-autocorrelation functions of such observables are ε -approximate Koopman eigenvalues, exhibiting a well-defined characteristic oscillatory frequency (estimated using the Koopman generator) and a slowly decaying modulating envelope. The results hold for measure-preserving, ergodic dynamical systems of arbitrary spectral character, including mixing systems with continuous spectrum and no non-constant Koopman eigenfunctions in L^2 . Numerical examples reveal a coherent observable of the Lorenz 63 system whose autocorrelation function remains above 0.5 in modulus over approximately 10 Lyapunov timescales.

Keywords: Kernel integral operators, Delay-coordinate maps, Koopman operators, Feature extraction, Ergodic dynamical systems

1 Introduction

1.1 Background

In the papers [30, 32, 33], A. J. Majda and the author proposed a decomposition technique for multivariate time series, called nonlinear Laplacian spectral analysis (NLSA), combining aspects of delay-coordinate maps of dynamical systems with kernel methods for machine learning. NLSA treats the sampled time series as an observable of a dynamical system and embeds it into a higher-dimensional space using Takens' method of delays [50, 53, 59]. Nonlinear features (principal components) are then extracted as eigenvectors of a normalized kernel matrix constructed from the delay-embedded data, adopting the

perspective of geometrical learning techniques such as Laplacian eigenmaps [7] and diffusion maps [16]. One of the principal empirical findings in [30,32,33] was that the leading NLSA modes exhibit a coherent temporal evolution, capturing distinct timescales from multiscale input data. Examples include systems of ordinary differential equations with metastable regime behavior [32], as well as simulated and observed climate data [31,58].

Meanwhile, in independent work [8], Berry et al. developed an analysis technique called diffusion-mapped delay coordinates (DMDC) which is based on a related delay-coordinate kernel construction, and gave a theoretical interpretation of the timescale separation capability of the DMDC modes using the Oseledets multiplicative ergodic theorem and Lyapunov metrics of dynamical systems. In particular, they showed that under smoothness and hyperbolicity assumptions on the dynamics, and for an appropriately weighted delay-embedding scheme, as the number of delays increases, the leading eigenfunctions recovered through the diffusion maps algorithm vary predominantly along the Oseledets subspace associated with the most stable Lyapunov exponent of the system. They then argued that the evolution of these eigenfunctions, viewed as reduced coordinates for the system state, can be well modeled as a gradient flow driven by a non-autonomous perturbation from the remaining degrees of freedom. In this picture, diffusion maps captures the leading eigenfunctions of the generator of a stochastic process, exhibiting distinct timescales associated with the corresponding eigenvalues.

Besides DMDC and NLSA, several other feature extraction techniques utilizing delay-coordinate maps have been proposed, including early methods such as singular spectrum analysis (SSA) [11,63] and more recent techniques where connections with operator-theoretic ergodic theory have been emphasized [2,12,49]. While NLSA and DMDC differ from these methods in the use of nonlinear kernels (which allow recovery of nonlinear features), the general consensus stemming from this body of the literature is that incorporating delays in feature extraction methodologies facilitates the recovery of dynamically relevant, coherent patterns. Note that this property is distinct from topological state space reconstruction from partial observations (which was the original purpose of delay-coordinate maps [50]) and can be beneficial even under fully observed scenarios. Techniques for coherent feature extraction blending aspects of geometrical integral operators and evolution operators have also received significant attention in the context of non-autonomous dynamical systems [6,27,37].

In [19,29], an interpretation of the timescale separation seen in features recovered from delay-coordinate-mapped data was given through a spectral analysis of kernel integral operators and Koopman evolution operators of dynamical systems [5,26,38]. Specifically, it was shown that for a measure-preserving ergodic dynamical system, as the number of delays increases, the commutator between kernel integral operators constructed from delay-embedded data (subject to mild requirements) and the Koopman operator converges to zero in operator norm, meaning that these operators acquire common eigenspaces in the infinite-delay limit. Since (i) kernel integral operators associated with sufficiently regular (e.g., continuous) kernels are compact and thus have finite-dimensional eigenspaces corresponding to nonzero eigenvalues and (ii) the eigenspaces of Koopman operators of ergodic dynamical systems are one-dimensional, it follows that in the infinite-delay limit, the eigenspaces of the kernel integral operators employed for feature extraction are finite unions of Koopman eigenspaces. The latter are each characterized by a distinct timescale associated with the corresponding eigenvalue of the generator. In applications,

it is oftentimes observed that the eigenspaces of kernel integral operators with large numbers of delays are numerically two-dimensional, meaning that they are associated with a single pair of Koopman eigenfrequencies of equal modulus and different signs. Sampled along orbits of the dynamics, such kernel eigenfunctions have the structure of pure sinusoids, which can be thought of as exhibiting an “ideal” form of timescale separation.

A useful aspect of the results in [19,29] is that they hold for broad classes of measure-preserving, ergodic systems (including systems with non-smooth attractors) and choices of kernel. Thus, they provide relevant information about the asymptotic behavior of a variety of feature extraction techniques utilizing delays, including the methods [2,8,11,12,30,32,33,63] outlined above. Importantly, the integral operators employed can be consistently approximated in a spectral sense from time series data using well-developed theory [44,60,61].

1.2 Motivation and contributions of this work

Despite their generally broad applicability, the results in [19,29] offer limited insight on the behavior of kernel-based feature extraction techniques utilizing delay-coordinate maps for an important class of dynamical systems, namely systems with mixing behavior (or so-called mixed-spectrum systems with both quasiperiodic and mixing components). Indeed, a necessary and sufficient condition for a measure-preserving dynamical system to be mixing is that the generator on the L^2 space associated with the invariant measure has a simple eigenvalue at zero, with a constant corresponding eigenfunction, and no other eigenvalues. As a prototypical example, consider the Lorenz 63 (L63) system [43] on $\mathbb{R}^3$, which is rigorously known to possess an ergodic invariant measure μ supported on the famous “butterfly” attractor with mixing dynamics [45,62]. According to [19], for such a system the kernel integral operator in the infinite-delay limit acquires an infinite-dimensional nullspace containing all $L^2(\mu)$ observables orthogonal to the constant, allowing features with arbitrarily broad frequency spectra (i.e., no timescale separation or coherence). Moreover, data-driven spectral approximation results such as [44,60,61] do not hold for the potentially infinite-dimensional nullspaces of compact operators.

Yet, as illustrated in Fig. 1, the eigenfunctions of integral operators based on a sufficiently long delay-embedding window, T , exhibit a form of coherence, which can be thought of as a relaxation of the periodic behavior of Koopman eigenfunctions. In particular, for sufficiently large T , the time series associated with the kernel eigenfunctions near the top of the spectrum have the structure of amplitude-modulated waves, with a well-defined carrier frequency and a low-frequency modulating envelope. In effect, the pure sinusoids generated by Koopman eigenfunctions can be thought of as special cases of these patterns with constant modulating envelopes. A similar behavior was observed in [54], who found that with increasing number of delays, NLSA provides increasingly coherent representations of the El Niño Southern Oscillation of the climate system, as well as other patterns of climate variability.

The main contribution of this work is to provide a characterization of the coherence properties of eigenfunctions of integral operators constructed from delay-embedded observables of measure-preserving, ergodic dynamical systems of arbitrary (quasiperiodic, mixing, or mixed-spectrum) spectral characteristics, underpinning the behavior in Fig. 1. We will do so by studying a class of complex-valued observables z , whose real and imag-

positivity conditions that depend on the observation map and the form of the kernel, consistent with the results of [19]. On the other hand, for systems with mixing dynamics, the behavior of ε , and thus the coherence of z , is influenced by an interplay between the delay-embedding window length (promoting coherence) and the decay of the eigenvalues of K_T with increasing T (inhibiting coherence). Nevertheless, it is possible that ε is made small by increasing T , so long as the eigenvalues associated with z remain sufficiently isolated in the spectrum of K_T .

The plan of this paper is as follows. In Sect. 2, we describe the class of dynamical systems under study and state our results, including Theorem 1 which is the main theoretical contribution of this work. Section 3 contains a proof of Theorem 1, and Sect. 4 describes the data-driven formulation of our framework. We illustrate our results with numerical examples for the L63 system in Sect. 5 and state our conclusions in Sect. 6. Auxiliary results and definitions on spectral approximation of integral operators are collected in Appendix A.

2 Main results

2.1 Dynamical system under study

Consider a continuous-time, continuous dynamical flow $\Phi^t : \Omega \rightarrow \Omega$, $t \in \mathbb{R}$, on a metric space Ω possessing an invariant, ergodic Borel probability measure μ , supported on a compact set $X \subseteq \Omega$. We assume that the support X of the invariant measure is contained in a forward-invariant, C^1 compact manifold M such that $\Phi^t|_M$ is C^1 , but do not require that X has differentiable structure. The system is observed through a continuous function $F : \Omega \rightarrow Y$, where Y is a Banach space, and the restriction of F to M is C^1 .

This setup encompasses a large class of autonomous dynamical systems encountered in applications. For instance, as a prototypical ODE example with quasiperiodic behavior, one can consider an ergodic rotation $\Phi^t : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ on the 2-torus, in which case $\Omega = M = X = \mathbb{T}^2$ and μ is the Haar measure. The L63 system from Fig. 1 is an example of a smooth dissipative flow on $\Omega = \mathbb{R}^3$, with a rigorously known mixing attractor $X \subset \Omega$ [45, 62] and compact absorbing balls $M \supset X$ [42]. The assumptions stated above also hold for classes of dissipative PDE models possessing inertial manifolds [17].

Within this class of models, our goal is as follows: Given time-ordered data $y_0, y_1, \dots, y_{N-1} \in Y$ with $y_n = F(x_n)$, sampled along a dynamical trajectory $x_n = \Phi^{n\Delta t}(x_0)$ at an interval $\Delta t > 0$, identify a collection of functions $\zeta_j : \Omega \rightarrow \mathbb{C}$ which evolve coherently under the dynamics. Intuitively, by that we mean that the dynamically evolved functions $\zeta_j \circ \Phi^t$ should be relatable to ζ_j in a natural way for t lying in a “large” interval containing zero. From the perspective of learning theory, the functions ζ_j are principal components/features, which are to be identified through an unsupervised learning problem that favors coherence. Note that this objective differs significantly from the classical proper orthogonal decomposition (POD) [4, 36, 41], whose goal is to extract features on the basis of explained variance. Once identified, such coherent features are useful in a variety of contexts, including dimension reduction of high-dimensional time series and predictive modeling [1, 14]. In these approaches, a basic premise is that features related to the spectrum of the underlying dynamical system should reveal physically meaningful dynamical processes (e.g., fundamental oscillations of the climate system [54, 58]), while having favorable predictability properties.

2.2 Pseudospectral criteria for coherence

To establish a mathematically precise notion of dynamical coherence of observables, consider the evolution group of unitary Koopman operators $U^t : L^2(\mu) \rightarrow L^2(\mu)$, acting on observables by composition with the flow, $U^t f = f \circ \Phi^t$ [5, 26, 38, 39]. By Stone's theorem on one-parameter unitary groups [57], the group $\{U_t\}_{t \in \mathbb{R}}$ is generated by a skew-adjoint operator $V : D(V) \rightarrow L^2(\mu)$ with a dense domain $D(V) \subset L^2(\mu)$. As an operator, V corresponds to an extension of the directional derivative on $C^1(M)$ functions associated with the vector field $\tilde{V}$ generating Φ^t , namely $\mathcal{V}f := \tilde{V} \cdot \nabla f$. In particular, for any $f \in D(V)$, $t \mapsto U^t f$ is continuously differentiable in $L^2(\mu)$ and

$$\frac{d}{dt} U^t f = V U^t f = U^t \mathcal{V}f. \quad (1)$$

It is a standard result from ergodic theory [26] that whenever V possesses an eigenfunction $z \in L^\infty(\mu)$ with $\|z\|_{L^2(\mu)} = 1$ and corresponding eigenvalue $i\omega$ (where the eigenfrequency ω is real by skew-adjointness of V), then $|z(x)| = 1$ for μ -a.e. $x \in \Omega$. Thus, we have the periodic evolution

$$U^t z = e^{tV} z = e^{i\omega t} z, \quad (2)$$

and at least measure theoretically, $U^t z$ can be considered to take values on the unit circle. This means, in particular, that for μ -a.e. $x \in \Omega$, the time series $t \mapsto z(\Phi^t(x)) = e^{i\omega t} z(x)$ behaves as a Fourier function on $\mathbb{R}$ with frequency ω . Due to these facts, we think of Koopman eigenfunctions of measure-preserving ergodic dynamical systems as exhibiting an “ideal” form of coherence. Indeed, starting from work in the late 1990s on data-driven spectral analysis of Koopman operators [47, 48] and the related transfer operators [23, 24] spectral decomposition of evolution operators has emerged as a popular approach for coherent feature extraction in dynamical systems.

Yet, despite their attractive properties, Koopman eigenfunctions in $L^2(\mu)$ are not an appropriate theoretical paradigm for coherent features of dynamical systems with complex (mixing) behavior. Indeed, a necessary and sufficient condition for a measure-preserving, ergodic flow to be mixing is that the generator V on $L^2(\mu)$ has a simple eigenvalue 0, with a constant corresponding eigenfunction, and no other eigenvalues. Thus, in this case Koopman eigenfunctions only yield the trivial (constant) feature.

Systems with so-called mixed spectra exhibit an intermediate behavior, in the sense that they do exhibit non-constant eigenfunctions satisfying (2), but these eigenfunctions span only a strict subspace of $L^2(\mu)$ and provide no information about the mixing component of the dynamics. Specifically, it is a classical result [35] that $L^2(\mu)$ admits an orthogonal decomposition

$$L^2(\mu) = H_p \oplus H_c \quad (3)$$

into closed, U^t -invariant subspaces H_p and H_c , such that every observable in H_p is a linear combination of Koopman eigenfunctions (and thus exhibits a quasiperiodic evolution associated with the point spectrum of the generator), whereas $H_c = H_p^\perp$ is a subspace orthogonal to every Koopman eigenfunction and thus associated with the continuous spectrum of the generator. In particular, every observable $g \in H_c$ exhibits a form of mixing behavior (called weak-mixing) characterized by a loss of cross-correlation with any observable $f \in L^2(\mu)$, viz.

$$\lim_{t \rightarrow \infty} C_{fg}(t) = 0, \quad \text{where} \quad C_{fg}(t) := \frac{1}{t} \int_0^t |\langle f, U^s g \rangle| ds. \quad (4)$$

Here, $\langle \cdot, \cdot \rangle$ denotes the $L^2(\mu)$ inner product, $\langle f, g \rangle = \int_{\Omega} f^* g \, d\mu$, taken conjugate linear in the first argument. The issue with feature extraction by pure Koopman eigenfunctions is that the recovered features cannot capture observables in H_c and their mixing behavior.

Here, as a natural relaxation of (2), we seek observables satisfying the Koopman eigenvalue equation in an approximate sense. Specifically, we seek nonzero observables $z \in L^2(\mu)$ satisfying

$$\|U^t z - e^{i\omega t} z\|_{L^2(\mu)} \leq \varepsilon \|z\|_{L^2(\mu)}, \quad (5)$$

for some $\varepsilon > 0$, $\omega \in \mathbb{R}$. Every such observable z is said to be an ε -approximate eigenfunction of U^t , and the complex number $e^{i\omega t}$ is said to lie in the ε -approximate point spectrum of this operator [13]. In addition, we require that the same bound ε holds for all t in an interval $[0, \tau]$ with $\tau > 0$. Observables satisfying these conditions with $\varepsilon \ll 1$ and $\tau \gg 2\pi/\omega$ then behave to a good approximation as Koopman eigenfunctions of measure-preserving ergodic dynamical systems. Note, in particular, that the eigenfunctions $\phi_{1,T}$ and $\phi_{2,T}$ depicted in Fig. 1b are strongly suggestive of this behavior if they are interpreted as the real and imaginary parts of z , i.e., $z = \phi_{1,T} + i\phi_{2,T}$. In the sequel, we will refer to $(e^{i\omega t}, z)$ satisfying (5) as an ε -approximate eigenpair of U^t . It can be shown that because U^t is a normal operator, $(e^{i\omega t}, z)$ is an eigenpair if and only if it is an ε -approximate eigenpair for every $\varepsilon > 0$.

