

Spatiotemporal pattern extraction by spectral analysis of vector-valued observables

Dimitrios Giannakis

Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

Abbas Ourmazd and Joanna Slawinska

Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

Zhizhen Zhao

*Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

(Dated: November 21, 2017)

We develop a data-driven framework for extracting complex spatiotemporal patterns generated by ergodic dynamical systems. Our approach, called Vector-valued Spectral Analysis (VSA), is based on an eigendecomposition of a kernel integral operator acting on a Hilbert space of vector-valued observables of the system, taking values in a space of functions (scalar fields) on a spatial domain. This operator is constructed by combining aspects of the theory of operator-valued kernels for machine learning with delay-coordinate maps of dynamical systems. Specifically, delay-coordinate maps performed pointwise in the spatial domain induce an operator acting on functions on that domain for each pair of dynamical states. Unlike conventional eigendecomposition techniques, which decompose the input data into pairs of temporal and spatial modes with a separable, tensor product structure, the patterns recovered by VSA can be manifestly non-separable, requiring only a modest number of modes to represent signals with intermittency in both space and time. In addition, our kernel construction naturally quotients out any symmetries present in the data. We show that in the limit of infinitely many delays the kernel integral operator employed in our scheme commutes with a Koopman operator governing the evolution of vector-valued observables under the dynamics; as a result, in that limit our recovered patterns lie in simultaneous eigenspaces of these operators associated with the point spectrum of the dynamical system. We present applications of VSA to the Kuramoto-Sivashinsky model, which demonstrate considerable performance gains in efficient and meaningful decomposition over eigendecomposition techniques utilizing scalar-valued kernels.

I. INTRODUCTION

Spatiotemporal pattern formation is ubiquitous in physical, biological, and engineered systems, ranging from molecular-scale reaction-diffusion systems, to engineering- and geophysical-scale convective flows, and astrophysical flows, among many examples [1–4]. The mathematical models for such systems are generally formulated by means of partial differential equations (PDEs), or coupled ordinary differential equations, with dissipation playing an important role in the development of low-dimensional effective dynamics on attracting subsets of the state space [5–7]. In light of this property, many pattern forming systems are amenable to analysis by empirical, data-driven techniques, complementing the scientific understanding gained from first-principles approaches.

Historically, many of the classical proper orthogonal decomposition (POD) techniques for spatiotemporal pattern extraction have been based on the spectral properties of temporal and spatial covariance operators estimated from snapshot data [8, 9]. In Singular Spectrum Analysis (SSA) and related algorithms [10, 11], the combination of this approach with delay-coordinate maps of dynamical systems [12–16] generally improves the representation of the information content of the data in terms of a few meaningful modes. More recently, advances in

machine learning and applied harmonic analysis [17–29] have led to the development of techniques recovering temporal and spatial patterns through the eigenfunctions of kernel integral operators (e.g., heat operators) defined intrinsically in terms of a Riemannian geometric structure of the data. In particular, in a family of techniques called Nonlinear Laplacian Spectral Analysis (NLSA) [30–32], as well as in independent work in [33], the diffusion maps algorithm [20] was combined with delay-coordinate maps to extract spatiotemporal patterns through the eigenfunctions of a kernel integral operator adept at capturing distinct and physically meaningful timescales in individual eigenmodes from multiscale high-dimensional signals.

At the same time, spatial and temporal patterns have been extracted from eigenfunctions of Koopman [34–39] and Perron-Frobenius [40, 41] operators governing the evolution of observables and probability measures, respectively, in dynamical systems [42–44]. In [36], it was shown that the dynamic mode decomposition (DMD) algorithm [45, 46] can, under some conditions, approximate spatial patterns called Koopman modes, corresponding to projections of the data onto eigenfunctions of the Koopman operator. The DMD algorithm can be implemented as a singular value decomposition of a linear operator relating time-shifted snapshots to a reference snapshot dataset. As shown in [37], this has connections with linear inverse modeling (LIM) techniques [47]. In [48], an extended formulation of DMD was developed using as an

approximation space a dictionary of functions providing greater flexibility and stronger convergence guarantees than the snapshot-based dictionary employed in DMD. In [49–51], an approximation scheme targeting the generator of the Koopman group (as opposed to Koopman operators themselves) was developed using a data-driven orthonormal basis of the Hilbert space of the dynamical system acquired through diffusion maps.

In general, Koopman and Perron-Frobenius techniques target operators defined intrinsically for the dynamical system generating the data, and thus can, in principle, recover temporal and spatial patterns of higher physical interpretability and utility in predictive modeling than POD and kernel integral operator based approaches. In practice, however, the Koopman and Perron-Frobenius operators tend to have significantly more complicated spectral properties (e.g., non-isolated eigenvalues and/or continuous spectrum), hindering the stability and convergence of data-driven approximation techniques. In [50, 51] these issues were addressed at least in the case of systems with pure point spectra and mixed spectra of their Koopman operators. There, it was shown that the eigenfunctions recovered by kernel integral operators defined on delay-coordinate mapped data (e.g., the covariance and heat operators in SSA and NLSA algorithms, respectively) in fact converge to Koopman eigenfunctions in the limit of infinitely many delays, indicating a deep connection between these two major branches of data analysis algorithms.

All of the techniques described above recover from the data set of temporal patterns and a corresponding set of spatial patterns, referred to in [8] as “chronos” and “topos” modes, respectively. In particular, for a dynamical system with a state space X developing patterns in a physical domain Y , each chronos mode, ϕ_k , corresponds to a scalar- (real- or complex-) valued function on X , and the corresponding topos mode, ψ_k , corresponds to a scalar-valued function on Y . Spatiotemporal reconstructions of the data with these approaches thus correspond to linear combinations of tensor product patterns of the form $\phi_k\psi_k$, mapping pairs of points (x, y) in the product space $\Omega = X \times Y$ to the number $\phi_k(x)\psi_k(y)$. For a dynamical system possessing a compact invariant set $A \subseteq X$ with an ergodic invariant measure (e.g., a dissipative system) and states sampled “on the attractor”, the chronos modes become scalar-valued functions on A , which may be of significantly smaller dimension than X , increasing the feasibility of robust approximation of these modes from finite datasets. As discussed below, if the system additionally possesses a so-called physical measure [52], these patterns can also be approximated from data lying near the invariant set, as would be the case in a typical experimental scenario.

Evidently, for spatiotemporal patterns $f(x, y)$ of high complexity, tensor product patterns of the form $\phi_k(x)\psi_k(y)$ with separable dependence on x and y can be highly inefficient in capturing the properties of the input signal. That is, the number of such patterns needed to

recover f via a linear superposition $\sum_k \phi_k\psi_k$ would generally be large, and none of the individual patterns $\phi_k\psi_k$ would be representative of f . In essence, the problem is similar to that of approximating a non-separable space-time signal in a tensor product basis of temporal and spatial basis functions. Another issue with tensor product decompositions based on scalar-valued eigenfunctions, which has received considerable attention [9, 53] is that in the presence of nontrivial spatial symmetries, the recovered patterns are oftentimes pure symmetry modes (e.g., Fourier modes in a periodic domain with translation invariance), with minimal dynamical significance and physical interpretability.

Here, we present a framework for spatiotemporal pattern extraction, called Vector-valued Spectral Analysis (VSA), designed to alleviate the shortcomings mentioned above. Instead of producing separable space-time modes, our approach recovers patterns in a Hilbert space of vector-valued observables assigning a scalar field (spatial pattern) on Y for every dynamical state in A , without imposing a separable structure. We recover data-driven patterns in this space through the eigenfunctions of a kernel integral operator, P_Q , constructed using the theory of operator-valued kernels [54–56]. As in NLSA algorithms, P_Q operates on data in delay-coordinate space with Q delays, and has analogous relationships with Koopman operators as those established in [50, 51]. However, unlike NLSA and other techniques recovering scalar-valued temporal patterns, the kernel p_Q of our integral operator is a function of both the state of the dynamical system in A and the spatial position in Y ; in particular, for every pair of dynamical states in A , there is a bounded linear operator acting on functions on Y (as opposed to a real number as in the case of NLSA and comparable algorithms) constructed from p_Q .

As we will show, every eigenfunction of P_Q can be equivalently characterized as a function-valued function on A , taking values in a Hilbert space of scalar functions on Y , or as a scalar-valued function on the set $B_Q \subseteq \mathbb{R}^Q$ of delay-embedded sequences corresponding to all dynamical states in A and spatial points in Y . In particular, B_Q can be understood as the base space of a topological bundle with the product space $M = A \times Y \subseteq \Omega$ acting as the total space and the observation function F_Q in delay-coordinate space as the projection map. This structure naturally quotients out any spatiotemporal symmetries present in the delay-embedded data (which may depend on Q), reducing the dimension of M when such symmetries are present. Moreover, the eigenfunctions of P_Q can have a manifestly non-separable structure. In fact, we will see that for small to moderate numbers of delays the leading non-constant eigenspace of P_Q can capture a large portion of the input signal. Thus, our kernel acts as a matched filter for the spatiotemporal signal generated by the dynamical system. This property has useful applications requiring estimation of level sets of functions from noisy data, such as topological data analysis [57], and is mathematically related to nonlocal averaging tech-

niques in image processing. Besides spatiotemporal data, this bundle construction may be useful in other scenarios, such as analysis of data generated by dynamical systems with varying parameters [58].

Another limit of interest corresponds to infinitely many delays, $Q \rightarrow \infty$, where a result that generalizes the results in [50, 51] is found to hold. Namely, as $Q \rightarrow \infty$, K_Q and P_Q converge to operators, K_∞ and P_∞ , respectively, that commute with the lift of the Koopman operator of the dynamical system on M . This result implies that K_∞ and P_∞ have common eigenfunctions with the lifted Koopman operator, and thus act as filters recovering the component of the signal associated with the discrete Koopman spectrum. We demonstrate the utility of our approach in the context of the Kuramoto-Sivashinsky (KS) model [59, 60] in chaotic regimes.

The plan of this paper is as follows. In Section II, we introduce the class of systems under study, and establish our notation. Section III presents the VSA framework for spatiotemporal pattern extraction using operator-valued kernels. Section IV discusses certain connections between our framework and the theory of topological bundles, which are useful for gaining intuition on the behavior of our technique, particularly in the presence of symmetries. In Section V, we examine the asymptotic behavior of VSA in the limit of no delays or infinitely-many delays, and establish connections with nonlocal averaging and Koopman operators, respectively. Section VI describes the data-driven implementation of our techniques. That section includes Theorem 6 and Corollary 7, which establish convergence of our data-driven algorithms to their counterparts in section III. In Section VII, we present our applications to the KS system. Our primary conclusions are described in Section VIII. The paper includes four appendices containing a summary of the properties of Koopman eigenfunctions and technical results. All numerical results presented here were obtained via a Matlab code which is available for download from the first author's personal website, <http://cims.nyu.edu/~dimitris>.

II. NOTATION AND PRELIMINARIES

Let $\Phi^t : X \mapsto X$, $t \in \mathbb{R}$, be a continuous-time flow on a differentiable (of class C^1) manifold X , generated by a Lipschitz vector field $v : C^1(X) \mapsto C(X)$, possessing a compact invariant set $A \subseteq X$ and an ergodic invariant probability measure μ with support equal to A . As with every such manifold, X can be equipped with a metric $\bar{d}_X : X \times X \mapsto \mathbb{R}_+$; in what follows we will not require an explicit specification of \bar{d}_X , but its existence will be used in the proof of Theorem 6. Let also (Y, \bar{d}_Y) be a compact, connected metric space equipped with a Borel probability measure ν . In what follows, (X, Φ^t) will be the dynamical system generating the data and Y the spatial domain in which measurements are taken. The measure ν provides a notion of normalized volume on Y . In addition, we

consider that we observe the system at a fixed sampling time interval $\tau > 0$ through a function $\bar{F} \in C(X; C(Y))$; that is, for every $x \in X$, $F_x = \bar{F}(x)$ is a continuous scalar-valued function on Y , and the dependence of F_x on x is continuous. Associated with \bar{F} is a continuous scalar-valued F function on $\Omega = X \times Y$ such that $F(x, y) = F_x(y)$. The product space Ω can be equipped with a metric $\bar{d}_\Omega : \Omega \mapsto \Omega$ induced by \bar{d}_X and \bar{d}_Y .

The assumptions stated above are sufficient to ensure that the VSA framework introduced below is theoretically well posed. Additional assumptions are needed to ensure that our scheme is amenable to data-driven approximation using finite numbers of sampled points in X and Y . These assumptions will be stated in Section VI A.

Next, we introduce groups of Koopman operators [42–44], governing the evolution of scalar and vector-valued observables under the dynamics on X . For that, we first specify appropriate function spaces. In the case of scalar-valued observables, $f : X \mapsto \mathbb{C}$, the standard space to consider is the Hilbert space $L^2(A, \mu)$ associated with the invariant measure. Hereafter, we abbreviate our notation for this space using $H_A = L^2(A, \mu)$, and also use the notations $\langle f, g \rangle_{H_A} = \int_A f^* g d\mu$ and $\|f\|_{H_A} = \langle f, f \rangle_{H_A}^{1/2}$ for its inner product and norm, respectively. Similarly, we consider scalar-valued observables on the spatial domain Y and the product space $M = A \times X$ lying in the Hilbert spaces $H_Y = L^2(Y, \nu)$ and $H_M = L^2(M, \rho)$, respectively, where $\rho = \mu \times \nu$ is the product probability measure between the invariant measure of the dynamics and the volume measure of the spatial domain. The inner products and norms of these spaces are defined analogously to those of H_A , and will be denoted using H_Y and H_M subscripts, respectively. Because the measures μ and ν are both finite (and thus H_A and H_Y are both separable), H_M is isomorphic to the tensor product space $H_A \otimes H_Y$; that is, every element f of H_M can be expressed as a countable sum $f = \sum_{k=0}^{\infty} g_k h_k$ with $g_k \in H_A$ and $h_k \in H_Y$.

The Koopman operator on H_A at time $t \in \mathbb{R}$ is the unitary, bounded operator $U^t : H_A \mapsto H_A$ that acts on observables via composition with the flow map, namely $U^t f = f \circ \Phi_t$, or equivalently, $U^t f(x) = f(\Phi^t(x))$, where the equality holds up to null sets in A with respect to μ . Note that the unitarity of U^t , $U^{t*} U^t = U^t U^{t*} = I_{H_A}$, is a consequence of Φ^t preserving μ , and we also have $U_t^{-1} = U_{-t}$. The Koopman operators on H_A trivially lift to the unitary Koopman operators $\tilde{U}^t = U^t \otimes I_{H_Y}$ on H_M , which are composition operators associated with the flow $\tilde{\Phi}^t = \Phi^t \otimes I_Y$. Note that “dynamics” on the spatial domain Y are trivial under this flow, as they are described by the identity map I_Y . Moreover, $\tilde{\Phi}^t$ is manifestly non-ergodic, as every set of the form $A \times S$ with $\nu(S) > 0$ is $\tilde{\Phi}^t$ -invariant and has positive μ -measure. Further details on the properties of Koopman operators are included in Appendix A.

Koopman operators give rise to a distinguished class of observables through their eigenfunctions. In the case

of U^t , the Koopman eigenvalue problem reads

$$U^t z = \Lambda^t z, \quad \Lambda^t \in \mathbb{C}, \quad z \in H_A.$$

As discussed in Appendix A, in ergodic dynamical systems the eigenvalues Λ^t are simple, lie on the unit circle, and take the form $\Lambda^t = e^{i\alpha t}$, where α are real frequencies. Thus, under the action of U^t , every such eigenfunction evolves by multiplication by a periodic phase factor; this makes Koopman eigenfunctions highly predictable observables. In addition, Koopman eigenfunctions have the important property of being intrinsic to the dynamical system; i.e., they do not depend on additional structures such as a Riemannian geometry often utilized in manifold learning algorithms. In measure-preserving systems, Koopman eigenfunctions corresponding to distinct eigenvalues are orthogonal by unitarity of U^t . Hereafter, we will assume that all eigenfunctions of U^t are continuous; while this is not necessarily the case, we are not aware of any continuous-time flows with discontinuous Koopman eigenfunctions (though discrete-time dynamical systems with discontinuous Koopman eigenfunctions are known to exist [61]).