2.3 Integral operators induced by delay-coordinate maps

Motivated by the delay-embedding techniques described in Sect. 1, we seek observables satisfying (5) through eigenfunctions of integral operators on $L^2(\mu)$ based on delay-coordinate maps. To construct appropriate such operators, consider first the distance-like function $d : \Omega \times \Omega \rightarrow \mathbb{R}_+$ induced by the norm of Y and the observable F ,

$$d(x, x') = \|F(x) - F(x')\|_Y,$$

and for every $T > 0$ define $d_T : \Omega \times \Omega \rightarrow \mathbb{R}$ with

$$d_T^2(x, x') = \frac{1}{T} \int_0^T d^2(\Phi^t(x), \Phi^t(x')) \, dt. \quad (6)$$

The function d_T can be equivalently thought of as being induced from the norm of $Y_T := L^2([0, T]; Y)$ under the continuous-time delay-coordinate mapping $F_T : \Omega \rightarrow Y_T$ with $F_T(x)(t) = F(\Phi^t(x))$; that is,

$$d_T^2(x, x') = \|F_T(x) - F_T(x')\|_{Y_T}^2 / T.$$

By convention, we set $d_0 = d$.

Using d_T and a positive, C^1 , bounded shape function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ with bounded derivative, we then consider the family of symmetric kernel functions $k_T : \Omega \times \Omega \rightarrow \mathbb{R}_+$, such that

$$k_T(x, x') = h(d_T^2(x, x')). \quad (7)$$

As a concrete example, we will nominally work with the choice $h(u) = e^{-u/\sigma^2}$, where σ is a positive bandwidth parameter. This leads to the radial Gaussian kernel $k_T(x, x') = e^{-d_T^2(x, x')/\sigma^2}$, which is a common starting point in manifold learning techniques [7, 16] approximating heat kernels on Riemannian manifolds as $\sigma \rightarrow 0$. While here we do not assume that X has manifold structure, which would allow us to use these results,

it should be noted when $Y = \mathbb{R}^m$, Gaussian kernels have an important property that holds irrespective of the regularity of the support of the sampling distribution of the data; namely, they are strictly positive-definite [56]. See [28] for additional examples of kernels commonly employed in machine learning applications.

Every kernel from (7) induces an integral operator $K_T : L^2(\mu) \rightarrow L^2(\mu)$ such that

$$K_T f = \int_{\Omega} k_T(\cdot, x) f(x) d\mu(x). \quad (8)$$

By symmetry and continuity of k_T and compactness of X , K_T is a self-adjoint, Hilbert–Schmidt integral operator with Hilbert–Schmidt norm equal to $\|k_T\|_{L^2(\mu \times \mu)}$. As a result, there exists an orthonormal basis $\{\phi_{0,T}, \phi_{1,T}, \dots\}$ of $L^2(\mu)$ consisting of eigenfunctions of K_T corresponding to the eigenvalues $\lambda_{0,T} \geq \lambda_{1,T} \geq \dots$. The latter are all real and have finite multiplicity whenever nonzero by compactness of K_T . In addition, by continuous differentiability of k_T and compactness of X , every element of in the range of K_T has a representative in $C^1(M)$. In particular, every eigenfunction $\phi_{j,T}$ with nonzero corresponding eigenvalue has the continuous representative

$$\varphi_{j,T} = \frac{1}{\lambda_{j,T}} \int_{\Omega} k_T(\cdot, x) \phi_{j,T}(x) d\mu(x), \quad (9)$$

whose restriction on M is C^1 . Note that $\varphi_{j,T}$ is an everywhere-defined function on Ω , as opposed to the left-hand side of (8) which is an $L^2(\mu)$ -element defined only up to sets of μ -measure zero. We let $\sigma_p(K_T) = \{\lambda_{0,T}, \lambda_{1,T}, \dots\}$ denote the point spectrum of K_T .

In the following subsection, we will show that appropriate linear combinations of eigenfunctions $\phi_{j,T}$ are ε -approximate eigenfunctions of the Koopman operator, satisfying (5) for a threshold ε that decreases as T increases, but increases as $\lambda_{j,T}$ decreases. The continuous representatives of these eigenfunctions will then provide the coherent features ζ_j .

Remark 1 In this section, we have opted to work with delay-coordinate maps in continuous time as this will facilitate the derivation of ε -approximate spectral bounds valid for continuous time intervals. We will later pass to the more common discrete-time formulation based on the sampling interval Δt , which will introduce quadrature errors in (6) that vanish as $\Delta t \rightarrow 0$. In addition, aside from the class of radial kernels in (7), our results hold with straightforward modifications to other classes of kernels with $T \rightarrow \infty$ limits in $L^2(\mu \times \mu)$. Examples include the covariance kernels employed by SSA (which can be obtained by polarization of (7) using a linear shape function), Markov-normalized kernels [10,15,16], and variable-bandwidth kernels [9]. It is also possible to replace the kernel family k_T in (7), which is obtained by a application of a fixed shape function to the T -dependent functions d_T^2 , by a family $\tilde{k}_T$ obtained by averaging a fixed continuous kernel function $k : \Omega \times \Omega \mapsto \mathbb{R}$, i.e., $\tilde{k}_T(x, x') = \int_0^t k(\Phi^t(x), \Phi^t(x')) dt / T$. See [19] for further details.

2.4 Dynamically coherent eigenfunctions

According to the theory of delay-coordinate maps, e.g., [25,51,53], for a sufficiently long window, the delay-coordinate map F_T becomes homeomorphic on the compact support X of the invariant measure for a large class of dynamical systems and observation functions F , even if $F|_X$ is not injective. This property has been widely employed in techniques for state space reconstruction [50] and forecasting [52]. Our interest here, however, is not so

much on topological reconstruction, but rather on the effect of delay-coordinate maps on the spectral properties of kernel integral operators on $L^2(\mu)$, irrespective of the injectivity properties of F . To that end, we begin with a proposition that summarizes some of the results on the limiting behavior of operators in the family K_T from (8), reported in [19].

Proposition 1 *As $T \rightarrow \infty$, the following hold:*

- (i) *The distance-like functions d_T converge in $L^2(\mu \times \mu)$ norm to a function d_∞ , which is invariant under the Koopman operator $U^t \otimes U^t$ of the product dynamical system on $\Omega \times \Omega$ for any $t \in \mathbb{R}$. Correspondingly, the kernel functions k_T also converge in $L^2(\mu \times \mu)$ to a $U^t \otimes U^t$ -invariant kernel k_∞ .*
- (ii) *The sequence of operators K_T converges in $L^2(\mu)$ operator norm to the Hilbert–Schmidt integral operator K_∞ associated with k_∞ .*
- (iii) *For every $t \in \mathbb{R}$, K_∞ and the Koopman operator U^t commute.*
- (iv) *The continuous spectrum subspace H_c lies in the nullspace of K_∞ .*

While we refer the reader to [19] for a proof of this proposition, we note here that Claim (i) follows from the fact that with the definition in (6), d_T^2 corresponds to a continuous-time Birkhoff average of the continuous function $d^2 \in C(\Omega \times \Omega)$ under the product dynamical flow $\Phi^t \times \Phi^t$. The existence and $U^t \otimes U^t$ -invariance of d_∞ is then a consequence of the pointwise ergodic theorem. The remaining claims of Proposition 1 can be deduced by the $U^t \otimes U^t$ -invariance of k_∞ . It is also worthwhile noting that, since Φ^t is mixing with respect to μ if and only if $\Phi^t \times \Phi^t$ is ergodic with respect to $\mu \times \mu$, it follows that d_∞ is constant in $L^2(\mu \times \mu)$ sense if and only if the dynamics Φ^t is μ -mixing. In that case, d_∞ is $\mu \times \mu$ -a.e. constant by ergodicity, and thus, K_∞ is a kernel integral operator with constant kernel. This implies that the nullspace of K_∞ consists of all $L^2(\mu)$ functions orthogonal to the constant. The latter comprise precisely the subspace H_c under mixing dynamics, and we conclude that $\ker K_\infty = H_c$. This last relationship is a special case of Proposition 1(iv) for mixing systems.

For our purposes, the main corollaries of Proposition 1, which follow from Claims (iii) and (ii), respectively, in conjunction with compactness of K_T and K_∞ are:

Corollary 1 *Every eigenspace E of K_∞ corresponding to a nonzero eigenvalue is a finite union of Koopman eigenspaces, and the restriction $V|_E$ of the generator is unitarily diagonalizable. It further follows from skew-adjointness of the generator and ergodicity that E is even-dimensional if and only if it is orthogonal to constant functions (i.e., the nullspace of V).*

Corollary 2 *For every nonzero eigenvalue λ_j of K_∞ , the sequence of eigenvalues $\lambda_{j,T}$ of K_T satisfies $\lim_{T \rightarrow \infty} \lambda_{j,T} = \lambda_j$. Moreover, the orthogonal projections onto the corresponding eigenspaces converge in operator norm. Conversely, if a sequence λ_T of eigenvalues of K_T has a $T \rightarrow \infty$ nonzero limit λ_∞ , then λ_∞ is necessarily an eigenvalue of K_∞ .*

Suppose now that E is a two-dimensional eigenspace of K_∞ corresponding to a nonzero eigenvalue λ , where we have suppressed the j subscript for simplicity of notation. Then, by Corollary 1, E is a union of two Koopman eigenspaces orthogonal to $\ker V$. Let also $\{\phi, \psi\}$ be an orthonormal basis of E , where the eigenfunctions ϕ and ψ are real (such a basis can always be found since the kernel k_∞ is real) and $L^2(\mu)$ -orthogonal to the constants. Then,

it follows by skew-adjointness and reality of V that

$$\langle \phi, V\phi \rangle = \langle \psi, V\psi \rangle = 0,$$

whereas

$$\omega := \langle \psi, V\phi \rangle = -\langle \phi, V\psi \rangle$$

is real. In addition, ω is nonzero since E is a V -invariant subspace of $L^2(\mu)$ orthogonal to $\ker V$. Defining $z = (\phi + i\psi)/\sqrt{2}$, we get

$$Vz = \langle \phi, Vz \rangle \phi + \langle \psi, Vz \rangle = -i\omega\phi + \omega\psi = i\omega z,$$

so we conclude that z is a Koopman eigenfunction corresponding to eigenfrequency ω . By construction, this eigenfunction has unit $L^2(\mu)$ norm, so for any $t \in \mathbb{R}$ we have

$$\alpha_t := \langle z, U^t z \rangle = e^{i\omega t},$$

and if we interpret α_t as an instantaneous autocorrelation function for z (cf. the time-averaged cross-correlation in (4)), it follows that we can recover Koopman eigenvalues from the time-autocorrelation functions of the corresponding eigenfunctions. It also follows from the generator equation (1) that ω can be determined from the derivative of the autocorrelation function at 0, $i\omega = \dot{\alpha}_t|_{t=0}$.

Our main result, stated in the form of the following theorem, is essentially a generalization of these basic observations to ε -approximate eigenfunctions of U^t constructed from eigenfunctions of K_T with finite delay-embedding window T :

Theorem 1 *With the assumptions and notation of Sects. 2.1–2.3, let ϕ and ψ be mutually orthogonal, unit-norm, real eigenfunctions of K_T corresponding to eigenvalues λ_T and ν_T , respectively, with $0 < \lambda_T \leq \nu_T$. Assume that λ_T, ν_T are simple if distinct and twofold-degenerate if equal. Define*

$$z = \frac{1}{\sqrt{2}}(\phi + i\psi), \quad \alpha_t = \langle z, U^t z \rangle, \quad \omega = \langle \psi, V\phi \rangle \equiv \frac{1}{i} \langle z, Vz \rangle \equiv \frac{1}{i} \dot{\alpha}_t|_{t=0},$$

where ω is real, and set

$$\gamma_T = \min_{u \in \sigma_p(K_T) \setminus \{\lambda_T, \nu_T\}} \{\min\{|\lambda_T - u|, |\nu_T - u|\}\},$$

$$\delta_T = \frac{1}{\sqrt{2}}(\nu_T - \lambda_T), \quad \tilde{\delta}_T = \frac{\delta_T}{\nu_T}.$$

Then, the following hold for every $t \geq 0$:

- (i) *The autocorrelation function α_t lies in the $\tilde{\varepsilon}_t$ -approximate point spectrum of U^t , and z is a corresponding $\tilde{\varepsilon}_t$ -approximate eigenfunction for the bound*

$$\tilde{\varepsilon}_t = s_t + \sqrt{S_t},$$

where

$$s_t = \frac{1}{\gamma_T} \left(\frac{C_1 t}{T} + 3\delta_T \right), \quad S_t = \frac{C_2 \|\mathcal{V}\| (1 + \tilde{\delta}_T)}{\lambda_T} \int_0^t s_u \, du.$$

Here, $\|\mathcal{V}\|$ is the norm of the dynamical vector field, viewed as a bounded operator $\mathcal{V} : C^1(M) \rightarrow C(M)$, and C_1 and C_2 are constants that depend only on the observation map F . Explicitly, we have

$$C_1 = 2\|h\|_{C^1(\mathbb{R}_+)} \|d^2\|_{C(X \times X)}, \quad C_2 = 2\|h\|_{C^1(\mathbb{R}_+)} \|d^2\|_{C^1(M \times M)}.$$

- (ii) The modulus $|\omega|$ is independent of the choice of real orthonormal basis $\{\phi, \psi\}$ for the eigenspace(s) corresponding to λ_T and ν_T . Moreover, the phase factor $e^{i\omega t}$ is related to the autocorrelation function according to the bound

$$|\alpha_t - e^{i\omega t}| \leq 2\sqrt{S_t}.$$

Note that s_t and S_t in Theorem 1 are increasing functions of $t \geq 0$. This, in conjunction with the fact that $\|U^t z - e^{i\omega t} z\|_{L^2(\mu)} \leq \|U^t z - \alpha_t z\|_{L^2(\mu)} + |\alpha_t - e^{i\omega t}|$, leads to the following corollary, which shows how to attain the bound in (5) uniformly over a bounded time interval.

Corollary 3 *The phase factor $e^{i\omega t}$ lies in the ε_t -approximate point spectrum of U^t , and z is a corresponding ε_t -approximate eigenfunction for the bound*

$$\varepsilon_t = s_t + 3\sqrt{S_t}.$$

Moreover, for every $\tau \geq 0$, $(e^{i\omega t}, z)$ is an ε_τ -approximate eigenpair of U^t for all $t \in [0, \tau]$. This eigenpair has the continuous representative $\zeta \in C(\Omega)$ given by

$$\zeta = \frac{1}{\sqrt{2}} \int_{\Omega} k_T(\cdot, x) \left(\frac{\phi(x)}{\lambda_T} + i \frac{\psi(x)}{\nu_T} \right) d\mu(x),$$

which acts as an everywhere-defined, continuous coherent feature on the state space Ω .