The eigenvalue problem for the lifted Koopman operator on H_M ,

$$\tilde{U}^t \tilde{z} = \tilde{\Lambda}^t \tilde{z}, \quad \tilde{\Lambda}^t \in \mathbb{C}, \quad \tilde{z} \in H_M, \quad (1)$$

is structurally very similar to that for U^t ; in fact, it is a direct consequence of the definition of \tilde{U}^t that its eigenvalues are the same as those of U^t . However, in contrast to the one-dimensional eigenspaces of U^t , the eigenspaces of \tilde{U}^t are all infinite dimensional. In particular, given any eigenfunction of U^t at eigenvalue Λ^t and any spatial pattern $\psi \in H_Y$, the function $\tilde{z} = z\psi$ is an eigenfunction of \tilde{U}^t at the same eigenvalue.

Here, in addition to spaces of scalar-valued observables, we will be interested in vector-valued observables lying in the space $H = L^2(A, \mu; H_Y)$. This space consists of (equivalence classes of) functions on A taking values in H_Y , which are square-integrable with respect to the invariant measure μ . That is, every observable $f \in H$ satisfies $\int_X \|f(x)\|_{H_Y}^2 d\mu(x) < \infty$. H is a Hilbert space for the inner product $\langle f, g \rangle_H = \int_A \langle f(x), g(x) \rangle_{H_Y} d\mu(x)$. Given a point $x \in A$ (an initial condition), we can associate a spatiotemporal pattern to every continuous observable $f \in H$ through the relationship

$$t \mapsto f(\Phi^t(x)), \quad t \in \mathbb{R}. \quad (2)$$

This generalizes the standard notion of temporal patterns (including the chronos modes) constructed from scalar valued observables in H_A or H_M via analogous relationships to (2). As with the isomorphism between H_M and $H_A \otimes H_Y$ noted above, the finiteness of the measures μ and ν implies the isomorphisms $H \simeq H_M \simeq H_A \otimes H_Y$. Thus, we can equivalently characterize the observables of interest in this work either as vector-valued functions in H , scalar-valued functions in H_M , or elements of the tensor product space $H_A \otimes H_Y$.

III. VECTOR-VALUED SPECTRAL ANALYSIS FORMALISM

We now describe the VSA framework for spatiotemporal pattern extraction through operator-valued kernels. Due to the isomorphism between H and H_M stated in Section II, we first introduce our approach through a scalar-valued kernel construction on the product space M , and then pass to an equivalent operator-valued kernel formulation on the invariant set A .

A. Scalar-valued kernels and their associated integral operators

For the purposes of this work, a (scalar-valued) kernel is a function $k : \Omega \times \Omega \mapsto \mathbb{R}_+$ mapping pairs of points in Ω to a positive real number. Intuitively, k can be thought of as a measure of similarity between points in Ω . Mathematically, we require that it has the following properties:

1. k is continuous on $\Omega \times \Omega$. Hence, restricted to any compact subset of $\Omega \times \Omega$ (in particular, $M \times M$), it is bounded.
2. Restricted to $M \times M$, k is bounded away from zero; that is, there exists a constant $c > 0$ such that $k(\omega, \omega') \geq c$ for all $\omega, \omega' \in M$.

Sometimes, but not always, we consider symmetric kernels, i.e., kernels k having the additional property $k(\omega, \omega') = k(\omega', \omega)$ for all $\omega, \omega' \in \Omega$. It should be noted that while the assumptions above are sufficient to ensure the theoretical well-posedness of our schemes, in Section VI below we will make additional assumptions in order to construct convergent data-driven approximations to these schemes. In particular, we will require that k be continuous on a compact superset of M of positive Lebesgue measure.

Associated with every kernel k is an integral operator $K : H_M \mapsto H_M$ defined via $K : f \mapsto Kf$, where

$$Kf = \int_M k(\cdot, \omega) f(\omega) d\rho(\omega). \quad (3)$$

By continuity of k , K is a compact operator whose range is included in the space of continuous functions on M . If, in addition, k is symmetric, K is also self-adjoint. As a result, in the symmetric case, the eigenfunctions of K are continuous, and its eigenvalues are real, bounded, and can only accumulate at zero. Moreover, the nonzero eigenvalues of K have finite multiplicities (i.e., the corresponding eigenspaces are finite-dimensional).

B. Kernels from pseudometric functions

We will focus on kernels constructed from scaled pseudometric functions. Below, a pseudometric will be a symmetric function $d : \Omega \times \Omega \mapsto \mathbb{R}$, continuous on $M \times M$,

and having the properties $d(\omega, \omega') \geq 0$ and $d(\omega, \omega) = 0$ for all $\omega, \omega' \in \Omega$. Note that $d(\omega, \omega')$ can vanish even if $\omega \neq \omega'$, which would not be the case for a proper metric. Moreover, d is not required here to obey the triangle inequality.

We first consider the class of pseudometric functions d_Q , $Q \in \mathbb{N}_+$, obtained from delay-coordinate mapped data with Q delays. Specifically, given $\omega = (x, y)$ and $\omega' = (x', y')$ with $x, x' \in X$ and $y, y' \in Y$, we define

$$d_Q^2(\omega, \omega') = \frac{1}{Q} \sum_{q=0}^{Q-1} |F(\Phi^{-q\tau}(x), y) - F(\Phi^{-q\tau}(x'), y')|^2.$$

Note that $d_Q^2(\omega, \omega')$ corresponds to a time-averaged squared pseudo-distance, evaluated by fixing the spatial points $y, y' \in Y$, and sampling the observation map F on Q temporally-equispaced points along the dynamical trajectories starting from $x, x' \in X$.

We will also be interested in the limiting behavior of d_Q as the number of delays increases, viz.

$$d_\infty(\omega, \omega') = \lim_{Q \rightarrow \infty} d_Q(\omega, \omega'). \quad (4)$$

Unlike d_Q at finite Q , d_∞ may not be continuous on $M \times M$. However, it can be shown that it is a well-defined function in $L^2(M \times M, \rho \times \rho)$; in particular, the limit in (4) exists for almost every pair of points $(\omega, \omega') \in M \times M$ with respect to the product measure $\rho \times \rho$ [51]. Moreover, $d_\infty(\omega, \omega')$ may vanish despite $d_Q(\omega, \omega')$ being non-vanishing (e.g., if ω' lies in the stable manifold of ω). Additional details on the properties of d_∞ can be found in Appendix C.

Next, as is common in machine learning algorithms [26–28, 31, 62] (including NLSA), we introduce a positive scaling function $s_Q : \Omega \mapsto \mathbb{R}$, and form a rescaled pseudometric $\tilde{d}_Q : \Omega \times \Omega \mapsto \mathbb{R}$, given by

$$\tilde{d}_Q(\omega, \omega') = d_Q(\omega, \omega') \sqrt{s_Q(\omega) s_Q(\omega')}. \quad (5)$$

We require that, restricted to M , s_Q is continuous (and therefore bounded). Under these conditions, \tilde{d}_Q has the required properties of pseudometrics stated above, and we can similarly study the limit

$$\tilde{d}_\infty = \lim_{Q \rightarrow \infty} \tilde{d}_Q; \quad (6)$$

see Appendix C for additional details.

Intuitively, one can think of s_Q as a modification of the original pseudometric d_Q to account for variations in the sampling density and/or time tendency (“velocity”) of the data. For instance, we can choose s_Q such that it is large in regions of Ω where the sampling density is large, and small in regions where the sampling density is small. With this choice, \tilde{d}_Q becomes relatively smaller (larger) than d_Q for pairs of points in regions of Ω with small (large) sampling density. As a result, \tilde{d}_Q adaptively provides higher “resolution” in regions where we

have plentiful observations and smaller resolution in regions where the observations are sparse. Bandwidth functions employed in the literature include functions based on near-neighbor distances [28, 62], local time tendencies of the data [26, 31], and kernel density estimates [27, 28]. In certain cases involving data on smooth Riemannian manifolds, the transformation $d_Q \mapsto \tilde{d}_Q$ associated with these techniques can be interpreted as a conformal change of metric [26, 28]; that is, a change of volume on the data manifold without change of angles.

Here, we employ a class of scaling functions introduced in [50], which has both velocity- and density-dependent contributions. Explicit formulas and discussion on this class of bandwidth functions are included in Appendix B. Note that in [26, 28, 31, 62] the convention is to work with rescaled distance functions of the form

$$\tilde{d}_Q(\omega, \omega') = \frac{d_Q(\omega, \omega')}{\sqrt{s'_Q(\omega) s'_Q(\omega')}}, \quad (7)$$

which is equivalent to (5) with $s'_Q = 1/s_Q$, so long as s_Q is bounded away from zero. Here, we have opted to work with (5) in order to allow for the possibility that s_Q vanishes. As we will see below, this may be useful in the context of spatiotemporal data where the signal can exhibit constant or near-constant values in certain parts of the domain Y (e.g., if F is the temperature field in a convection flow, and held to fixed values at the boundary of Y).

Having specified the class of pseudometrics \tilde{d}_Q , we also select a continuous shape function $h : \mathbb{R} \mapsto \mathbb{R}_+$, and define

$$k_Q(\omega, \omega') = h(\tilde{d}_Q(\omega, \omega')), \quad k_\infty(\omega, \omega') = h(\tilde{d}_\infty(\omega, \omega')).$$

By symmetry of pseudometrics, k_Q and k_∞ are both symmetric kernels. In what follows, we will always use Gaussian shape functions, $h(u) = e^{-u^2/\epsilon}$, parameterized by a positive bandwidth parameter ϵ . Explicitly, we have

$$k_Q(\omega, \omega') = e^{-\tilde{d}_Q^2(\omega, \omega')/\epsilon}, \quad k_\infty(\omega, \omega') = e^{-\tilde{d}_\infty^2(\omega, \omega')/\epsilon}. \quad (8)$$

We denote the kernel integral operators from (3) associated with k_Q and k_∞ by K_Q and K_∞ , respectively. Note that in the context of exponentially decaying kernels such as (8) the function s'_Q from (7) can be interpreted as a bandwidth function since, for a fixed base point $\omega \in \Omega$, a large (small) value $s'_Q(\omega)$ increases (decreases) the size of the neighborhood where $k_Q(\omega, \cdot)$ is appreciably greater than zero.

The Gaussian class of shape functions has been widely used in machine learning applications [17–20, 29], including NLSA. Among their attractive properties is their localizing behavior at small ϵ (not true for quadratic shape functions associated with the covariance kernels employed in POD), which allows the associated kernel integral operators to approximate, as $\epsilon \rightarrow 0$, heat kernels on Riemannian manifolds. The latter have, in turn, a

number of useful properties for dimension reduction and clustering of data with a manifold geometric structure. The exponential decay of k_Q and k_∞ also implies that K_Q and K_∞ can be well approximated by sparse matrices. The latter property is particularly important in the problems studied here involving spatiotemporal data, as the number of sampled states in X and spatial points in Y are generally both large (see Section VI). It is important to note that in the present work we do not assume that M is a smooth manifold (since for nonlinear dissipative systems with fractal attractors M is highly non-smooth), so we do not consider the limiting behavior of our kernels as $\epsilon \rightarrow 0$.

C. Markov normalization

Following the approach taken in NLSA algorithms and in [33, 49–51], we construct from K_Q and K_∞ associated compact Markov operators, $P_Q : H_M \mapsto H_M$ and $P_\infty : H_M \mapsto H_M$, respectively, by applying the normalization procedure introduced in the diffusion maps algorithm [20] and further developed in the context of general exponentially decaying kernels in [29]. Below, we describe the procedure for a general kernel integral operator K of the class in (3) associated with a kernel k .

We begin by defining the functions

$$r = K1_M, \quad l = K(1_M/r)$$

where 1_M is the function on M everywhere equal to 1. By the properties of kernels listed in section III A, both of the above functions are continuous, positive functions on M , bounded away from zero. We then define the kernel integral operator P (again of the class in (3)), associated with the kernel p , where

$$p(\omega, \omega') = \frac{k(\omega, \omega')}{l(\omega)r(\omega')}. \quad (9)$$

In [29], the division of $k(\omega, \omega')$ by $l(\omega)$ and $r(\omega')$ to form $p(\omega, \omega')$ is referred to as left and right normalization, respectively. Because r and l are both positive and bounded away from zero (which is in turn a consequence of property 2 of k), p meets both conditions on kernels stated in Section III A. Moreover, it has the Markov property,

$$\int_M p(\omega, \cdot) d\rho = 1, \quad \forall \omega \in M,$$

by construction. As result, P is a compact Markov operator on H_M , preserving constant functions. In addition, since p is bounded away from zero, P is ergodic; i.e., it has a simple eigenvalue $\lambda_0 = 1$, and the corresponding eigenfunction, ϕ_0 , is constant. Here, we are interested in the kernels p_Q and p_∞ , and the corresponding Markov operators P_Q and P_∞ , constructed by applying the diffusion maps normalization procedure to k_Q and k_∞ , respectively.

In general, the kernels of the class in (9), including p_Q and p_∞ , are not symmetric, and as a result the corresponding Markov operators are not self-adjoint. However, for a kernel p constructed by applying the diffusion maps normalization procedure to a symmetric kernel k (including the kernels introduced in Section III A), the corresponding operator P is related to a self-adjoint compact operator, $\hat{P} : H_M \mapsto H_M$ by a similarity transformation. In particular, let f be a bounded function in $L^\infty(M, \rho)$, and $T_f : H_M \mapsto H_M$ the corresponding multiplication operator by f . That is, for $g \in H_M$, $T_f g$ is the function equal to $f(\omega)g(\omega)$ for ρ -a.e. $\omega \in M$. Defining S as the self-adjoint kernel integral operator from (3) with the symmetric kernel

$$\hat{p}(\omega, \omega') = \frac{k(\omega, \omega')}{\hat{d}(\omega)\hat{d}(\omega')}, \quad \hat{d} = \sqrt{lr},$$

one can verify that \hat{P} can be obtained from P through the similarity transformation

$$\hat{P} = T_{\hat{d}} \circ P \circ T_{\hat{d}}^{-1}, \quad \hat{d} = \sqrt{lr}. \quad (10)$$

Due to (10), P and \hat{P} have the same eigenvalues λ_k , which are real by self-adjointness of \hat{P} , and admit the ordering $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots$, because P is an ergodic Markov operator. Moreover, the eigenvalues have the properties $|\lambda_j| \leq 1$, $\lim_{j \rightarrow \infty} \lambda_j = 0$, and the eigenspaces corresponding to the nonzero λ_j are finite-dimensional.

Since \hat{P} is self-adjoint, there exists an orthonormal basis $\{\hat{\phi}_j\}_{j=0}^\infty$ of H_M consisting of real eigenfunctions $\hat{\phi}_j$ of \hat{P} corresponding to the λ_j , which are continuous by the assumed continuity of kernels. Moreover, due to (10), for every element $\hat{\phi}_j$ of this basis, the continuous functions $\phi_j = T_{\hat{d}}\hat{\phi}_j = \hat{d}\hat{\phi}_j$ and $\phi'_j = T_{\hat{d}}^{-1}\hat{\phi}_j = \hat{\phi}_j/\hat{d}$ are eigenfunctions of P and P^* , respectively, corresponding to the same eigenvalue λ_j . The sets $\{\phi_j\}_{j=0}^\infty$ and $\{\phi'_j\}_{j=0}^\infty$ are (non-orthogonal) bases of H_M satisfying the bi-orthogonality relation $\langle \phi'_i, \phi_j \rangle_{H_M} = \delta_{ij}$. In particular, every $f \in H_M$ can be expanded as $f = \sum_{j=0}^\infty c_j \phi_j$ with $c_j = \langle \phi'_j, f \rangle_{H_M}$, and we have $Pf = \sum_{j=0}^\infty \lambda_j c_j \phi_j$. For notational simplicity, in what follows we will use the same symbols, λ_j ($\hat{\phi}_j$) to represent the eigenvalues (eigenfunctions) of either \hat{P}_Q or \hat{P}_∞ ; the symmetric operators constructed by applying (10) to P_Q and P_∞ , respectively. Moreover, the eigenfunctions of P_Q and P_∞ will both be denoted by ϕ_j .