Theorem 1 will be proved in Sect. 3. We now discuss some of the intuitive aspects of the results. First, it should be noted that the bounds established are not sharp, as there are systems for which one can readily construct integral operators K_T with finite embedding windows T and common eigenspaces with the Koopman operator. Examples include operators derived from translation-invariant kernels on tori under quasiperiodic dynamics [20, 29], e.g., the heat kernel associated with the flat metric. For such kernels, there exist eigenfunctions z which are also Koopman eigenfunctions, and the corresponding autocorrelation coefficients α_t lie in the ε -approximate point spectra of U^t for any $\varepsilon > 0$ and $t \in \mathbb{R}$. Still, even without sharp bounds, Theorem 1 provides useful information on the spectral properties of integral operators utilizing delay-coordinate maps that promote or inhibit dynamical coherence, as follows.

1. As one might expect, the bounds in Theorem 1 become weaker as the regularity of the observation map F and kernel shape function h decreases, in the sense that $\tilde{\varepsilon}_t$ and ε_t are increasing functions of the C^1 norms of d^2 and h . It should be noted that many commonly used kernels for feature extraction [7, 9, 10, 16, 28], including the kernels employed in this work, are parameterized by bandwidth parameters controlling the concentration of the kernel about the diagonal (e.g., the parameter σ in (35) ahead). For such kernels, the C^1 norm of h typically increases without bound as the bandwidth parameter decreases.
2. For fixed t , the strength of the bounds is an interplay between the length T of the embedding window, the eigenvalue λ_T , the gap γ_T (measuring the isolation of the eigenspaces corresponding to λ_T and ν_T from the rest of the point spectrum of K_T), and the gaps $\delta_T, \tilde{\delta}_T$ (measuring the extent at which λ_T and ν_T fail to be twofold-degenerate). Inspecting the dependence of the functions s_t and S_t on these terms indicates that, in general, the bounds become stronger as the window length

T increases and/or the gaps $\delta_T, \tilde{\delta}_T$ decrease, whereas they weaken as λ_T and/or the gap γ_T decreases. Of course, these terms cannot be independently controlled as T varies, and the expected coherence of z on the basis of Theorem 1 will depend on their combined effect. It should be noted that Theorem 1 does not make an assertion about existence of $T \rightarrow \infty$ limits for the ε -approximate eigenpairs $(e^{i\omega t}, z)$, although as we discuss below, there are particular cases for which such limits exist.

3. Suppose that the eigenvalue sequence λ_T has a nonzero $T \rightarrow \infty$ limit λ_∞ . Then, by Proposition 1, λ_∞ is a nonzero eigenvalue of the compact operator K_∞ . By the same proposition, if the eigenspace E corresponding to λ_∞ does not contain constant functions it is even-dimensional, so the gap coefficients δ_T and $\tilde{\delta}_T$ converge to 0. If, further, E is two-dimensional, the gap γ_T converges to a nonzero value. In such cases, Theorem 1 and Corollary 3 imply that for any $\tau \geq 0$ and $\varepsilon > 0$, there exists $T_* > 0$ such that for all $T > T_*$, (5) holds for all $t \in [0, \tau]$. This implies in turn that for such a sequence λ_T , there is a subsequence of frequencies ω converging to an eigenfrequency of the generator (where we consider a subsequence to account for possible sign flips due the choice of functions ϕ and ψ at each T). Moreover, the corresponding observables z similarly approximate Koopman eigenfunctions.
4. Suppose now that the dynamics is mixing with respect to the invariant measure μ . Then, all eigenvalues λ_T with non-constant corresponding eigenfunctions converge to 0 as $T \rightarrow \infty$, and therefore, the gaps γ_T, δ_T , and $\tilde{\delta}_T$ also converge to 0. In that case, the asymptotic behavior of ε_t as $T \rightarrow \infty$ depends on the behavior of

$$\eta_T := \gamma_T \lambda_T T, \quad (10)$$

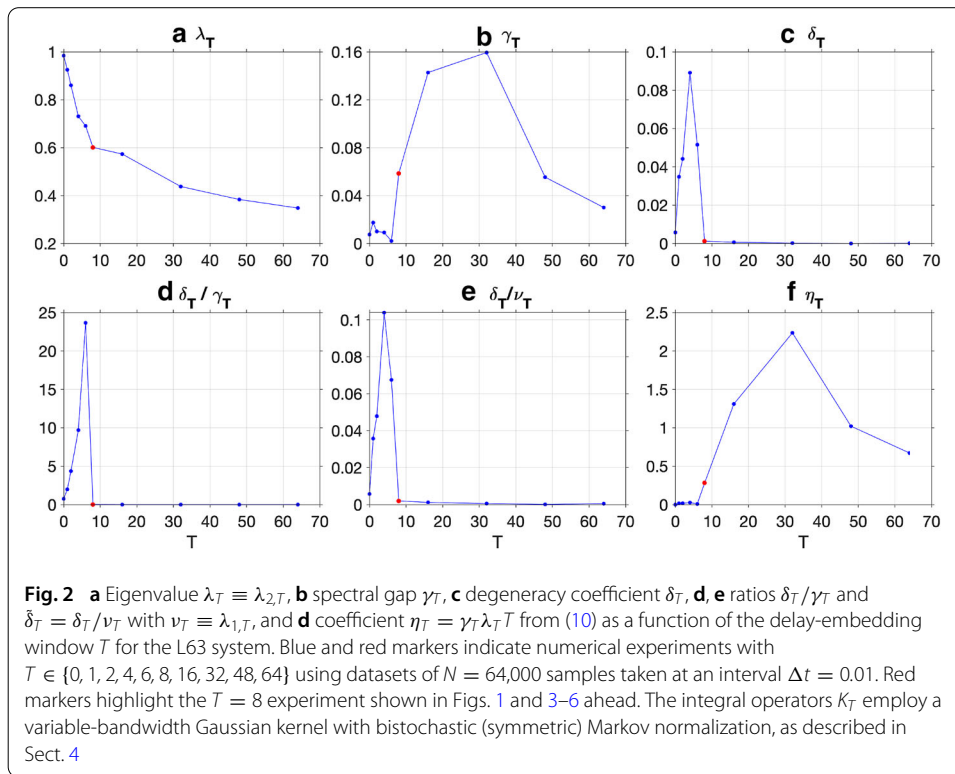
as well as the ratios δ_T/γ_T and $\tilde{\delta}_T \equiv \delta_T/\nu_T$, on the chosen eigenvalue sequences λ_T and ν_T . If η_T converges to 0 as $T \rightarrow \infty$, then ε_t diverges in that limit for any $t > 0$, failing to provide a useful bound. However, the possibility still remains that the rate of decay of γ_T and λ_T is slow enough such that η_T attains large values over a suitable range of T , allowing ε_t to remain small on a large interval $[0, \tau] \ni t$ (so long as γ_T/δ_T and $\tilde{\delta}_T$ are also small). In Fig. 2, numerical η_T values for the L63 system are found to lie above the value corresponding to the $T = 8$ results in Fig. 1 out to at least $\simeq 70$ Lyapunov times, before eventually decaying. In addition, δ_T/γ_T and $\tilde{\delta}_T$ are also small after initial transients have died out. Together, these results demonstrate that the bounds from Theorem 1 are practically relevant for a wide range of delay-embedding windows for the L63 system. An intriguing question (lying outside the scope of this work) is whether there are mixing dynamical systems and integral operators for which η_T actually diverges as $T \rightarrow \infty$.

3 Proof of Theorem 1

3.1 Proof of Claim (i)

Noting that z and z^* are mutually orthogonal unit vectors in $L^2(\mu)$, and $U^0 = \text{Id}$, we begin by writing down the expansion

$$U^t z = \alpha_t z + \beta_t z^* + r_t, \quad (11)$$



where $\alpha_t = \langle z, U^t z \rangle$ (as in the statement of the theorem), $\beta_t = \langle z^*, U^t z \rangle$, r_t is a residual orthogonal to both z and z^* , and

$$\begin{aligned} |\alpha_t| &\leq 1, \quad |\beta_t| \leq 1, \quad \|r_t\|_{L^2(\mu)} \leq 1, \\ \alpha_0 &= 1, \quad \beta_0 = \|r_0\|_{L^2(\mu)} = 0. \end{aligned} \quad (12)$$

It then follows that

$$\|U^t z - \alpha_t z\|_{L^2(\mu)} \leq |\beta_t| + \|r_t\|_{L^2(\mu)}, \quad (13)$$

and we will prove the first claim of the theorem by bounding $|\beta_t|$ and $\|r_t\|_{L^2(\mu)}$.

To that end, note first that by skew-symmetry and reality of V , and by definition of the $L^2(\mu)$ inner product,

$$\langle z^*, Vz \rangle = -\langle Vz^*, z \rangle = -\langle (Vz)^*, z \rangle = -\langle Vz, z^* \rangle^* = -\langle z^*, Vz \rangle,$$

so $\langle z^*, Vz \rangle = 0$. Moreover,

$$\langle z, Vz \rangle^* = \langle z^*, (Vz)^* \rangle = \langle z^*, Vz^* \rangle = -\langle Vz^*, z^* \rangle = -\langle z^*, Vz^* \rangle^* = -\langle z, Vz \rangle,$$

so $\langle z, Vz \rangle$ and $\langle z^*, Vz^* \rangle$ are purely imaginary. In fact, it follows from the definition of z that

$$\langle z, Vz \rangle / i = \langle \psi, V\phi \rangle = \omega, \quad (14)$$

and from the definition of the generator that

$$\frac{1}{i} \langle z, Vz \rangle = \lim_{t \rightarrow 0} \frac{1}{it} \langle z, (U^t - \text{Id})z \rangle = \lim_{t \rightarrow 0} \frac{1}{it} (\alpha_t - 1) = \dot{\alpha}_t|_{t=0},$$

so we can use $\langle z, Vz \rangle / i$ and $\dot{\alpha}_t|_{t=0} / i$ as alternative definitions of the frequency ω as in the statement of Theorem 1.

Using these relationships, the generator equation in (1), and the bound for $|\beta_t|$ in (12), we obtain

$$\begin{aligned} \frac{d}{dt} |\beta_t|^2 &= 2 \operatorname{Re} \left(\beta_t^* \frac{d\beta_t}{dt} \right) = 2 \operatorname{Re} \left(\beta_t^* \frac{d}{dt} \langle z^*, U^t z \rangle \right) = 2 \operatorname{Re} (\beta_t^* \langle z^*, V U^t z \rangle) \\ &= -2 \operatorname{Re} (\beta_t^* \langle V z^*, U^t z \rangle) = -2 \operatorname{Re} (\beta_t^* \langle V z^*, \alpha_t z + \beta_t z^* + r_t \rangle) \\ &= -2 \operatorname{Re} (\beta_t^* \langle V z^*, r_t \rangle) \leq 2 |\beta_t| |\langle V z^*, r_t \rangle| \leq 2 \|V z\|_{L^2(\mu)} \|r_t\|_{L^2(\mu)}. \end{aligned}$$

Therefore, the squared modulus $|\beta_t|^2$ is bounded by a solution of the differential inequality

$$\frac{d}{dt} |\beta_t|^2 \leq \|V z\|_{L^2(\mu)} \|r_t\|_{L^2(\mu)}, \quad \beta_0 = 0, \quad (15)$$

where we have used (12) to set the initial conditions. Note that we were able to use the generator equation in order to arrive at this relation since $z \in \operatorname{ran} K_T$, and every element in $\operatorname{ran} K_T$ has a $C^1(M)$ representative and thus lies in the domain of the generator, $D(V)$.

Inspecting (13) and (15) indicates that the norm of the residual $\|r_t\|_{L^2(\mu)}$ bounds $\|U^t z - \alpha_t z\|_{L^2(\mu)}$ both directly, in (13), and indirectly by bounding the rate of growth of $|\beta_t|^2$, in (15). In addition, $\frac{d}{dt} |\beta_t|^2$ depends on the norm $\|V z\|_{L^2(\mu)}$. The following two lemmas are useful for estimating these terms.

Lemma 1 *With the notation and assumptions of Theorem 1, for every $t \geq 0$ and $T > 0$ the commutator $[U^t, K_T]$ satisfies*

$$\|[U^t, K_T]\| \leq \frac{2 \|h\|_{C^1(\mathbb{R}_+)} \|d\|_{C(X \times X)}^2 t}{T},$$

where $\|\cdot\|$ denotes $L^2(\mu)$ operator norm.

Proof The proof follows closely that of Lemma 19 in [19], which established a similar result for discrete-time sampling and $C(X)$ operator norm. In particular, it is a direct consequence of the definition of the delay-coordinate distance d_T in (6) that for any $x, x' \in X$ and $t \geq 0$,

$$\begin{aligned} d_T^2(\Phi^t(x), \Phi^t(x')) &= \frac{1}{T} \int_0^T d^2(\Phi^{t+u}(x), \Phi^{t+u}(x')) du \\ &= d_T^2(x, x') + \frac{1}{T} \left(\int_T^{T+t} du - \int_0^t du \right) d^2(\Phi^u(x), \Phi^u(x')). \end{aligned}$$

Therefore,

$$|d_T^2(\Phi^t(x), \Phi^t(x')) - d_T^2(x, x')| \leq \frac{2 \|d\|_{C(X \times X)}^2 t}{T},$$

and using the above and the definition of the kernel k_T in (7), we get

$$\begin{aligned} |k_T(\Phi^t(x), \Phi^t(x')) - k_T(x, x')| &= |h(k_T(\Phi^t(x), \Phi^t(x')))) - h(k_T(x, x'))| \\ &\leq \|h\|_{C^1(\mathbb{R}_+)} |d_T^2(\Phi^t(x), \Phi^t(x')) - d_T^2(x, x')| \\ &\leq \frac{2\|h\|_{C^1(\mathbb{R}_+)} \|d\|_{C(X \times X)}^2 t}{T}. \end{aligned} \quad (16)$$

It then follows that for any $f \in L^2(\mu)$

$$\begin{aligned} \|U^t K_T f - K_T U^t f\|_{L^2(\mu)} &= \left\| \int_{\Omega} (k_T(\Phi^t(\cdot), x)f(x) - k_T(\cdot, x)f(\Phi^t(x))) \, d\mu(x) \right\|_{L^2(\mu)} \\ &= \left\| \int_{\Omega} (k_T(\Phi^t(\cdot), \Phi^t(x)) - k_T(\cdot, x)) U^t f(x) \, d\mu(x) \right\|_{L^2(\mu)} \\ &\leq \|k_T(\Phi^t(\cdot), \Phi^t(\cdot)) - k_T\|_{C(X \times X)} \|U^t f\|_{L^1(\mu)} \\ &\leq \|k_T(\Phi^t(\cdot), \Phi^t(\cdot)) - k_T\|_{C(X \times X)} \|f\|_{L^2(\mu)}. \end{aligned}$$

Note that to obtain the second and last lines in the displayed equations above, we used the fact that μ is an invariant probability measure under the flow Φ^t . Using this result and (16), we arrive at

$$\|U^t, K_T\| = \|U^t K_T - K_T U^t\| \leq \frac{2\|h\|_{C^1(\mathbb{R}_+)} \|d\|_{C(X \times X)}^2 t}{T},$$

proving the lemma. $\square$

Lemma 2 *With the notation and assumptions of Theorem 1, the family of operators $\{A_T = VK_T \mid T > 0\}$ is uniformly bounded on $L^2(\mu)$ with*

$$\|A_T\| \leq \|\mathcal{V}\| \|h\|_{C^1(\mathbb{R}_+)} \|d^2\|_{C^1(M \times M)}.$$