D. Operator-valued kernels and the associated spatiotemporal patterns

Having constructed the kernel integral operators of section III B, we now make use of the isomorphism between H_M and the Hilbert space H of vector-valued observables to establish analogous operators associated with operator-valued kernels. Let $\mathcal{L}(H_Y)$ be the Banach space

of bounded linear operators on H_Y , equipped with the operator norm. For our purposes, an operator-valued kernel will be a continuous map $l : X \times X \mapsto \mathcal{L}(H_Y)$, mapping each pair of states $(x, x') \in X \times X$ to a bounded linear operator in $l(x, x') := L_{xx'} \in \mathcal{L}(H_Y)$. Note that we do not impose a positive-definiteness property employed in the definition of operator-valued kernels in [54–56]; that property is important in the construction of associated reproducing kernel Hilbert spaces of vector-valued observables, which we do not address in this work (though it would be an interesting topic for future work).

To every such operator-valued kernel there corresponds a kernel integral operator $L : H \mapsto H$ defined by $L : f \mapsto Lf$, where

$$Lf = \int_X l(\cdot, x')f(x') d\mu(x'), \quad x \in A, \quad (11)$$

and the above integral is a Bochner integral for Hilbert-space valued observables taking values in H_Y .

Given a scalar kernel k from Section III A, we can construct an associated operator-valued kernel \tilde{k} by observing that the functions $\tilde{k}_{xx'} : Y \times Y \mapsto \mathbb{R}_+$, defined for every $(x, x') \in X \times X$ as $\tilde{k}_{xx'}(y, y') = k((x, y), (x', y'))$, are kernels on the spatial domain Y . As a result, associated with every $k_{xx'}$ is a compact (hence bounded) kernel integral operator $\tilde{K}_{xx'} \in \mathcal{L}(H_Y)$, defined analogously to (3) via

$$\tilde{K}_{xx'}f = \int_Y \tilde{k}_{xx'}(\cdot, y')f(y') d\nu(y').$$

The operator-valued kernel \tilde{k} associated with k is then defined as $\tilde{k}(x, x') = \tilde{K}_{xx'}$. It follows by construction of \tilde{k} that if ϕ_j is an eigenfunction of the integral operator from (3) associated with k corresponding to the eigenvalue λ_j , then the vector-valued function $\vec{\phi}_j \in H$ defined as

$$\vec{\phi}_j(x)(y) = \phi_j((x, y)), \quad x \in A, \quad y \in Y, \quad (12)$$

is an eigenfunction of \tilde{K} corresponding to the same eigenvalue λ_j . Note that by continuity of k and \tilde{k} , both ϕ_j and $\vec{\phi}_j$ are continuous.

In this work, we extract spatiotemporal patterns using the eigenfunctions $\vec{\phi}_0, \vec{\phi}_1, \dots$ of the kernel integral operator \tilde{P}_Q of the class (11), whose kernel \tilde{p}_Q is the operator-valued kernel associated with the Markov kernel p_Q from Section III C. We also consider the analogous limiting operator \tilde{P}_∞ with kernel \tilde{p}_∞ associated with the Markov kernel p_∞ , characterizing the behavior of \tilde{P}_Q in the limit of infinitely many delays. As with their counterparts P_Q and P_∞ acting on scalar-valued functions in H_M , there exists a (non-orthogonal) basis $\{\vec{\phi}_j\}_{j=0}^\infty$ of H consisting of real eigenfunctions of \tilde{P}_Q related to those of P_Q via (12). Moreover, to each such eigenfunction $\vec{\phi}_j$ there corresponds a spatiotemporal pattern obtained via (2). In Section VII, we will demonstrate that these patterns can provide an efficient and physically meaningful decomposition of complex spatiotemporal signals.

IV. CONNECTIONS WITH TOPOLOGICAL BUNDLES

To gain intuition on the behavior of the kernel integral operators introduced in Section III, it is useful to examine a natural topological bundle structure that M acquires through the observation map in delay-coordinate space. First, we recall that a topological bundle is a triplet (E, B, π) , where E and B are topological spaces, called the total and base spaces, respectively, and $\pi : E \mapsto B$ is a continuous surjective map. The preimage $\pi^{-1}(b)$ of a point $b \in B$ is called the fiber of π over b . A topological bundle is called a fiber bundle if all fibers are homeomorphic to a reference fiber, R . If it happens that $E = B \times R$, the fiber (E, B, π) is said to be trivial. While not every fiber bundle will have this property, it is the case that fiber bundles are locally trivial; that is, for every $b \in B$ there exists a neighborhood $N \subseteq B$ such that $\pi^{-1}(N)$ is homeomorphic to $N \times R$, and π agrees with the natural projection onto the first factor of that Cartesian product; see, e.g., [63] for more details.

A. Bundles for finitely many delays

In what follows, we will be interested in topological bundles where the total space is the product space $M = A \times Y$, and the base space and the projection map are constructed by application of delay-coordinate maps to F . In particular, for $Q \in \mathbb{N}_+$, let $F_Q : \Omega \mapsto \mathbb{R}^Q$ be the continuous map defined by

$$F_Q((x, y)) = \left(F((x, y)), F((\Phi^{-\tau}(x)), y), \dots, F((\Phi^{-(q-1)\tau}(x)), y) \right).$$

Let also $B_Q = F_Q(M)$ be the image of M under F_Q , and $\pi_Q : M \mapsto B_Q$ the map defined uniquely by $\pi_Q(\omega) = F_Q(\omega)$ for all $\omega \in M$. Then, (M, B_Q, π_Q) is a topological bundle. Note that the invariant measure ρ on M induces a Borel measure $\rho_Q = \pi_{Q*}(\rho)$ on B_Q , where π_{Q*} is the pushforward map on measures associated with π_Q .

The usefulness of this construction is that we can express the pseudometric d_Q on M as a pullback of a Euclidean metric on B_Q . In particular, let $d'_Q : \mathbb{R}^Q \times \mathbb{R}^Q \mapsto \mathbb{R}$ be the Euclidean metric on \mathbb{R}^Q scaled by the square root of the number of delays, i.e.,

$$d'_Q(b, b') = \|b - b'\|_{\mathbb{R}^Q} / Q^{1/2}, \quad b, b' \in \mathbb{R}^Q.$$

Then, for all $\omega, \omega' \in \Omega$ we have

$$d_Q(\omega, \omega') = d'_Q(F_Q(\omega), F_Q(\omega')), \quad (13)$$

which expresses the fact that d_Q is a pullback of d'_Q . The same conclusion holds for the restrictions of d_Q and d'_Q to $M \times M$ and $B_Q \times B_Q$, respectively, as stated above. Now, as with every pseudometric, d_Q partitions Ω (and therefore M) into equivalence classes of points for which d_Q

vanishes. In particular, the equivalence class associated with $\omega \in \Omega$ is given by $[\omega]_Q = \{\omega' \in \Omega \mid d_Q(\omega, \omega') = 0\}$. It follows from (13) that the set of all such equivalence classes in M is homeomorphic (topologically equivalent) to B_Q ; that is, for every $\omega \in M$, the equivalence class $[\omega]_Q$ is given by the preimage $\pi_Q^{-1}(\pi_Q(\omega))$.

By similar arguments, we can also conclude that the kernel p_Q can be expressed as a pullback of the kernel $p'_Q : \mathbb{R}^Q \times \mathbb{R}^Q \mapsto \mathbb{R}_+$ constructed by applying the Markov normalization procedure described in Section III C to the kernel $k'_Q = e^{-d_Q^2/\epsilon}$ on $\mathbb{R}^Q \times \mathbb{R}^Q$. That is, we have

$$p_Q(\omega, \omega') = p'_Q(F_Q(\omega), F_Q(\omega')), \quad \omega, \omega' \in \Omega.$$

As a result, the eigenfunctions ϕ_j of P_Q correspond to pullbacks of the eigenfunctions of the kernel integral operator $P_Q : L^2(B_Q, \rho_Q) \mapsto L^2(B_Q, \rho_Q)$ whose kernel is p'_Q . Explicitly, we have

$$\phi_j = \varphi_j \circ \pi_Q, \quad (14)$$

where φ_j is an eigenfunction of P'_Q satisfying

$$P'_Q \varphi_j = \lambda_j \varphi_j, \\ P'_Q \varphi_j(b) = \int_{B_Q} p'_Q(b, b') \varphi_j(b') d\rho_Q(b'), \quad b \in B_Q.$$

In general, the “useful” patterns that can be extracted from a kernel integral operator are represented by eigenfunctions at nonzero corresponding eigenvalues (recall that such eigenfunctions lie in finite-dimensional eigenspaces, and thus can be robustly approximated from data). Therefore, it is useful to examine the closed subspace

$$\mathcal{H}_Q = \overline{\text{span}\{\phi_j : \lambda_j > 0\}} \subseteq H_M,$$

whose every element is expressible as a countable linear combination of such eigenfunctions of P_Q . Note that \mathcal{H}_Q can be equivalently defined as the closure of the range of P_Q , or as a pullback under π_Q of the closure of the range of P'_Q . In particular, whether or not \mathcal{H}_Q is a strict subspace of H_M depends on the number of delays Q , and importantly, it also depends on the presence of symmetries in the data.