Proof We use the notation $\mathcal{V}_1 : C^1(M \times M) \rightarrow C(M \times M)$ to represent the differential operator on $C^1(M \times M)$ which acts by the dynamical vector field $\mathcal{V} : C^1(M) \rightarrow C(M)$ along the first coordinate, i.e.,

$$\mathcal{V}_1 f(x, x') = \lim_{t \rightarrow 0} \frac{f(\Phi^t(x), x') - f(x, x')}{t} = \mathcal{V} f_{x'}(x),$$

where $f_{x'} = f(\cdot, x') \in C^1(M)$. Note that $\mathcal{V}_1$ and $\mathcal{V}$ have equal operator norms, $\|\mathcal{V}_1\| = \|\mathcal{V}\|$. Moreover, $\mathcal{V}_1$ commutes with the induced action by the product dynamical flow $\Phi^t \otimes \Phi^t$ on $C^1(M \times M)$, in the sense that

$$\mathcal{V}_1(f \circ \Phi^t \otimes \Phi^t) = (\mathcal{V}_1 f) \circ \Phi^t \otimes \Phi^t, \quad \forall t \geq 0, \quad \forall f \in C^1(M \times M).$$

Using these facts, we obtain

$$\begin{aligned} \|\mathcal{V}_1 d_T^2\|_{C(X \times X)} &= \left\| \frac{1}{T} \int_0^T \mathcal{V}_1(d^2 \circ \Phi^t \otimes \Phi^t) \, dt \right\|_{C(X \times X)} \\ &= \left\| \frac{1}{T} \int_0^T (\mathcal{V}_1 d^2) \circ \Phi^t \otimes \Phi^t \, dt \right\|_{C(X \times X)} \\ &\leq \|\mathcal{V}_1 d^2\|_{C(X \times X)} \leq \|\mathcal{V}_1\| \|d^2\|_{C^1(M \times M)} = \|\mathcal{V}\| \|d^2\|_{C^1(M \times M)}, \end{aligned}$$

and thus,

$$\begin{aligned}\|\mathcal{V}_1 k_T\|_{C(X \times X)} &= \|\mathcal{V}_1(h \circ d_T^2)\|_{C(X \times X)} \leq \|h\|_{C^1(\mathbb{R}_+)} \|\mathcal{V}_1 d_T^2\|_{C(X \times X)} \\ &\leq \|h\|_{C^1(\mathbb{R}_+)} \|\mathcal{V}\| \|d^2\|_{C^1(M \times M)}.\end{aligned}\quad (17)$$

Now, because k_T lies in $C^1(M \times M)$, for every $f \in L^2(\mu)$ we have

$$A_T f = VK_T f = V \int_{\Omega} k_T(\cdot, x) f(x) d\mu(x) = \int_{\Omega} \mathcal{V}_1 k_T(\cdot, x) f(x) d\mu(x),$$

so A_T is a kernel integral operator on $L^2(\mu)$ whose kernel $\mathcal{V}_1 k_T$ is continuous on $X \times X$. The $L^2(\mu)$ operator norm of A_T therefore satisfies

$$\|A_T\| \leq \|\mathcal{V}_1 k_T\|_{C(X \times X)},$$

and the claim of the lemma follows from (17). $\square$

With these results in place, we proceed to bound $\|r_t\|_{L^2(\mu)}$. First, acting with K_T on both sides of (11), we obtain

$$\begin{aligned}K_T U^t z &= \alpha_t K_T z + \beta_t K_T z^* + K_T r_t \\ &= \frac{\alpha_t}{\sqrt{2}}(\lambda_T \phi + i v_T \psi) + \frac{\beta_t}{\sqrt{2}}(\lambda_T \phi - i v_T \psi) + K_T r_t \\ &= \lambda_T(\alpha_t z + \beta_t z^*) + i \delta_T(\alpha_t - \beta_t) \psi + K_T r_t \\ &= \lambda_T U^t z + i \delta_T(\alpha_t - \beta_t) \psi + (K_T - \lambda_T) r_t \\ &= \frac{1}{\sqrt{2}} U^t (K_T \phi + i \lambda_T \psi) + i \delta_T(\alpha_t - \beta_t) \psi + (K_T - \lambda_T) r_t \\ &= \frac{1}{\sqrt{2}} U^t (K_T \phi + i K_T \psi) + i \delta_T(\alpha_t - \beta_t - U^t) \psi + (K_T - \lambda_T) r_t \\ &= U^t K_T z + i \delta_T(\alpha_t - \beta_t - U^t) \psi + (K_T - \lambda_T) r_t.\end{aligned}$$

Therefore,

$$(K_T - \lambda_T) r_t = -[U^t, K_T] z + i \delta_T(U^t - \alpha_t - \beta_t) \psi.$$

which, in conjunction with (12), leads to

$$\|(K_T - \lambda_T) z\|_{L^2(\mu)} \leq \|[U^t, K_T]\| + 3\delta_T. \quad (18)$$

On the other hand,

$$\begin{aligned}\|(K_T - \lambda_T) r_t\|_{L^2(\mu)} &= \sum_{\lambda_{j,T} \in \sigma_p(K_T) \setminus \{\lambda_T, v_T\}} (\lambda_{j,T} - \lambda_T)^2 |\langle \phi_{j,T}, r_t \rangle|^2 \\ &\geq \sum_{\lambda_{j,T} \in \sigma_p(K_T) \setminus \{\lambda_T, v_T\}} \gamma_T^2 |\langle \phi_{j,T}, r_t \rangle|^2 = \gamma_T^2 \|r_t\|_{L^2(\mu)}^2,\end{aligned}\quad (19)$$

and using (18), (19), and Lemma 1, we arrive at the bound

$$\|r_t\|_{L^2(\mu)} \leq \frac{1}{\gamma_T} \left(\frac{2\|h\|_{C^1(\mathbb{R}_+)} \|d\|_{C(X \times X)}^2}{T} + 3\delta_T \right) = s_t, \quad (20)$$

where the function s_t was defined in the statement of Theorem 1.

Next, it follows from Lemma 2 that

$$\begin{aligned}\|Vz\|_{L^2(\mu)} &= \frac{1}{\sqrt{2}} \|V(\phi + i\psi)\|_{L^2(\mu)} = \frac{1}{\sqrt{2}} \left\| VK_T \left(\frac{\phi}{\lambda_T} + i \frac{\psi}{\nu_T} \right) \right\|_{L^2(\mu)} \\ &= \frac{1}{\lambda_T} \left\| A_T \left(z + \frac{i}{\sqrt{2}} \left(\frac{1}{\nu_T} - \frac{1}{\lambda_T} \right) \psi \right) \right\|_{L^2(\mu)} \\ &\leq \frac{\|A_T\|(1 + \tilde{\delta}_T)}{\lambda_T} = \frac{\|\mathcal{V}\| \|h\|_{C^1(\mathbb{R}_+)} \|d^2\|_{C^1(M \times M)} (1 + \tilde{\delta}_T)}{\lambda_T}.\end{aligned}\quad (21)$$

Inserting the estimates for $\|r_t\|_{L^2(\mu)}$ and $\|Vz\|_{L^2(\mu)}$ in (20) and (21), respectively, into (15), and using the definition of the constant C_2 in the statement of the theorem, then leads to the differential inequality

$$\frac{d}{dt} |\beta_t|^2 \leq \frac{C_2 \|\mathcal{V}\| (1 + \tilde{\delta}_T)}{\lambda_T} s_t, \quad \beta_0 = 0,$$

and integrating we obtain

$$|\beta_t|^2 \leq \frac{C_2 \|\mathcal{V}\| (1 + \tilde{\delta}_T)}{\lambda_T} \int_0^t s_u \, du = \frac{2C_2(1 + \tilde{\delta}_T)}{\lambda_T \gamma_T} \left(\frac{C_1 t^2}{T} + 3\delta_T t \right) = S_t, \quad (22)$$

where the function S_t is defined in the statement of the theorem. Substituting (20) and (22) into (13) then leads to $\|U^t z - \alpha_t z\|_{L^2(\mu)} \leq s_t + \sqrt{S_t}$, proving Claim (i) of the theorem.

3.2 Proof of Claim (ii)

First, to verify that ω is independent of the choice of mutually orthonormal basis functions ϕ and ψ , it is sufficient to consider the following two cases:

- Case I: λ_T and ν_T are simple eigenvalues. In this case, the claim is obvious since any unit-norm eigenvectors ϕ' and ψ' corresponding to λ_T and ν_T , respectively, are related to ϕ and ψ by

$$\phi' = c_\phi \phi, \quad \psi' = c_\psi \psi,$$

where $c_\phi, c_\psi \in \{-1, 1\}$.

- Case II: $\lambda_T = \nu_T$ are twofold-degenerate eigenvalues. To verify the claim, let $\{\phi', \psi'\}$ be any real, orthonormal basis of the corresponding eigenspace, E . Then, there exists a 2×2 orthogonal matrix O such that

$$\begin{pmatrix} \phi' \\ \psi' \end{pmatrix} = O \begin{pmatrix} \phi \\ \psi \end{pmatrix}, \quad O = \begin{pmatrix} O_{\phi\phi} & O_{\phi\psi} \\ O_{\psi\phi} & O_{\psi\psi} \end{pmatrix}.$$

Since $\langle \phi, V\phi \rangle = \langle \psi, V\psi \rangle = 0$ (by skew-adjointness and reality of V , in conjunction with reality of ϕ and ψ), we have

$$\begin{aligned}|\langle \psi', V\phi' \rangle| &= |\langle O_{\psi\phi}\phi + O_{\psi\psi}\psi, O_{\phi\phi}V\phi + O_{\phi\psi}V\psi \rangle| \\ &= |(O_{\psi\phi}O_{\phi\psi} - O_{\phi\phi}O_{\psi\psi})\omega| \\ &= |\det O| |\omega| = |\omega|,\end{aligned}$$

proving that $|\omega|$ is independent of the choice of real orthonormal basis of E .

Next, to bound $|\alpha_t - e^{i\omega t}|$, we follow a differential inequality approach similar to that used to bound $|\beta_t|$ in Sect. 3.1. In particular, let $a_t = \alpha_t - e^{i\omega t}$. We have

$$|a_t|^2 = |\alpha_t|^2 + 1 - 2\operatorname{Re}(\alpha_t e^{-i\omega t}),$$

and therefore

$$\frac{d}{dt} |a_t|^2 \leq \frac{d}{dt} |\alpha_t|^2 + 2 \left| \operatorname{Re} \frac{d}{dt} (\alpha_t e^{-i\omega t}) \right|. \quad (23)$$

To place a bound on the first term in the right-hand side of (23), observe that

$$\begin{aligned} \frac{d}{dt} |\alpha_t|^2 &= 2 \operatorname{Re} \left(\alpha_t^* \frac{d\alpha_t}{dt} \right) = 2 \operatorname{Re} (\alpha_t^* \langle z, U^t Vz \rangle) = -2 \operatorname{Re} (\alpha_t^* \langle Vz, U^t z \rangle) \\ &= -2 \operatorname{Re} (\alpha_t^* \langle Vz, \alpha_t z + \beta_t z^* + r_t \rangle) \\ &= -2 \operatorname{Re} (|\alpha_t|^2 \langle Vz, z \rangle + \alpha_t^* \beta_t \langle Vz, z^* \rangle + \alpha_t^* \langle Vz, r_t \rangle) \\ &= -2 \operatorname{Re} (\alpha_t^* \langle Vz, r_t \rangle) \leq 2 |\alpha_t| |\langle Vz, r_t \rangle| \leq 2 |\langle Vz, r_t \rangle|. \end{aligned} \quad (24)$$

Note that to obtain the equality in the second-to-last line, we used the facts that $\langle z^*, Vz \rangle$ and $\langle z, Vz \rangle$ are vanishing and purely imaginary, respectively (see Sect. 3.1). Moreover, we used the bound $|\alpha_t| \leq 1$ in (12) and Lemma 2 to arrive at the inequality in the last line. Similarly, using (14) and the fact that $\langle z^*, Vz \rangle = 0$ leads to

$$\begin{aligned} \left| \operatorname{Re} \frac{d}{dt} (\alpha_t e^{-i\omega t}) \right| &= \left| \operatorname{Re} (\langle Vz, \alpha_t z + \beta_t z^* + r_t \rangle e^{-i\omega t} - i\omega \alpha_t e^{-i\omega t}) \right| \\ &= \left| \operatorname{Re} (\langle Vz, r_t \rangle e^{-i\omega t}) \right| \leq |\langle Vz, r_t \rangle|, \end{aligned} \quad (25)$$

and inserting (24) and (25) into (23), we obtain

$$\frac{d}{dt} |a_t|^2 \leq 4 |\langle Vz, r_t \rangle| = \frac{4C_2 \|\mathcal{V}\| (1 + \tilde{\delta}_T)}{\lambda_T} s_t.$$

Integrating this differential inequality subject to the initial condition $a_0 = \alpha_0 - 1 = 0$ then leads to

$$|\alpha_t - e^{i\omega t}|^2 = |a_t|^2 \leq 4S_t,$$

and the bound in Claim (ii) of Theorem 1 follows. This completes our proof of the theorem.

4 Data-driven approximation

In this section, we consider how to approximate the eigenvalues and eigenfunctions of the integral operator K_T , as well as the frequency ω and autocorrelation function α_t , from the time series data $y_0, \dots, y_{N-1}$ sampled at the interval Δt , as described in Sect. 2.1. Aside from errors associated by approximating continuous-time delay-coordinate maps by (their more familiar) discrete-time analogs, error analyses for the approximation scheme described below have been performed elsewhere [19, 21, 34]. Here, we limit ourselves to a high-level description of the construction and its convergence in the large-data limit, relegating technical details to these references.

4.1 Construction of the data-driven approximation scheme

The main steps in the construction of the approximation scheme are as follows:

Step 1 (Discrete-time delay-coordinate map) Replace the continuous-time delay-coordinate map $F_T : \Omega \rightarrow L^2([0, T]; Y)$ by the discrete-time map $F_{Q, \Delta t} : \Omega \rightarrow Y^Q$ given by

$$F_{Q, \Delta t}(x) = (F(x), F(\Phi^{\Delta t}(x)), \dots, F(\Phi^{(Q-1)\Delta t}(x))). \quad (26)$$

Here, Q is an integer parameter corresponding to the number of delays. The map $F_{Q,\Delta t}$ with $\Delta t = T/Q$ then induces a continuous distance-like function $d_{T,\Delta t} : \Omega \times \Omega \rightarrow \mathbb{R}_+$,

$$d_{T,\Delta t}^2(x, x') = \frac{1}{Q} \|F_{Q,\Delta t}(x) - F_{Q,\Delta t}(x')\|_{Y^Q}^2 = \frac{1}{Q} \sum_{q=0}^{Q-1} d^2(\Phi^{(q-1)\Delta t}(x), \Phi^{(q-1)\Delta t}(x')),$$

which is meant to approximate continuous-time function d_T from (6). Specifically, standard properties of quadrature using the rectangle rule [22] lead to the estimates

$$\|d_T^2 - d_{T,\Delta t}^2\|_{C(M \times M)} \leq \frac{\|d^2\|_{C^1(M \times M)} \Delta t}{2} = \frac{\|d^2\|_{C^1(M \times M)} T}{2Q}, \quad (27)$$

$$\|d_T^2 - d_{T,\Delta t}^2\|_{C^1(M \times M)} = o(\Delta t^0). \quad (28)$$

Similarly, we approximate the continuous-time kernel k_T in (7) by $k_{T,\Delta t} := h \circ d_{T,\Delta t}^2$. Note that (28) merely indicates that as $\Delta t \rightarrow 0$, $d_{T,\Delta t}^2$ converges to d_T^2 in C^1 norm. A stronger bound can be obtained if d_T has higher than C^1 regularity, e.g., $\|d_T^2 - d_{T,\Delta t}^2\|_{C^1(M \times M)} = O(\Delta t^\alpha)$ if it lies in $C^{1,\alpha}(M \times M)$ for some $\alpha > 0$.