To examine this in more detail, suppose that for every spatial point $y \in Y$ the observation map F_y is generic in the sense of delay-coordinate mappings [12–16]; that is, there exists a minimal positive integer Q_y^* such that for all $Q \geq Q_y^*$ the image $B_{Q,y} = F_Q(A_y)$ of the set $A_y = A \times \{y\} \subset M$ is homeomorphic to A . In that case, we can topologically reconstruct the attractor from time-lagged measurements taken at y . Assume now that $Q^* = \sup_{y \in Y} Q_y^*$ is finite, and observe that, for any Q , the base space B_Q of the bundle (M, B_Q, π_Q) is given by $B_Q = \cup_{y \in Y} B_{Q,y}$. For $Q \geq Q^*$, if the sets $B_{Q,y}$ are disjoint, B_Q is homeomorphic to $M = A \times Y$, and thus has a Cartesian product structure. In other words, in that case, (M, B_Q, π_Q) becomes a trivial bundle with A acting

as the reference fiber. Moreover, under additional conditions on the form of the kernel k_Q and the properties of the measure ρ , P_Q and P'_Q may have no zero eigenvalues, in which case $\mathcal{H}_Q = H_M$.

On the other hand, in a number of cases, including when nontrivial symmetries are present, some of the $B_{Q,y}$ for $Q \geq Q^*$ may intersect or overlap. In such scenarios, B_Q will generally not have a Cartesian product structure, and (M, B_Q, π_Q) may not be trivial (in fact, it may not even be a fiber bundle). Moreover, \mathcal{H}_Q will be a strict subspace of H_M . For data analysis purposes, this property may actually be desirable since functions in \mathcal{H}_Q are pullbacks of L^2 functions on B_Q , and the latter is in turn homeomorphic to a “smaller”, quotient space, constructed from the $[\cdot]_Q$ equivalence classes on M as described above. In effect, by factoring out redundancies in the data, functions on B_Q can be approximated more robustly than functions on M , while still capturing the salient properties of the signal.

B. Spatial symmetries

We now examine one particular scenario encountered frequently in applications leading to the map π_Q being non-invertible, namely the presence of spatial symmetries. Throughout this section, we consider that the state space manifold X is actually a subset of the Hilbert space H_Y . That would be the case, for instance, if H_Y is the solution space of a dissipative PDE defined on a spatial domain Y , and X is an inertial manifold of that PDE [7]; the KS system studied in Section VII is an example of such a system. In this setting, the vector-valued observation map $\vec{F} : X \mapsto C(Y)$ reduces to the inclusion map from X to H_Y . For notational simplicity, given $x \in X$, we will identify $\vec{F}(x)$ with x . We will also use the symbol \mathfrak{X} to represent the space of continuous vector fields on X .

1. Group actions and their corresponding invariant pseudometrics

Let G be a topological group with a left action on Y . By that, we mean that there exists a map $\Gamma_Y : G \times Y \mapsto Y$ with the following properties:

1. Γ_Y is continuous, and $\Gamma_Y(g, \cdot) : Y \mapsto Y$ is a homeomorphism for all $g \in G$.
2. Γ_Y is compatible with the group structure of G ; that is, for all $y \in Y$ and $g, g' \in G$,

$$\Gamma_Y(gg', y) = \Gamma_Y(g, \Gamma_Y(g', y)),$$

and $\Gamma_Y(e, y) = y$, where e is the identity element of G .

Henceforth, given $g \in G$, we will abbreviate the map $\Gamma_Y(g, \cdot) : Y \mapsto Y$ by Γ_Y^g . Note that the (continuous) inverse of this map is given by $\Gamma_Y^{g^{-1}}$.

The action of G on Y induces a continuous left action on H_Y such that the action map $\Gamma_{H_Y}^g : H_Y \mapsto H_Y$ sends $u \in H_Y$ to $u \circ \Gamma_Y^{g^{-1}}$. As usual, we will consider G to be a dynamical symmetry (e.g., [9]) if the following hold for all $g \in G$:

1. The state space manifold X is invariant under $\Gamma_{H_Y}^g$. Thus, we obtain a left group action Γ_X^g on X by restriction of Γ_{H_Y} .
2. Γ_X^g is differentiable for all $g \in G$, and the dynamical vector field v is invariant under the pushforward map $\Gamma_{X^*}^g : \mathfrak{X} \mapsto \mathfrak{X}$. That is, for every $g \in G$, $x \in X$, and $f \in C^1(X)$, we have

$$\Gamma_{X^*}^g(v|_x)(f) = v|_{\Gamma_X^g(x)}(f),$$

or, equivalently,

$$v|_x(f \circ \Gamma_X^g) = v|_{\Gamma_X^g(x)}(f).$$

Note that the well-definition of $\Gamma_{X^*}^g$ as a map on vector fields relies on the fact that Γ_X^g is a diffeomorphism (which is in turn a consequence of the fact that X_T^g is a differentiable group action).

Together, these two conditions lead to the following important commutation (equivariance) property between the group action on X and the dynamical evolution map, namely,

$$(\Gamma_X^g \circ \Phi^t)(x) = (\Phi^t \circ \Gamma_X^g)(x), \quad (15)$$

which holds for all $g \in G$, $t \in \mathbb{R}$, and $x \in X$. Equation (15) expresses the fact that if $t \mapsto \Phi^t(x)$ is a dynamical trajectory starting at $x \in X$, then $t \mapsto \Gamma_X^g(\Phi^t(x))$ is also a dynamical trajectory, starting at $\Gamma_X^g(x)$. In addition, upon evaluation at the point $y' = \Gamma_Y^g(y)$, where $y \in Y$ is arbitrary, (15) leads to the invariance property

$$\Phi^t(\Gamma_X^g(x))(\Gamma_Y^g(y)) = \Phi^t(x)(y), \quad (16)$$

which holds for all $g \in G$, $t \in \mathbb{R}$, and $x \in X$.

Next, consider the left G -action on the product space Ω induced by the corresponding actions on X and Y ; this is characterized by the action map $\Gamma_\Omega^g : \Omega \mapsto \Omega$, $g \in G$, given by

$$\Gamma_\Omega^g((x, y)) = (\Gamma_X^g(x), \Gamma_Y^g(y)) = (x \circ \Gamma_Y^{g^{-1}}, \Gamma_Y^g(y)).$$

It follows from (15) that this group action and the dynamical evolution map $\tilde{\Phi}^t : \Omega \mapsto \Omega$ (see Section II) satisfy

$$(\Gamma_\Omega^g \circ \tilde{\Phi}^t)(\omega) = (\tilde{\Phi}^t \circ \Gamma_\Omega^g)(\omega), \quad (17)$$

for all $g \in G$, $t \in \mathbb{R}$, and $\omega \in \Omega$.

The invariance property in (16) endows the pseudometrics d_Q and \tilde{d}_Q in delay-coordinate space with symmetries, which can be summarized as follows.