Step 2 (Sampling measure) Replace the Hilbert space $L^2(\mu)$ associated with the invariant measure with the finite-dimensional Hilbert space $L^2(\mu_N)$ associated with the sampling measure $\mu_N = \sum_{n=0}^{N-1} \delta_{x_n}/N$ on the dynamical trajectory $x_0, \dots, x_{N-1} \in \Omega$ underlying the data $y_0, \dots, y_{N-1}$. Here, δ_x denotes the Dirac measure supported at $x \in \Omega$. The space $L^2(\mu_N)$ consists of equivalence classes of measurable, complex-valued functions on Ω with common values at the sampled states $x_0, \dots, x_N$, and is equipped with the inner product

$$\langle f, g \rangle_N = \int_{\Omega} f^* g \, d\mu_N = \frac{1}{N} \sum_{n=0}^{N-1} f^*(x_n) g(x_n).$$

For simplicity of exposition, we will assume that all sampled states x_n are distinct (by ergodicity, this will be the case aside from trivial cases), so $L^2(\mu_N)$ is an N -dimensional Hilbert space, canonically isomorphic to $\mathbb{C}^N$ equipped with a normalized dot product. Under this isomorphism, an element $f \in L^2(\mu_N)$ is represented by a column vector $\vec{f} = (f_0, \dots, f_{N-1})^\top \in \mathbb{C}^N$ such that $f_n = f(x_n)$, and we have $\langle f, g \rangle_N = \vec{f} \cdot \vec{g}/N$. Moreover, a linear map $A : L^2(\mu_N) \rightarrow L^2(\mu_N)$ is represented by an $N \times N$ matrix A such that $A\vec{f}$ corresponds to the column vector representation of Af . We will also assume without loss of generality that the starting state x_0 (and thus the entire sampled dynamical trajectory) lies in the forward-invariant manifold M , but note that x_0 need not lie on the support X of the invariant measure. In light of these facts, our data-driven schemes can be numerically implemented using standard tools from linear algebra, and as we will see below, their formulation requires few structural modifications of their infinite-dimensional counterparts from Sect. 2.

Step 3 (Data-driven integral operator) Approximate the kernel integral operator $K_T : L^2(\mu) \rightarrow L^2(\mu)$ by the operator $K_{T,\Delta t,N} : L^2(\mu_N) \rightarrow L^2(\mu_N)$, where

$$K_{T,\Delta t,N} f = \int_{\Omega} k_{T,\Delta t}(\cdot, x) f(x) \, d\mu_N(x) = \frac{1}{N} \sum_{n=0}^{N-1} k_{T,\Delta t}(\cdot, x_n) f(x_n).$$

This operator is self-adjoint, and there exists a real orthonormal basis $\{\phi_{0,T,\Delta t,N}, \dots, \phi_{N-1,T,\Delta t,N}\}$ of $L^2(\mu_N)$ consisting of its eigenvectors, with corresponding eigenvalues

$\lambda_{0,T,\Delta t,N} \geq \lambda_{1,T,\Delta t,N} \geq \cdots \geq \lambda_{N-1,T,\Delta t,N}$. The data-driven operator $K_{T,\Delta t,N}$ is understood as an approximation of K_T in the following spectral sense:

- Let $\lambda_{j,T,\Delta t,N}$ be a nonzero eigenvalue of $K_{T,\Delta t,N}$. Then, $\lambda_{j,T,\Delta t,N}$ is employed as an approximation of eigenvalue $\lambda_{j,T}$ of K_T .
- Eigenfunction $\phi_{j,T,\Delta t,N} \in L^2(\mu_N)$ has a continuous representative

$$\varphi_{j,T,\Delta t,N} = \frac{1}{\lambda_{j,T,\Delta t,N}} \int_{\Omega} k_{T,\Delta t}(\cdot, x) \phi_{j,T,\Delta t,N}(x) d\mu_N(x), \quad (29)$$

defined everywhere on Ω . The restriction of $\varphi_{j,T,\Delta t,N}$ to M is a continuously differentiable function, employed as an approximation of $\varphi_{j,T}$ from (9).

Numerically, the eigenvalues and eigenvectors of $K_{T,\Delta t,N}$ are computed by solving the eigenvalue problem for the $N \times N$ kernel matrix $\mathbf{K} = [k_{T,\Delta t,N}(x_m, x_n)]_{mn}/N$, which is the matrix representation of $K_{T,\Delta t,N}$ according to Step 2 above. For kernels with rapidly decaying shape functions (e.g., the Gaussian kernels employed in Sect. 5 below), the leading eigenvalues and eigenvectors of $\mathbf{K}$ are well approximated by the corresponding eigenvalues and eigenvectors of a sparse matrix obtained by zeroing out small entries of $\mathbf{K}$, considerably reducing computational cost. See, e.g., Appendix A in [29], or Appendix B in [19] for further details on numerical implementation.

Step 4 (Shift operator) For each time $t = q \Delta t$, $q \in \mathbb{N}_0$, approximate the Koopman operator $U^t : L^2(\mu) \rightarrow L^2(\mu)$ by the q -step shift operator $U_N^q : L^2(\mu_N) \rightarrow L^2(\mu_N)$, defined as

$$U_N^q f(x_n) = \begin{cases} f(x_{n+q}), & 0 \leq n \leq N-1-q, \\ 0, & n > N-1-q. \end{cases}$$

It should be noted that, unlike $U^t f = f \circ \Phi^t$, the shift operator U_N^q is not a composition operator by the underlying dynamical flow—this is because Φ^t does not preserve μ_N -null sets, and thus $\circ \Phi^t$ does not lift to an operator on equivalence classes of functions in $L^2(\mu_N)$. In fact, while U^t is unitary, U_N^q is a nilpotent operator with $U_N^N = 0$. Still, despite these differences, one can interpret U_N^q as an approximation of the Koopman operator in the following sense:

- Let $\mathcal{U}^t : C(M) \rightarrow C(M)$, $t \geq 0$, denote the Koopman operator on continuous functions on the forward-invariant manifold M . Let also $\iota_N : C(M) \rightarrow L^2(\mu_N)$ be the canonical linear operator mapping $C(M)$ functions to their corresponding equivalence classes in $L^2(\mu_N)$, respectively. Then, for any fixed $q \in \mathbb{N}_0$ and continuous function $f \in C(M)$, we have

$$U_N^q \circ \iota_N f = \iota_N \circ \mathcal{U}^{q \Delta t} f + r_N, \quad (30)$$

where $r_N \in L^2(\mu_N)$ are residuals whose norm converges to 0, $\lim_{N \rightarrow \infty} \|r_N\|_{L^2(\mu_N)} = 0$. In contrast, the Koopman operator on $L^2(\mu)$ satisfies $U^t \circ \iota f = \iota \circ \mathcal{U}^t f$ for any (fixed) $t \in \mathbb{R}$, where $\iota : C(M) \rightarrow L^2(\mu)$ is the canonical inclusion map.

Step 5 (Finite-difference operator) Approximate the generator $V : D(V) \rightarrow L^2(\mu)$ by the finite-difference operator $V_{\Delta t, N} : L^2(\mu_N) \rightarrow L^2(\mu_N)$, where

$$V_{\Delta t, N} = \frac{U_N - \text{Id}}{\Delta t}.$$

Explicitly, we have

$$V_{\Delta t, N} f(x_n) = \begin{cases} (f(x_{n+1}) - f(x_n)) / \Delta t, & 0 \leq n \leq N-2, \\ -f(x_{N-1}) / \Delta t, & n = N-1. \end{cases}$$

This operator can be understood as an approximation of the generator in the following sense:

- Let $\mathcal{V}_{\Delta t} : C(M) \rightarrow C(M)$ be the finite-difference approximation of the dynamical vector field $\mathcal{V} : C^1(M) \rightarrow C(M)$, given by

$$\mathcal{V}_{\Delta t} = \frac{\mathcal{U}^{\Delta t} - \text{Id}}{\Delta t}.$$

Then, for any $f \in C(M)$, we have

$$V_{\Delta t, N} \circ \iota_N f = \iota_N \circ \mathcal{V}_{\Delta t} f + r_{\Delta t, N},$$

where $\lim_{N \rightarrow \infty} \|r_{\Delta t, N}\|_{L^2(\mu_N)} = 0$. If, in addition, f lies in $C^1(M)$, then

$$\mathcal{V}_{\Delta t} f = \mathcal{V} f + r_{\Delta t}, \quad (31)$$

where the residual $r_{\Delta t}$ converges uniformly to 0 as the sampling interval decreases, $\lim_{\Delta t \rightarrow 0^+} \|r_{\Delta t}\|_{C(M)} = 0$. Note that the generator V on $L^2(\mu)$ satisfies $V \circ \iota f = \iota \circ \mathcal{V} f$ for any $f \in C^1(M)$.

Step 6 (Coherent features) In order to construct coherent observables analogously to Theorem 1, pick two consecutive, nonzero, simple eigenvalues of $K_{T, \Delta t, N}$, which we denote $\lambda_{T, \Delta t, N}$ and $\nu_{T, \Delta t, N}$ suppressing j subscripts and consider corresponding real normalized eigenfunctions $\phi_{T, \Delta t, N}$ and $\psi_{T, \Delta t, N}$, respectively. Alternatively, a single twofold-degenerate nonzero eigenvalue can be used. Then, form the complex unit vector $z_{T, \Delta t, N} = (\phi_{T, \Delta t, N} + i\psi_{T, \Delta t, N}) / \sqrt{2} \in L^2(\mu_N)$, and compute its continuous representative

$$\zeta_{\Delta t, N} = \frac{1}{\sqrt{2}} \int_{\Omega} k_{T, \Delta t, N}(\cdot, x) \left(\frac{\phi_{\Delta t, N}}{\lambda_{\Delta t, N}} + i \frac{\psi_{\Delta t, N}}{\nu_{\Delta t, N}} \right) d\mu_N(x). \quad (32)$$

The function $\zeta_{\Delta t, N}$ is employed as a data-driven coherent feature, analogous to ζ in Corollary 3. Note, in particular, that $\zeta_{\Delta t, N}$ is expressible as a finite linear combination of kernel sections $k(\cdot, x_n)$ and thus can be empirically evaluated at any point in Ω . Moreover, we construct data-driven analogs of the autocorrelation function α_t for $t = q \Delta t$ and the oscillatory frequency ω by computing

$$\alpha_{q, \Delta t, N} = \langle z_{\Delta t, N}, U_N^q z_{\Delta t, N} \rangle_N, \quad \omega_{\Delta t, N} = \langle \psi_{\Delta t, N}, V_{\Delta t, N} \phi_{\Delta t, N} \rangle_N, \quad (33)$$

respectively.

4.2 Convergence in the large-data limit

We are interested in establishing convergence of the data-driven coherent observable $\zeta_{\Delta t, N}$, autocorrelation function $\alpha_{q, \Delta t, N}$, and oscillatory frequency $\omega_{\Delta t, N}$ to their counterparts from Sect. 2 in a limit of large data, $N \rightarrow \infty$, and vanishing sampling interval,

$\Delta t \rightarrow 0$. For that, we follow a similar approach to [19,34], who employ spectral approximation results for kernel integral operators by Von Luxburg et al. [44]. The principal elements of this approach are as follows.

Operators on continuous functions Since the operators K_T and $K_{T,\Delta t,N}$ act on different Hilbert spaces, we use the space of continuous functions on the forward-invariant manifold M as a universal comparison space to establish spectral convergence. In particular, since the kernels k_T and $k_{T,\Delta t,N}$ are all continuous, one can consider integral operators $\mathcal{K}_T : C(M) \rightarrow C(M)$ and $\mathcal{K}_{T,\Delta t,N} : C(M) \rightarrow C(M)$, defined analogously to $K_T : L^2(\mu) \rightarrow L^2(\mu)$ and $K_{T,\Delta t,N} : L^2(\mu_N) \rightarrow L^2(\mu_N)$, respectively. We then have $\iota \circ \mathcal{K}_T = K_T \circ \iota$ and $\iota_N \circ \mathcal{K}_{T,\Delta t,N} = K_{T,\Delta t,N} \circ \iota_N$, and it is straightforward to verify that $\lambda_{j,T}$ (resp. $\lambda_{j,T,\Delta t,N}$) is a nonzero eigenvalue of K_T (resp. $K_{T,\Delta t,N}$) if and only if it is a nonzero eigenvalue of $\mathcal{K}_T$ (resp. $\mathcal{K}_{T,\Delta t,N}$). Moreover, if $\phi_{j,T} \in L^2(\mu)$ (resp. $\phi_{j,T,\Delta t,N} \in L^2(\mu_N)$) is a corresponding eigenfunction of K_T (resp. $K_{T,\Delta t,N}$), then $\phi_{j,T} \in C(M)$ from (9) (resp. $\phi_{j,T,\Delta t,N} \in C(M)$ from (29)) is a corresponding eigenfunction of $\mathcal{K}_T$ (resp. $\mathcal{K}_{T,\Delta t,N}$). It can further be shown that $\mathcal{K}_T$ is compact, and clearly, $\mathcal{K}_{T,\Delta t,N}$ has finite rank.

Ergodicity and physical measures Let $B_\mu \subseteq \Omega$ be the basin of the ergodic invariant measure μ , i.e., the set of initial conditions $x_0 \in \Omega$ such that the corresponding sampling measures μ_N weak-converge to μ ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_N} f = \mathbb{E}_\mu f, \quad \forall f \in C_b(\Omega), \quad (34)$$

for Lebesgue almost every sampling interval Δt . Here, $\mathbb{E}_\rho f = \int_\Omega f d\rho$ denotes expectation with respect to a measure ρ , and $C_b(\Omega)$ is the Banach space of continuous, real-valued functions on Ω equipped with the uniform norm. By ergodicity of the dynamical flow Φ^t , $B_\mu \cap X$ is a dense subset of the support X of μ . Moreover, for a class of dynamical systems possessing so-called physical measures [65] the basin B_μ has positive measure with respect to an ambient probability measure on state space Ω from which initial conditions are drawn, even if X is a null set with respect to that measure. In such situations, the data-driven scheme described in Sect. 4.1 converges from a “large” set of experimentally accessible initial conditions, which need not lie on the support of μ . Examples include the L63 system, where the ergodic invariant measure supported on the Lorenz attractor is a Sinai–Ruelle–Bowen (SRB) measure with a basin of positive Lebesgue measure in $\Omega = \mathbb{R}^3$ [62]. For simplicity of exposition, and without loss of generality with regards to asymptotic convergence, we will henceforth assume that the initial state x_0 lies in $B_\mu \cap M$. Moreover, $\Delta t \rightarrow 0$ limits will be assumed to be taken along a sequence such that (34) holds.

Spectral convergence Since our approach for coherent feature extraction employs on eigenvalues and eigenvectors of kernel integral operators, it is necessary to ensure that the family $\mathcal{K}_{T,\Delta t,N}$ converges to $\mathcal{K}_T$ in a sufficiently strong sense so as to imply spectral convergence. Here, we consider the iterated limit of $N \rightarrow \infty$ followed by $\Delta t \rightarrow 0$; under the former limit, empirical expectation values with respect to the sampling measures converge to expectation values with respect to the invariant measure (according to (34)), and under the latter limit, the kernels based on discrete-time delay-coordinate maps converge to their continuous-time counterparts (according to (27)). In particular, we have:

Proposition 2 *With notation and assumptions as above, let $\lambda_{j,T}$ be a nonzero eigenvalue of $\mathcal{K}_T$, where the ordering $\lambda_{0,T} \geq \lambda_{1,T} \geq \dots$ is in decreasing order and includes multiplicities. Let $\Pi_{j,T} : C(M) \rightarrow C(M)$ be the spectral projection to the corresponding eigenspace. Then, the following hold:*

- (i) *The j th eigenvalues $\lambda_{j,T,\Delta t,N}$ of $\mathcal{K}_{T,\Delta t,N}$ (ordered with the same convention as the eigenvalues of $\mathcal{K}_T$) converge to $\lambda_{j,T}$, in the sense of the iterated limit*

$$\lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \lambda_{j,T,\Delta t,N} = \lambda_{j,T}.$$

- (ii) *For any neighborhood $\Sigma \subseteq \mathbb{C}$ such that $\sigma(\mathcal{K}_T) \cap \Sigma = \{\lambda_{j,T}\}$, the spectral projections $\Pi_{\Sigma,T,\Delta t,N}$ of $\mathcal{K}_{T,\Delta t,N}$ onto Σ converge strongly to $\Pi_{j,T}$. In particular, for any eigenfunction $\varphi_{j,T} \in C(M)$ of $\mathcal{K}_T$ corresponding to eigenvalue $\lambda_{j,T}$ there exist eigenfunctions $\varphi_{j,T,\Delta t,N} \in C(M)$ of $\mathcal{K}_{T,\Delta t,N}$ corresponding to $\lambda_{j,T,\Delta t,N}$, such that*

$$\lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \|\varphi_{j,T,\Delta t,N} - \varphi_{j,T}\|_{C(M)} = 0.$$

Remark 2 Analogous spectral convergence results to Proposition 2 hold for integral operators with data-dependent kernels $k_{T,\Delta t,N}$, so long as these kernels have well defined $N \rightarrow \infty$ limits in $C(M)$ norm. Examples of such kernels include Markov-normalized kernels [10, 15, 16] and variable-bandwidth Gaussian kernels [9]. See, e.g., Theorem 7 in [34] for a spectral convergence result for data-dependent kernels related to the kernels employed in the numerical experiments in Sect. 5.

A corollary of Proposition 2 is that the properties the data-driven coherent observable $\zeta_{\Delta t,N}$ from (32) and the corresponding empirical autocorrelation function and oscillatory frequency in (33) converge to their counterparts from Theorem 1 and thus obey the same pseudospectral bounds associated with dynamical coherence.

Corollary 4 *Under the assumptions of Proposition 2, the following hold in the large-data limit, $\Delta t \rightarrow 0$ after $N \rightarrow \infty$, where ζ , α_t , and ω are defined in Theorem 1:*

- (i) *$\zeta_{\Delta t,N}$ converges to the coherent feature ζ , uniformly on the forward-invariant manifold M , i.e.,*

$$\lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \|\zeta_{\Delta t,N} - \zeta\|_{C(M)} = 0.$$

- (ii) *For any $q \in \mathbb{N}$, the empirical autocorrelation $\alpha_{q,\Delta t,N}$ converges to the autocorrelation function α_t at $t = q \Delta t$.*

- (iii) *The empirical oscillatory frequency $\omega_{\Delta t,N}$ converges to the frequency ω .*

Proof The uniform convergence of $\zeta_{\Delta t,N}$ to ζ in Claim (i) is a direct consequence of Proposition 2. Claim (ii) follows from the Claim (i), in conjunction with the residual estimate in (30), viz.

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \alpha_{q,\Delta t,N} &= \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \langle z_{\Delta t,N}, U_N^q z_{\Delta t,N} \rangle_N \\ &= \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \langle \iota_N \zeta_{\Delta t,N}, U_N^q \iota_N \zeta_{\Delta t,N} \rangle_N \\ &= \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \langle \iota_N \zeta_{\Delta t,N}, \iota_N \mathcal{U}^{q \Delta t} \zeta_{\Delta t,N} + r_N \rangle_N \\ &= \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \langle \iota_N \zeta_{\Delta t,N}, \iota_N \mathcal{U}^{q \Delta t} \zeta_{\Delta t,N} \rangle_N \end{aligned}$$

$$\begin{aligned}
&= \lim_{\Delta t \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \zeta_{\Delta t, N}(x_n) z_{\Delta t, N}(x_{n+q}) \\
&= \int_{\Omega} \zeta^* \mathcal{U}^{q \Delta t} \zeta \, d\mu = \langle \iota \zeta, \iota \mathcal{U}^{q \Delta t} \zeta \rangle = \langle z, U^{q \Delta t} z \rangle \\
&= \alpha_q \Delta t.
\end{aligned}$$

Claim (iii) follows similarly, using a finite-difference residual estimate in (31), in conjunction with the C^1 -norm convergence of $k_{T, \Delta t}$ to k_T as $\Delta t \rightarrow 0$ (see (28)). $\square$

4.3 Choice of kernel

Following [21], in the numerical experiments described below we employ integral operators $K_{T, \Delta t, N} : L^2(\mu_N) \rightarrow L^2(\mu_N)$ associated with a family of symmetric, Markov-normalized kernels $k_{T, \Delta t, N}$ constructed using the variable-bandwidth Gaussian kernels in conjunction with the bistochastic Markov normalization procedure proposed in [9] and [15], respectively. Specifically, to build $k_{T, \Delta t, N}$ we start from a radial Gaussian kernel $\bar{k}_{T, \Delta t} : \Omega \times \Omega \rightarrow \mathbb{R}_+$ on delay-coordinate mapped data,

$$\bar{k}_{T, \Delta t}(x, x') = \exp \left(-\frac{d_{T, \Delta t}^2(x, x')}{\bar{\sigma}^2} \right),$$

where $\bar{\sigma}$ is a positive bandwidth parameter determined numerically from the data (see, e.g., Algorithm 1 in [29]). Using this kernel, we compute the bandwidth function $\rho_{T, \Delta t, N} \in C(\Omega)$ given by

$$\rho_{T, \Delta t, N}(x) = \left(\int_{\Omega} \bar{k}_{T, \Delta t}(x, \cdot) \, d\mu_N \right)^{-1/m} = \left(\frac{1}{N} \sum_{n=0}^{N-1} \bar{k}_{T, \Delta t}(x, x_n) \right)^{-1/m}.$$

Here, $m > 0$ is an estimate of the dimension of the support X of the invariant measure, computed through the same procedure used to tune the kernel bandwidth $\bar{\sigma}$. We then build the variable-bandwidth kernel $\kappa_{T, \Delta t, N} : \Omega \times \Omega \rightarrow \mathbb{R}_+$, where

$$\kappa_{T, \Delta t, N}(x, x') = \exp \left(-\frac{d_{T, \Delta t}^2(x, x')}{\sigma^2 \rho_{T, \Delta t, N}(x) \rho_{T, \Delta t, N}(x')} \right). \quad (35)$$

In the above, σ is a positive bandwidth parameter determined automatically in a similar manner as $\bar{\sigma}$, though note that in general, σ and $\bar{\sigma}$ have different values.

By construction, $\kappa_{T, \Delta t, N}$ is continuous, positive, and bounded away from zero on $M \times M$. Intuitively, the function $\rho_{T, \Delta t, N}^{-m}$ can be thought of as a kernel estimate of the “sampling density” of the data relative to an ambient measure. The variable-bandwidth construction in (35) can then be thought of as a data-adaptive adjustment of the bandwidth σ , such that a data point x is assigned a smaller (larger) bandwidth $\sigma \rho_{T, \Delta t, N}(x)$ when the sampling density is higher (lower), thus reducing sensitivity to sampling errors. This intuition can be made precise if the support X has the structure of a Riemannian manifold and μ the structure of a smooth volume form. In that case, the variable-bandwidth kernel effects a conformal change of Riemannian metric on the data such that in the new geometry the invariant measure has constant density relative to the Riemannian volume form; see [29] for further details.

Next, we normalize the kernel $\kappa_{T, \Delta t, N}$ to obtain a symmetric Markov kernel $k_{T, \Delta t, N} : \Omega \times \Omega \rightarrow \mathbb{R}_+$ by first computing the strictly positive, continuous functions

$$u_{T, \Delta t, N} = \int_{\Omega} \kappa_{T, \Delta t, N}(\cdot, x) \, d\mu_N(x), \quad v_{T, \Delta t, N} = \int_{\Omega} \frac{\kappa_{T, \Delta t, N}(\cdot, x)}{u_{T, \Delta t, N}(x)} \, d\mu_N(x),$$

and then defining

$$k_{T,\Delta t,N}(x, x') = \int_{\Omega} \frac{\kappa_{T,\Delta t,N}(x, x'')\kappa_{T,\Delta t,N}(x'', x)}{u_{T,\Delta t,N}(x)v_{T,\Delta t,N}(x'')u_{T,\Delta t,N}(x')} d\mu_N(x''). \quad (36)$$

It can be readily verified that with this definition $k_{T,\Delta t,N}$ is a symmetric, strictly positive kernel with the Markov property, $\int_{\Omega} k_{T,\Delta t,N}(x, x') d\mu_N(x') = 1$, for all $x \in \Omega$. Moreover, $k_{T,\Delta t,N}$ is (strictly) positive-definite on the support of μ_N if $\kappa_{T,\Delta t,N}$ is (strictly) positive-definite. It can further be shown [34] that in the large-data limit, $\Delta t \rightarrow 0$ after $N \rightarrow \infty$, $k_{T,\Delta t,N}$ converges to an $L^2(\mu)$ -Markov, symmetric, continuous kernel k_T so an analogous spectral convergence result to Proposition 2 holds for this class of kernels (see also Remarks 1 and 2).

For the purposes of extraction of coherent observables of measure-preserving, ergodic dynamics, symmetric Markov kernels have the natural property of exhibiting a constant eigenfunction corresponding to the top eigenvalue, $\lambda_{0,T,\Delta t,N} = \lambda_{0,T} = 1$, with the remaining eigenfunctions capturing mutually orthogonal features orthogonal to the constant. See Appendix B in [21] for pseudocode for solving the eigenvalue problem for $K_{T,\Delta t,N}$, where explicit formation of the kernel in (36) is avoided through singular value decomposition of a non-symmetric kernel matrix.

5 Numerical examples

5.1 Dataset description

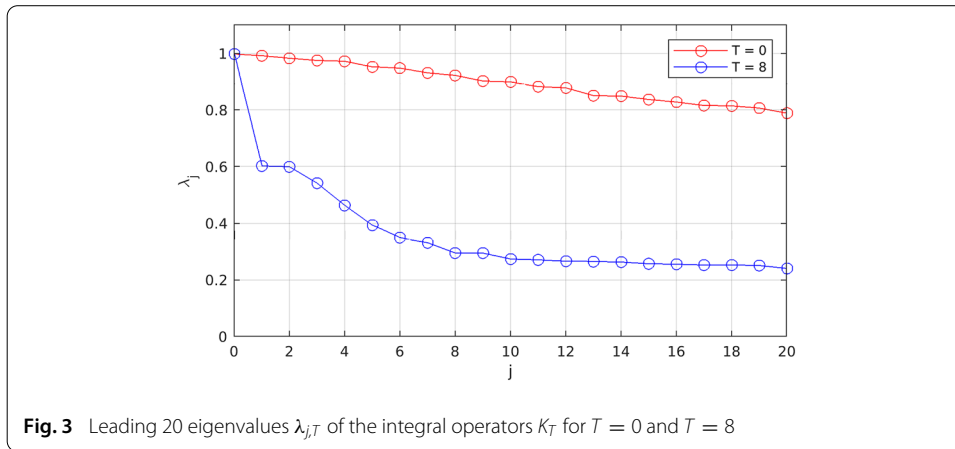
As an application of the results in Sects. 2–4, we study the properties of eigenfunctions of the integral operators $K_{T,\Delta t,N}$ induced by the L63 system on $\Omega = \mathbb{R}^3$ with the standard parameters,

$$\begin{aligned} \dot{x} &= \vec{V}(x), \quad x = (x^1, x^2, x^3) \in \mathbb{R}^3, \quad \vec{V}(x) = (V^1, V^2, V^3), \\ V^1 &= 10(x^2 - x^1), \quad V^2 = 28x^1 - x^2 - x^1x^3, \quad V^3 = x^1x^2 - 8x^3/3. \end{aligned}$$

We generate numerical trajectories $x_0, \dots, x_{\tilde{N}-1} \in \Omega$ sampled at an interval $\Delta t = 0.01$ natural time units using MATLAB's `ode45` solver. Numerical integration starts at an arbitrary point $\tilde{x} \in \mathbb{R}^3$, and we allow the state to settle near the Lorenz attractor over a spinup time of 640 time units before collecting the first sample x_0 . In anticipation of the fact that we will be using the delay-coordinate map in (26), we sample a total of $\tilde{N} = N + Q - 1$ states, where Q is the number of delays, and N is fixed at $N = 64,000$.

We consider two experiments, one with $Q = 1$ corresponding to no delays ($T = 0$) and another one with $Q = 800$ corresponding to a delay-embedding window of $T = Q \Delta t = 8$ natural time units. The latter, is approximately equal to 9 Lyapunov characteristic times $T_L = 1/\Lambda$, where $\Lambda \approx 0.91$ [55] is the positive Lyapunov exponent of the L63 system. The $T = 8$ embedding window is also approximately equal to 10 oscillations assuming a characteristic oscillatory timescale of $T_o = 0.8$. In both cases we set the observation map $F : \Omega \rightarrow Y$ to the identity map on $\mathbb{R}^3$, so the corresponding delay coordinate map $F_{Q,\Delta t}$ takes values in $Y^Q = \mathbb{R}^{3Q}$. Note that, after delay embedding, each experiment has $N = 64,000$ samples $y_n = F_{Q,\Delta t}$ available for analysis, which corresponds to 800 oscillatory timescales T_o .

As stated in Sect. 2.1, this L63 setup rigorously satisfies all the assumptions made in Theorem 1 [42,45,62]. In addition, since $\Delta t \ll T_o$, $(N - 1) \Delta t \gg T_o$, and in the $T = 8$ setup, $\Delta t \ll T$, we expect no significant sampling errors to be present in our numerical experiments; in particular, we expect the leading eigenfunctions of the data-driven integral



operators $K_{T,\Delta t N}$ to be good approximations of the corresponding eigenfunctions of the operators K_T from Theorem 1.

5.2 Coherent observables

We now discuss the properties of eigenfunctions of $K_{T,\Delta t N}$ constructed using the approach described in Sect. 4.3, some of which were already shown in Figs. 1 and 2. All results were obtained using the symmetric Markov kernels in (36) with $N = 64,000$, $\Delta t = 0.01$, and representative values of T in the range 0 to 64. For the rest of this section, we suppress Δt and N indices from our notation. Moreover, we do not distinguish between eigenfunctions $z \in L^2(\mu_N)$ and their continuous representatives $\zeta \in C(M)$, as our visualizations will be restricted to the training dataset $\{x_n\}_{n=0}^{N-1}$ for which $z(x_n) = \zeta(x_n)$.

We begin in Fig. 3 with a plot of the leading 20 eigenvalues $\lambda_{j,T}$ of K_T for $T = 0$ and $T = 8$, where both operators have the top eigenvalue $\lambda_{0,T} = 1$ by Markovianity of the kernels. When $T = 0$, K_T has a small spectral gap $\lambda_{0,T} - \lambda_{1,T} \approx 0.0075$, and the subsequent eigenvalues exhibit a gradual decay, reaching $\lambda_{j,T} \approx 0.8$ at $j = 20$. In contrast, when $T = 8$, K_T exhibits a significantly larger spectral gap $\lambda_{0,T} - \lambda_{1,T} \approx 0.4$, with a nearly degenerate corresponding eigenspace. In particular, we have $\lambda_{1,T} \approx 0.6032$ and $\lambda_{2,T} \approx 0.6015$, and the corresponding gap parameters from Theorem 1 take the values $\delta_T \approx 0.001$ and $\tilde{\delta}_T \approx 0.002$. Thus, on the basis of Theorem 1, the complex-valued observable $z = (\phi_{1,T} + i\phi_{2,T})/\sqrt{2}$ for $T = 8$ is a good candidate of a dynamically coherent feature evolving as an ε -approximate eigenfunction of the Koopman operator with small ε . The scatterplots and time series plots in Fig. 1 were already suggestive of this behavior, which we now examine in further detail. As a point of comparison, we consider the corresponding observable z constructed from the leading eigenfunctions of K_T at $T = 0$, which were also depicted in Fig. 1.