Proposition 1. *The pseudometric d_Q in delay-coordinate space satisfies*

$$d_Q(\omega, \omega') = d_Q(\Gamma_\Omega^g(\omega), \Gamma_\Omega^{g'}(\omega')),$$

for all $\omega, \omega' \in \Omega$ and $g, g' \in G$. In particular, it vanishes along every G -orbit of $\omega \in \Omega$:

$$d_Q(\omega, \Gamma_\Omega^g(\omega)) = 0, \quad \forall g \in G.$$

Proof. It follows from (16) that for any $\omega = (x, y) \in \Omega$, $\omega' = (x', y') \in \Omega$, and $g \in G$,

$$\begin{aligned} d_Q^2(\Gamma_\Omega^g(\omega), \omega') &= \frac{1}{Q} \sum_{q=0}^{Q-1} |\Phi^t(\Gamma_X^g(x))(\Gamma_Y^g(y)) - \Phi^t(x')(y')|^2 \\ &= \frac{1}{Q} \sum_{q=0}^{Q-1} |\Phi^t(x)(y) - \Phi^t(x')(y')|^2. \end{aligned}$$

The last equation and the fact that $d_Q(\omega, \omega') = d_Q(\omega', \omega)$ lead to the first claim of the Proposition. The second claim follows immediately from the first. \square

Proposition 2. *For the class of scaling functions s_Q in Appendix B, the scaled pseudometric \tilde{d}_Q from (5) exhibits the analogous invariance properties, namely*

$$\tilde{d}_Q(\omega, \omega') = \tilde{d}_Q(\Gamma_\Omega^g(\omega), \Gamma_\Omega^{g'}(\omega')),$$

for all $\omega, \omega' \in \Omega$ and $g, g' \in G$.

A proof of Proposition 2 can be found in Appendix B 4.

To examine the implications of these results, consider the quotient set $\Omega/G = \{\Gamma_\Omega(\omega) \mid \omega \in \Omega\}$, consisting of all G -orbits $\Gamma_\Omega(\omega) = \{\Gamma_\Omega^g \mid g \in G\}$ on Ω . It is a direct consequence of Proposition 1 that the set of $[\cdot]_Q$ equivalence classes on Ω with respect to the pseudometric d_Q contains Ω/G as a subset. Since the delay-coordinate observation map F_Q maps all elements of an arbitrary equivalence class $[\omega]_Q$ (and in particular, the elements of an equivalence class in Ω/G) to a single point $F_Q(\omega) \in \mathbb{R}^Q$, we interpret this behavior as our method “factoring out” the dynamical symmetry G from the data.

In particular, given any observable $f \in H_M$, it follows from the definition of the kernel integral operator K_Q that for any $\omega \in \Omega$ and $g \in G$,

$$K_Q f(\omega) = K_Q f(\Gamma_\Omega^g(\omega)).$$

Moreover, it is straightforward to check that the left- and right-normalization functions $r_Q = K_Q \mathbf{1}_M$ and $l_Q = K_Q(\mathbf{1}_M/r_Q)$ are both invariant under Γ_Ω^g , and as a result,

$$P_Q f(\omega) = P_Q f(\Gamma_\Omega^g(\omega)).$$

The last equation implies that the eigenfunctions ϕ_j of P_Q used for spatiotemporal pattern extraction are constant on G -orbits. Note that the properties described above do not rely on any prior knowledge of the symmetry represented by G , and would also apply for other types of group symmetries that the system exhibits besides the dynamical symmetries discussed here.

2. Spectral characterization

We can gain additional intuition on the relationships between the spectra of P_Q and symmetry operators associated with G acting on H_M , under the additional assumptions that Γ_X^g and Γ_Y^g preserve the measures μ and ν , respectively; that is $\mu(\Gamma_X^{g^{-1}}(S)) = \mu(S)$ for every $g \in G$ and measurable set $S \subseteq X$, and an analogous relationship holds for ν . An immediate consequence of these assumptions is that the product measure $\rho = \mu \times \nu$ is preserved by Γ_Ω^g . While these properties cannot be expected to hold in general (for example, the support A of μ may not be invariant under Γ_X^g), if they are indeed true, then associated with G are groups of unitary operators $R_A^g : H_A \mapsto H_A$, $R_Y^g : H_Y \mapsto H_Y$, and $R_M^g : H_M \mapsto H_M$, defined for every $g \in G$ via composition with the corresponding group actions, i.e.,

$$\begin{aligned} R_A^g f_A &= f_A \circ \Gamma_A^g, & R_Y^g f_Y &= f_Y \circ \Gamma_Y^g, \\ R_M^g f_M &= (R_A^g \otimes R_Y^g) f_M = f_M \circ \Gamma_\Omega^g, \end{aligned}$$

with $f_A \in H_A$, $f_Y \in H_Y$, and $f_M \in H_M$. Notice the analogy in the construction of the unitary symmetry groups $\{R_A^g\}_{g \in G}$ and $\{R_M^g\}_{g \in G}$ with the construction of the unitary Koopman groups $\{U^t\}_{t \in \mathbb{R}}$ and $\{\tilde{U}^t\}_{t \in \mathbb{R}}$, respectively. In particular, the maps $g \mapsto R_X^g$ and $t \mapsto U^t$ are representations of G and the Abelian group of the real numbers, respectively, on H_A , and similarly $g \mapsto R_\Omega^g$ and $t \mapsto \tilde{U}^t$ are representations of these groups on H_M . Note also that in many applications the symmetries associated with G will not be known a priori, so we will not assume that we have access to the R^g operators from data.

It is a direct consequence of (15) that the symmetry and Koopman operators on H_A commute; that is,

$$[R_A^g, U^t] = R_A^g U^t - U^t R_A^g = 0,$$

for all $g \in G$ and $t \in \mathbb{R}$. As is well known, commuting operators have common eigenspaces. In particular, given an eigenfunction z of U^t at eigenvalue Λ^t , we have

$$U^t R_A^g z = R_A^g U^t z = \Lambda^t R_A^g z; \quad (18)$$

this shows that $R_A^g z$ lies in the same Koopman eigenspace as z . As stated in Section II, and further discussed in Appendix A, in the presence of ergodicity (assumed here), all eigenspaces of U^t are one-dimensional. This, together with (18) implies that there exists a complex number γ_A^g such that

$$R_A^g z = \gamma_A^g z; \quad (19)$$

in other words, z is an eigenfunction of R_A^g at eigenvalue γ_A^g . Note that the γ_A^g are not necessarily simple eigenvalues.

Analogous commutation relationships hold for the Koopman and symmetry operators on H_M due to (17); that is,

$$[\tilde{U}^t, R_\Omega^g] = 0,$$

which implies that if $\tilde{z} \in H_M$ is an eigenfunction of \tilde{U}^t at eigenvalue Λ^t (recall that U^t and \tilde{U}^t have the same eigenvalues), then $R_\Omega^g \tilde{z}$ is also an eigenfunction of \tilde{U}^t at the same eigenvalue. Now, as discussed in Appendix A, all eigenfunctions of \tilde{U}^t have the tensor-product form

$$\tilde{z} = z \otimes \psi, \quad (20)$$

where $z \in H_A$ is the unique (up to normalization) eigenfunction of U^t at eigenvalue Λ^t , and ψ an arbitrary spatial pattern in H_Y . The latter, in conjunction with (19), leads to

$$R_M^g \tilde{z} = (R_A^g \otimes R_Y^g)(z \otimes \psi) = (\gamma_A^g z) \otimes (R_Y^g \psi),$$

which in turn implies that \tilde{z} from (20) is an eigenfunction of R_M^g if and only if ψ is an eigenfunction of R_Y^g . The R_M^g eigenvalue corresponding to \tilde{z} is then given by

$$\gamma_M^g = \gamma_Y^g / \gamma_A^g = \gamma_Y^g \gamma_A^{g*} = \gamma_Y^g \gamma_A^{g^{-1}}, \quad (21)$$

where γ_Y^g is the R_Y^g eigenvalue corresponding to ψ , and the last two equalities are due to the unitarity of R_A^g . We have thus obtained a complete characterization of the common eigenspaces of \tilde{U}^t and R_M^g .