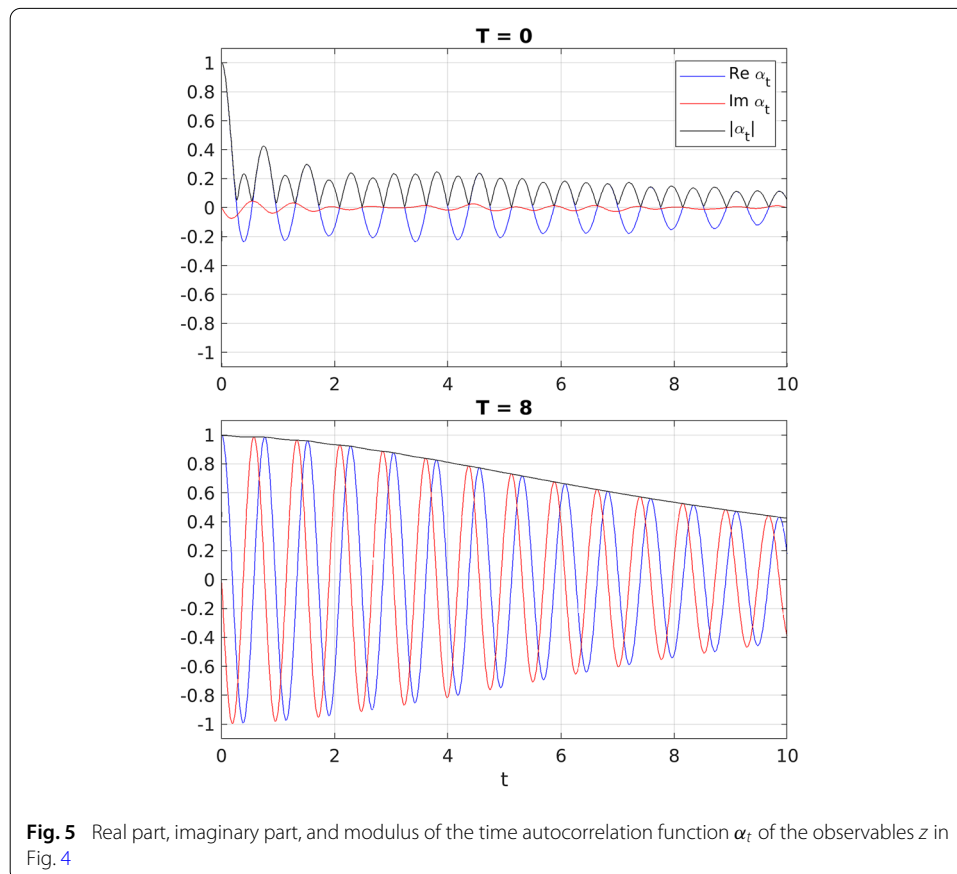
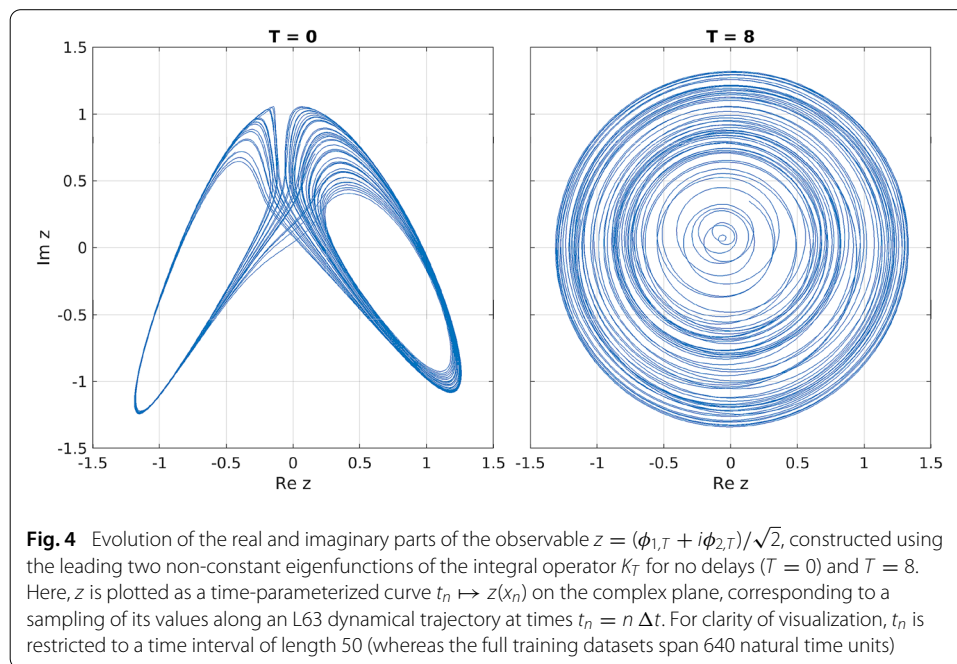
Figure 4 shows the evolution of the observables z as a time-parameterized curve $t_n \mapsto z(x_n)$ on the complex plane over a portion of the training data spanning 50 natural time units. In effect, these plots correspond to samplings of complex-valued functions on the Lorenz attractor along dynamical trajectories, akin to the time series plots in Fig. 1 which (up to a scaling by a factor of $\sqrt{2}$) correspond to the real and imaginary parts of z . The $T = 0$ evolution traces out what qualitatively resembles a two-dimensional projection of the attractor. In particular, we do not expect a dynamically coherent behavior for this

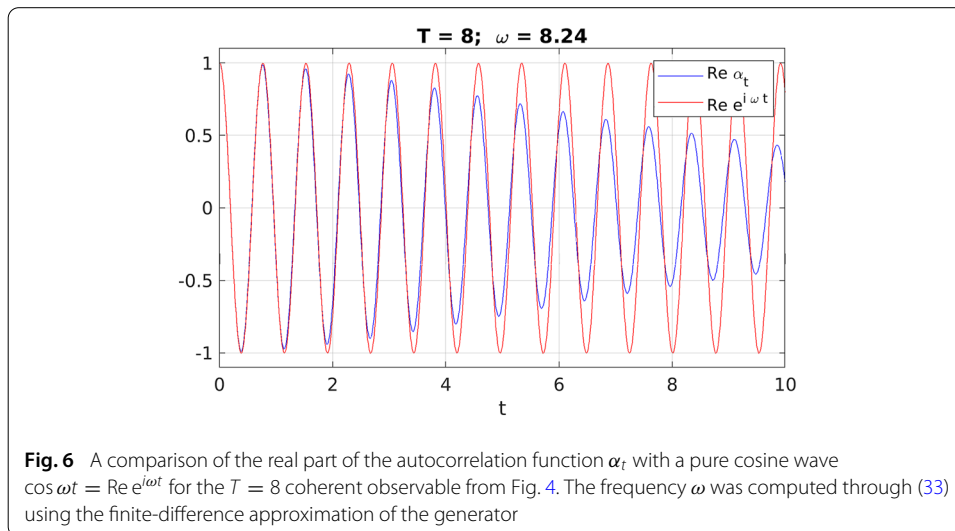
observable, as its evolution comprises of two cycles with a mixing region when $\operatorname{Re} z \simeq 0$ which is not too different from the raw L63 dynamics. This lack of coherence is manifestly visible in the scatterplots in Fig. 1 depicting the real and imaginary parts of z acted upon by the Koopman operator and can also be assessed more quantitatively through plots of the time-autocorrelation function α_t , shown in Fig. 5. There, the modulus $|\alpha_t|$ is seen to rapidly decay from its initial value $|\alpha_0| = 1$, reaching $|\alpha_t| \approx 0.05$ at $t \approx 0.3$, and never exceeds 0.4 after $\simeq 1$ Lyapunov time.

In contrast, the observable z constructed from the eigenfunctions of K_T at $T = 8$ exhibits a fundamentally different behavior, consistent with an approximate cycle that remains coherent over several Lyapunov timescales. In Fig. 4, the $T = 8$ dynamical trajectory lies in what appears to be a disk in the complex plane, executing a predominantly azimuthal motion with a slow radial motion (amplitude modulation) superposed. In particular, the real and imaginary parts of z have a 90° phase difference to a good approximation (at least when $|z|$ is not too small), and as indicated by the time series plots in Fig. 1, they have a nearly constant characteristic frequency. The coherent dynamical evolution stemming from this behavior is visually evident in the scatterplots of the real and imaginary parts of $U^t z$ in Fig. 1, which appear to “resist” mixing of level sets on significantly longer timescales than the $T = 0$ eigenfunctions.

More quantitatively, in Fig. 5, the evolution of the autocorrelation function α_t of z for $T = 8$ is consistent with an amplitude-modulated harmonic oscillator with a well-defined carrier frequency and slowly varying envelope function. In particular, the real and imaginary parts of α_t oscillate at a near-constant frequency and remain phase-locked to a 90° phase difference at least out to $t = 10$ natural time units, or $\simeq 10$ Lyapunov times. Meanwhile, the modulus $|\alpha_t|$ exhibits a significantly slower decay than what was observed for $T = 0$ and remains above 0.4 for all $t \in [0, 10]$. In Fig. 6, we compare the evolution of the autocorrelation function α_t with a pure sinusoid $e^{i\omega t}$ with frequency ω determined through the finite-difference-approximated generator using (33). The generator-based frequency, $\omega \approx 8.24$ (corresponding to a period of $2\pi/\omega \approx 0.76$), is seen to accurately capture the carrier frequency of the α_t signal, as expected from Theorem 1, with a slow build-up of phase decoherence that becomes noticeable by $t \simeq 10$.

Intriguingly, the $\omega \approx 8.24$ frequency identified here through eigenfunctions of K_T is close to an 8.18 approximate eigenfrequency identified in [21] through spectral analysis of a compact approximation to the generator V constructed using reproducing kernel Hilbert space (RKHS) techniques. The RKHS-based eigenfrequency has a corresponding approximate Koopman eigenfunction, z_{RKHS} , which has a qualitatively similar spatial structure on the L63 attractor as the approximate eigenfunction z identified here (compare Figure 5 in [21] with Fig. 1 of this paper). Moreover, both z and z_{RKHS} resemble an observable identified by Korda et al. [40] through a spectral analysis technique for Koopman operators utilizing Christoffel–Darboux kernels in frequency space (see Figure 13 in [40]). Having been identified via three independent data analysis techniques, it thus appears that the approximate eigenfrequency $\omega \simeq 8.2$ and the corresponding approximate eigenfunction with the structure depicted in Fig. 1 are robust features of the L63 system, warranting further investigation.





6 Concluding remarks

In this paper, we have studied how kernel integral operators constructed from delay-coordinate mapped data can identify, through their eigenfunctions, dynamically coherent features of measure-preserving, ergodic dynamical systems. We have shown that a class of eigenfunctions of such operators lead to complex-valued observables with an approximately cyclical evolution, behaving as ε -approximate eigenfunctions of the Koopman operator for a bound ε that decreases with the length of the embedding window. Such observables encapsulate a natural notion of dynamical coherence, so we have argued, in the sense of having high regularity on the attractor, a well-defined oscillatory frequency, and a slowly decaying time-autocorrelation amplitude. In addition, the spectral bounds were explicitly characterized as functions of the embedding window length, evolution time, and appropriate spectral gap parameters.

These results extend previous work on integral operators approximating the point spectrum of the Koopman operator in the infinite-delay limit [19, 29] to the setting of mixing dynamical systems with continuous Koopman spectra. Thus, they provide a theoretical interpretation of the efficacy of a number of data-driven techniques utilizing delay embeddings, including DMDC [8], HAVOK analysis [12], NLSA [30, 32, 33], and SSA [11, 63], in extracting coherent signals from complex systems. An attractive aspect of these methods is that they are amenable to consistent data-driven approximation from time series data based on techniques originally developed in the context of spectral clustering [44]. In particular, the data-driven schemes are rigorously applicable in situations where the invariant measure is supported on non-smooth sets, such as fractal attractors, without requiring addition of stochastic noise to regularize the dynamics.

As a numerical application, we have studied how eigenfunctions of kernel integral operators utilizing delay-coordinate maps identify coherent observables of the L63 model—a system known to have a unique SRB measure with mixing dynamics [45, 62], and thus the absence of non-constant Koopman eigenfunctions in L^2 . We found that for a sufficiently long embedding window (of approximately 8 Lyapunov times), the kernel-based approach, realized using a symmetric Markov kernel constructed by bistochastic normalization [15] of a variable-bandwidth Gaussian kernel [9], identifies through its two leading

non-constant eigenfunctions an observable of the L63 system exhibiting a highly coherent dynamical behavior. This observable has an oscillatory period of approximately 0.76 natural time units and remains coherent at least out to 10 natural time units (approximately 9 Lyapunov timescales) as measured by a 0.4 threshold of its time-autocorrelation function. Spatially, its real and imaginary parts have a structure that could be qualitatively described as a wavenumber 1 azimuthal oscillation about the holes in the two lobes of the attractor, a pattern that resembles observables previously identified through Koopman spectral analysis techniques appropriate for mixing dynamical systems [21, 40].

Possible applied directions stemming from this work include detection of coherence in prototype models for metastable regime behavior in atmospheric dynamics [18], as well as PDE models with intermittency in both space and time [46]. On the theoretical side, it would be interesting to explore connections between the spectral results presented here and geometrical characterizations of coherence, including the characterization given in DMDC based on the multiplicative ergodic theorem [8] and the dynamic isoperimetry approach proposed in [27]. It may also be fruitful to employ coherent eigenfunctions of integral operators based on delay-coordinate maps to construct approximation spaces for pointwise and/or spectral approximation of Koopman and transfer operators, including the extended dynamic mode decomposition (EDMD) technique [64] and the RKHS compactification approaches proposed in [21].

Acknowledgements

The author is grateful to Andrew Majda for his guidance and mentorship during a postdoctoral position at the Courant Institute from 2009 to 2012. He is especially grateful for his friendship and collaboration over the years. This research was supported by NSF Grant 1842538, NSF Grant DMS 1854383, and ONR YIP Grant N00014-16-1-2649.

A Proof of Proposition 2

It is convenient to introduce an intermediate integral operator $\mathcal{K}_{T,\Delta t} : C(M) \rightarrow C(M)$,

$$\mathcal{K}_{T,\Delta t}f = \int_{\Omega} k_{T,\Delta t}(\cdot, x)f(x) d\mu,$$

which integrates against the invariant measure μ using the discrete-time delay-coordinate map and split the analysis of the spectral convergence of $\mathcal{K}_{T,\Delta t,N}$ to $\mathcal{K}_T$ to two subproblems involving the convergence of (i) $\mathcal{K}_{T,\Delta t,N}$ to $\mathcal{K}_{T,\Delta t}$ as $N \rightarrow \infty$; and (ii) $\mathcal{K}_{T,\Delta t}$ to $\mathcal{K}_T$ as $\Delta t \rightarrow 0$. We now consider these two subproblems, starting from the second one.

Spectral convergence of $\mathcal{K}_{T,\Delta t}$ to $\mathcal{K}_T$ as $\Delta t \rightarrow 0$ The uniform convergence of the kernels $k_{T,\Delta t}$ to k_T , i.e., $\lim_{\Delta t \rightarrow 0} \|k_{T,\Delta t} - k_T\|_{C(M \times M)}$ (see (27)), implies convergence of $\mathcal{K}_{T,\Delta t}$ to $\mathcal{K}_T$ in $C(M)$ operator norm. It then follows from results on spectral theory of compact operators [3, 13] that the analogous claims to Proposition 2 hold for the eigenvalues and spectral projections, $\lambda_{\Sigma,T,\Delta t}$ and $\Pi_{\Sigma,T,\Delta t}$, respectively, of $\mathcal{K}_{T,\Delta t}$. That is, we have

$$\lim_{\Delta t \rightarrow 0} \lambda_{j,T,\Delta t} = \lambda_{j,T}, \quad \lim_{\Delta t \rightarrow 0} \Pi_{\Sigma,T,\Delta t}f = \Pi_{j,T}f, \quad \forall f \in C(M), \quad (37)$$

where $\Pi_{\Sigma,T,\Delta t} : C(M) \rightarrow C(M)$ is the spectral projection of $\mathcal{K}_{T,\Delta t}$ onto $\Sigma \subseteq \mathbb{C}$.

Spectral convergence of $\mathcal{K}_{T,\Delta t,N}$ to $\mathcal{K}_{T,\Delta t}$ as $N \rightarrow \infty$ Unlike the $\mathcal{K}_{T,\Delta t} \rightarrow \mathcal{K}_T$ case, the operators $\mathcal{K}_{T,\Delta t,N}$ need not converge to $\mathcal{K}_{T,\Delta t}$ in $C(M)$ operator norm. In essence, this is because the weak convergence of measures in (34) is not uniform with respect to f , even upon restriction to functions in $C(M)$. Nevertheless, as shown in [44], the continuity of the kernel k_T is sufficient to ensure that for a fixed $f \in C(M)$, a restricted form of uniform

convergence holds, namely

$$\lim_{N \rightarrow \infty} \sup_{g \in \mathcal{G}} |\mathbb{E}_{\mu_N} g - \mathbb{E}_{\mu} g| = 0, \quad (38)$$

where $\mathcal{G} \subset C(M)$ is the set of functions given by

$$\mathcal{G} = \{k_T(x, \cdot) f(\cdot) \mid x \in M\}.$$

A collection of functions satisfying (38) is known as a Glivenko–Cantelli class.

The Glivenko–Cantelli property turns out to be sufficient to ensure that as $N \rightarrow \infty$, the sequence of operators $\mathcal{K}_{T, \Delta t, N}$ exhibits a form of convergence to $K_{T, \Delta t}$, called collectively compact convergence which, despite being weaker than norm convergence, is sufficiently strong to imply the spectral convergence claims in Proposition 2. We state the relevant definitions for collectively compact convergence below and refer the reader to [13, 44] for additional details.

Definition 1 Let $A_N : E \rightarrow E$ be a sequence of bounded linear operators on a Banach space E , indexed by $N \in \mathbb{N}$.

- (i) A_N is said to converge to an operator $A : E \rightarrow E$ compactly if A_N converges to A strongly, and for every uniformly bounded sequence $f_N \in E$ the sequence $(A - A_N)f_N$ has compact closure.
- (ii) $\{A_N\}$ is said to be collectively compact if $\bigcup_{N \in \mathbb{N}} A_N B$ has compact closure in E , where B is the unit ball of E .
- (iii) A_N is said to converge to A collectively compactly if it converges pointwise, and there exists $N_0 \in \mathbb{N}$ such that for all $N > N_0$, $\{A_N\}_{N > N_0}$ is collectively compact.

It can be shown that operator norm convergence implies collectively compact convergence, and collectively compact convergence implies compact convergence. The latter is in turn sufficient for the following spectral convergence result:

Lemma 3 With the notation of Definition 1, suppose that A_N converges to A compactly. Let $\lambda \in \sigma_p(A)$ be an isolated eigenvalue of A with finite multiplicity m , and Σ an open neighborhood of λ such that $\sigma(A) \cap \Sigma = \{\lambda\}$. Then, the following hold:

- (i) There exists $N_0 \in \mathbb{N}$, such that for all $N > N_0$, $\sigma(A_N) \cap \Sigma$ is an isolated subset of the spectrum of A_N , containing at most m distinct eigenvalues whose multiplicities sum to m . Moreover, as $N \rightarrow \infty$, every element of $\sigma(A_{N > N_0}) \cap \Sigma$ converges to λ .
- (ii) As $N \rightarrow \infty$, the spectral projections of A_N onto $\sigma(A_N) \cap \Sigma$, defined in the sense of the holomorphic functional calculus, converge strongly to the spectral projection of A_N onto $\{\lambda\}$.

Using a similar approach as Proposition 13 in [44], which employs, in particular, the Glivenko–Cantelli property in (38), it can be shown that as $N \rightarrow \infty$, $\mathcal{K}_{T, \Delta t, N}$ converges collectively compactly to $\mathcal{K}_{T, \Delta t}$. Then, Lemma 3, in conjunction with the fact that $\mathcal{K}_{T, \Delta t}$ is compact (so every nonzero element of its spectrum is an isolated eigenvalue of finite multiplicity), implies that

$$\begin{aligned} \lim_{N \rightarrow \infty} \lambda_{j, T, \Delta t, N} &= \lambda_{j, T, \Delta t}, \quad \lim_{N \rightarrow \infty} \Pi_{\Sigma, T, \Delta t, N} f \\ &= \Pi_{\Sigma, T, \Delta t} f, \quad \forall \lambda_{j, T, \Delta t} \in \Sigma, \quad \forall f \in C(M), \end{aligned} \quad (39)$$

where Σ is the spectral neighborhood in the statement of the proposition. Proposition 2 is then proved by combining (37) and (39). $\square$

Received: 4 July 2020 Accepted: 4 December 2020

Published online: 28 January 2021

References

- Alexander, R., Zhao, Z., Szekely, E., Giannakis, D.: Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.* **74**, 1321–1342 (2017). <https://doi.org/10.1175/JAS-D-16-0147.1>
- Arbabi, H., Mezić, I.: Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. *SIAM J. Appl. Dyn. Syst.* **16**(4), 2096–2126 (2017). <https://doi.org/10.1137/17M1125236>
- Atkinson, K.E.: The numerical solution of the eigenvalue problem for compact integral operators. *Trans. Am. Math. Soc.* **129**(3), 458–465 (1967)
- 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>
- Baladi, V.: Positive Transfer Operators and Decay of Correlations. Advanced Series in Nonlinear Dynamics, vol. 16. World Scientific, Singapore (2000)
- Banisch, R., Koltai, P.: Understanding the geometry of transport: diffusion maps for Lagrangian trajectory data unravel coherent sets. *Chaos* **27**, 035804 (2017). <https://doi.org/10.1063/1.4971788>
- Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**, 1373–1396 (2003). <https://doi.org/10.1162/089976603321780317>
- Berry, T., Cressman, R., Gregurić-Ferenček, Z., Sauer, T.: Time-scale separation from diffusion-mapped delay coordinates. *SIAM J. Appl. Dyn. Syst.* **12**, 618–649 (2013). <https://doi.org/10.1137/12088183x>
- Berry, T., Harlim, J.: Variable bandwidth diffusion kernels. *Appl. Comput. Harmon. Anal.* **40**(1), 68–96 (2016). <https://doi.org/10.1016/j.acha.2015.01.001>
- Berry, T., Sauer, T.: Local kernels and the geometric structure of data. *Appl. Comput. Harmon. Anal.* **40**(3), 439–469 (2016). <https://doi.org/10.1016/j.acha.2015.03.002>
- 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](https://doi.org/10.1016/0167-2789(86)90031-x)
- 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>
- Chatelin, F.: Spectral Approximation of Linear Operators. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia (2011)
- Chen, N., Majda, A.J., Giannakis, D.: Predicting the cloud patterns of the Madden–Julian Oscillation through a low-order nonlinear stochastic model. *Geophys. Res. Lett.* **41**(15), 5612–5619 (2014). <https://doi.org/10.1002/2014gl060876>
- Coifman, R., Hirn, M.: Bi-stochastic kernels via asymmetric affinity functions. *Appl. Comput. Harmon. Anal.* **35**(1), 177–180 (2013). <https://doi.org/10.1016/j.acha.2013.01.001>
- Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006). <https://doi.org/10.1016/j.acha.2006.04.006>
- 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>
- Crommelin, D.T., Majda, A.J.: Strategies for model reduction: comparing different optimal bases. *J. Atmos. Sci.* **61**, 2206–2217 (2004). [https://doi.org/10.1175/1520-0469\(2004\)061<2206:sfmrnd>2.0.co;2](https://doi.org/10.1175/1520-0469(2004)061<2206:sfmrnd>2.0.co;2)
- Das, S., Giannakis, D.: Delay-coordinate maps and the spectra of Koopman operators. *J. Stat. Phys.* **175**(6), 1107–1145 (2019). <https://doi.org/10.1007/s10955-019-02272-w>
- Das, S., Giannakis, D.: Koopman spectra in reproducing kernel Hilbert spaces. *Appl. Comput. Harmon. Anal.* **49**(2), 573–607 (2020). <https://doi.org/10.1016/j.acha.2020.05.008>
- Das, S., Giannakis, D., Slawinska, J.: Reproducing kernel Hilbert space quantification of unitary evolution groups (2020). In minor revision
- Davis, P.J., Rabinowitz, P.: Methods of Numerical Integration, 2nd edn. Academic Press, San Diego (1984)
- Dellnitz, M., Froyland, G.: On the isolated spectrum of the Perron–Frobenius operator. *Nonlinearity* **13**, 1171–1188 (2000). <https://doi.org/10.1088/0951-7715/13/4/310>
- Dellnitz, M., Junge, O.: On the approximation of complicated dynamical behavior. *SIAM J. Numer. Anal.* **36**, 491 (1999). <https://doi.org/10.1137/S0036142996313002>
- Deyle, E.R., Sugihara, G.: Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* **6**(3), e18295 (2011). <https://doi.org/10.1371/journal.pone.0018295>
- Eisner, T., Farkas, B., Haase, M., Nagel, R.: Operator Theoretic Aspects of Ergodic Theory, Graduate Texts in Mathematics, vol. 272. Springer, Berlin (2015)
- Froyland, G.: Dynamic isoperimetry and the geometry of Lagrangian coherent structures. *Nonlinearity* **28**, 3587–3622 (2015). <https://doi.org/10.1088/0951-7715/28/10/3587>
- Genton, M.C.: Classes of kernels for machine learning: a statistics perspective. *J. Mach. Learn. Res.* **2**, 299–312 (2001)
- Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* **62**(2), 338–396 (2019). <https://doi.org/10.1016/j.acha.2017.09.001>
- 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)
- Giannakis, D., Majda, A.J.: Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis. *Geophys. Res. Lett.* **39**, L10710 (2012). <https://doi.org/10.1029/2012GL051575>

32. 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>
33. Giannakis, D., Majda, A.J.: Nonlinear Laplacian spectral analysis: capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. *Stat. Anal. Data Min.* **6**(3), 180–194 (2013). <https://doi.org/10.1002/sam.11171>
34. Giannakis, D., Ourmazd, A., Slawinska, J., Zhao, Z.: Spatiotemporal pattern extraction by spectral analysis of vector-valued observables. *J. Nonlinear Sci.* **29**(5), 2385–2445 (2019). <https://doi.org/10.1007/s00332-019-09548-1>
35. Halmos, P.R.: *Lectures on Ergodic Theory*. 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. Karrasch, D., Keller, J.: A geometric heat-flow theory of Lagrangian coherent structures. *J. Nonlinear Sci.* **30**, 1849–1888 (2020). <https://doi.org/10.1007/s00332-020-09626-9>
38. 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>
39. Koopman, B.O., von Neumann, J.: Dynamical systems of continuous spectra. *Proc. Natl. Acad. Sci.* **18**(3), 255–263 (1931). <https://doi.org/10.1073/pnas.18.3.255>
40. Korda, M., Putinar, M., Mezić, I.: Data-driven spectral analysis of the Koopman operator. *Appl. Comput. Harmon. Anal.* **48**(2), 599–629 (2020). <https://doi.org/10.1016/j.acha.2018.08.002>
41. Kosambi, D.D.: Statistics in function space. *J. Ind. Math. Soc.* **7**, 76–88 (1943)
42. Law, K., Shukla, A., Stuart, A.M.: Analysis of the 3DVAR filter for the partially observed Lorenz'63 model. *Discrete Contin. Dyn. Syst.* **34**(3), 1061–10178 (2013). <https://doi.org/10.3934/dcds.2014.34.1061>
43. Lorenz, E.N.: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963)
44. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. *Ann. Stat.* **26**(2), 555–586 (2008). <https://doi.org/10.1214/009053607000000640>
45. Luzzatto, S., Melbourne, I., Paccaut, F.: The Lorenz attractor is mixing. *Commun. Math. Phys.* **260**(2), 393–401 (2005)
46. Majda, M., McLaughlin, D.W., Tabak, E.G.: A one-dimensional model for dispersive wave turbulence. *J. Nonlinear Sci.* **6**, 9–44 (1997). <https://doi.org/10.1007/BF02679124>
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: spectral methods. In: *Proceedings of the 39th IEEE Conference on Decision and Control*, pp. 1224–1231. IEEE, Sydney, Australia (1999). <https://doi.org/10.1109/CDC.2000.912022>
49. Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior. *Phys. D* **197**, 101–133 (2004). <https://doi.org/10.1016/j.physd.2004.06.015>
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. Robinson, J.C.: A topological delay embedding theorem for infinite-dimensional dynamical systems. *Nonlinearity* **18**(5), 2135–2143 (2005). <https://doi.org/10.1088/0951-7715/18/5/013>
52. Sauer, T.: Time series prediction by using delay coordinate embedding. In: Weigend, A.S., Gerhsenfeld, N.A. (eds.) *Time Series Prediction: Forecasting the Future and Understanding the Past*. SFI Studies in the Sciences of Complexity, vol. 15, pp. 175–193. Addison-Wesley, Boston (1993)
53. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. *J. Stat. Phys.* **65**(3–4), 579–616 (1991). <https://doi.org/10.1007/bf01053745>
54. Slawinska, J., Giannakis, D.: Indo-Pacific variability on seasonal to multidecadal time scales. Part I: intrinsic SST modes in models and observations. *J. Clim.* **30**(14), 5265–5294 (2017). <https://doi.org/10.1175/JCLI-D-16-0176.1>
55. Sprott, J.C.: *Chaos and Time-Series Analysis*. Oxford University Press, Oxford (2003)
56. Steinwart, I.: On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.* **2**, 67–93 (2001)
57. Stone, M.H.: On one-parameter unitary groups in Hilbert space. *Ann. Math.* **33**(3), 643–648 (1932)
58. Székely, E., Giannakis, D., Majda, A.J.: Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data. *Clim. Dyn.* **46**(5), 1473–1502 (2016). <https://doi.org/10.1007/s00382-015-2658-2>
59. Takens, F.: Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence. Lecture Notes in Mathematics*, vol. 898, pp. 366–381. Springer, Berlin (1981). <https://doi.org/10.1007/bfb0091924>
60. Trillos, N.G., Gerlach, M., Hein, M., Slepčev, D.: Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace–Beltrami operator. *Found. Comput. Math.* (2019). <https://doi.org/10.1007/s10208-019-09436-w>. In press
61. Trillos, N.G., 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>
62. Tucker, W.: The Lorenz attractor exists. *C. R. Acad. Sci. Paris Ser. I* **328**, 1197–1202 (1999)
63. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Phys. D* **35**, 395–424 (1989). [https://doi.org/10.1016/0167-2789\(89\)90077-8](https://doi.org/10.1016/0167-2789(89)90077-8)
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.* **25**(6), 1307–1346 (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>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.