We now examine the existence of common eigenspaces between \tilde{U}^t , R_M^g , and the kernel integral operators K_Q , P_Q , noting that the latter are accessible from data whereas R_M^g may be unknown. First, in general, \tilde{U}^t does not commute with K_Q or P_Q . As will be discussed in more detail in Section V B, this can be attributed to the fact that the distance pseudometric \tilde{d}_Q is not invariant under the dynamical flow $\tilde{\Phi}^t$. There, we will also see that the pseudometric d_∞ is invariant under \tilde{U}^t , and an asymptotic commutation result between \tilde{U}^t and K_∞ , P_∞ will follow (Theorem 4). On the other hand, the invariance of \tilde{d}_Q with respect to Γ_Ω^g established in Proposition 2 holds for all Q , leading to the following result:

Theorem 3. *For any $g \in G$, the operator R_M^g commutes with both K_Q and P_Q . Moreover, every function in the range of K_Q and P_Q is invariant under R^g ; i.e.,*

$$R_M^g K_Q = K_Q, \quad R_M^g P_Q = P_Q.$$

Proof. We begin by verifying the claims of the theorem for K_Q . In particular, it follows from Proposition 2 and the fact that ρ is invariant under Γ_Ω^g that for any $g, g' \in G$,

$$\begin{aligned} K_Q f(\omega) &= \int_M k_Q(\omega, \omega') f(\omega') d\rho(\omega') \\ &= \int_M k_Q(\omega, \Gamma_\Omega^g(\omega')) f(\Gamma_\Omega^g(\omega')) d\rho(\omega') \\ &= \int_M k_Q(\Gamma_\Omega^{g^{-1}}(\omega), \omega') f(\Gamma_\Omega^g(\omega')) d\rho(\omega') \\ &= \int_M k_Q(\Gamma_\Omega^{g'}(\omega), \omega') f(\Gamma_\Omega^g(\omega')) d\rho(\omega') \\ &= R_M^{g'} K_Q R_M^g f(\omega). \end{aligned}$$

Setting $g' = g^{-1}$ in the above, and acting on both sides by R_M^g , leads to $R_M^g K_Q = K_Q R_M^{g'}$; i.e., $[R_M^g, K_Q] = 0$, as claimed. On the other hand, setting g to the identity element of G leads to $K_Q = R_M^{g'} K_Q$, which implies that $R_M^{g'}(\text{ran } K_Q) \subseteq \text{ran } K_Q$.

The corresponding results for P_Q follow by similar arguments, together with the fact that the diffusion maps normalization functions satisfy $r_Q \circ \Gamma_\Omega^g = r_Q$ and $l_Q \circ \Gamma_\Omega^g = l_Q$, as stated in Section IV B 1. \square

Theorem 3 is a direct manifestation of the fact that the operators K_Q and P_Q are based on kernels operating on the M/G equivalence classes. It provides a characterization of the eigenspaces of K_Q and P_Q (which are finite-dimensional by compactness of these operators) as finite-dimensional subspaces of H_M , associated with unit R_M^g eigenvalue. These subspaces are also representation spaces for the symmetry group G . In particular, let W_k , be the eigenspace of P_Q corresponding to eigenvalue λ_k , and $\{\phi_{1k}, \dots, \phi_{m_k k}\}$ with $m_k = \dim W_k$ an orthonormal basis of this space. Note that every such ϕ_{jk} corresponds to a VSA spatiotemporal pattern as described in Section III. By Theorem 3, $R_M^g \phi_{jk} = \phi_{jk}$ for every $g \in G$ and $j \in \{1, \dots, m_k\}$. As a result, restricted to W_k , R_Ω^g reduces to the identity operator.

Before closing this section, it is worthwhile noting that the eigenspaces W_k will not necessarily admit a tensor product decomposition $W_k = W_k^A \otimes W_k^Y$ in terms of G representation spaces W_k^A and W_k^Y that are subspaces of H_A and H_Y , respectively. In contrast, data analysis techniques operating in scalar-valued function spaces will oftentimes produce spatiotemporal patterns of the form $\varphi_k \otimes \psi_k \in H_M$, with φ_k and ψ_k lying in such representation spaces W_k^A and W_k^Y , respectively. For instance, in POD, φ_k and ψ_k would be eigenfunctions of the temporal and spatial covariance operators, $C_A : H_A \mapsto H_A$ and $C_Y : H_Y \mapsto H_Y$, respectively, corresponding to the same eigenvalue. A calculation analogous to that in the proof of Theorem 3 readily shows that C_A commutes with R_A^g , and C_Y commutes with R_Y^g . As a result, $\varphi_k \otimes \psi_k$ lies in the finite-dimensional, tensor product representation space $\tilde{W}_k = W_k^A \otimes W_k^Y$. As with W_k , \tilde{W}_k is also an eigenspace of R_Ω^g , but it will not necessarily correspond to eigenvalue 1.

C. Bundles in the limit of infinitely many delays

An analogous topological bundle construction to that described in Section IV A can also be performed in the limit of infinitely many delays, under the assumption that the pseudometric d_∞ is continuous on $M \times M$. Throughout this section, we will assume that the latter is indeed the case. As discussed in Appendix C, this condition will be satisfied under suitable decay conditions on the expansion coefficients of the observation map F in the Koopman eigenfunction basis.

To construct the base space, B_∞ , we start from the sequence space $\mathbb{R}^\mathbb{N}$, which we equip with the pseudometric $d'_\infty : \mathbb{R}^\mathbb{N} \times \mathbb{R}^\mathbb{N} \mapsto \mathbb{R}$ defined as

$$d'_\infty(b, b') = \lim_{q \rightarrow \infty} \left(\frac{1}{q} \sum_{n=0}^{q-1} |b_n - b'_n|^2 \right)^{1/2},$$

whenever the limit exists, and $d'_\infty(b, b') = 0$ otherwise, where $b = (b_0, b_1, \dots)$ and $b' = (b'_0, b'_1, \dots)$. We then define $\hat{\ell}^2$ as the set of equivalence classes in $\mathbb{R}^\mathbb{N}$ associated with d'_∞ , and equip this space with the corresponding induced metric (which we again denote by d'_∞) and its associated topology. It then follows that the map $F_\infty : \Omega \mapsto \hat{\ell}^2$ given by

$$F_\infty((x, y)) = [(F((x, y)), F((\Phi^{-\tau}(x), y)), F((\Phi^{-2\tau}(x), y)), \dots)]_{d'_\infty}$$

is continuous on M . Defining also $B_\infty = F_\infty(M)$ and the continuous map $\pi_\infty : M \mapsto B_\infty$ by restriction of the image space of F_∞ , it follows that $(M, B_\infty, \pi_\infty)$ is a topological bundle. As in the finite- Q case, we can view the d_∞ pseudometric as a pullback of the d'_∞ metric on $\hat{\ell}^2$, i.e.,

$$d_\infty(\omega, \omega') = d'_\infty(F_\infty(\omega), F_\infty(\omega'))$$

for any $\omega, \omega' \in M$.

It should be noted that in the limit of infinitely many delays the base space B_∞ may be smaller than B_Q at finite Q , in the sense that $d_\infty(\omega, \omega')$ may vanish even if $d_Q(\omega, \omega')$ is non-vanishing. For instance, as stated in Section III B, $d_\infty(\omega, \omega')$ vanishes for points $\omega = (x, y)$ and $\omega' = (x', y)$ such that x and x' lie in the same stable manifold. In general, it follows from the results in [51] that for fixed $y \in Y$ the image $F_\infty(A \times \{y\})$ is homeomorphic to a torus \mathbb{T}^{m_y} , whose dimension m_y may depend on y , but is bounded by the number of basic frequencies in the point spectrum of the Koopman operator U^t . Moreover, on that torus the dynamics is a quasiperiodic ergodic rotation with m_y basic frequencies. While studying the structure of $F_\infty(M) = \cup_{y \in Y} F_\infty(A \times \{y\})$ in more detail is beyond the scope of this work, it is worthwhile mentioning that the dynamics on that set may possess more symmetries than the full dynamical system on M . With increasing Q , VSA is expected to become increasingly efficient in capturing such symmetries.

V. ASYMPTOTIC BEHAVIOR

In this section, we study the asymptotic behavior of VSA in the limit of no delays and infinitely many delays, where we establish connections with matched filters and the point spectra of Koopman operators, respectively.

A. Limit of no delays

The limit of no delays, $Q = 1$, can be studied by means of the topological bundles introduced in Section IV A. In this case, the base space B_1 is equal to the image of M under F . Moreover, because F is continuous and M is connected (see Section II), B_1 is a closed interval $[b_1, b_2]$, where b_1 and b_2 are the minimum and maximum values of F over M , respectively. Moreover, the projection map π_1 is given by restriction of F onto M . Thus, the Markov kernel p_1 is given by pulling back a Markov kernel p'_1 on that interval; the Markov property of p'_1 holding with respect to the pushforward measure ρ'_1 . It then follows from (14), in conjunction with the fact that π_1 is given by restriction of F , that the eigenfunctions ϕ_j from Section III C are constant on the level sets of F .

On the basis of this observation, we can expect $\{\phi_j\}$ to be an efficient basis for representing the observation map restricted on M through an approximation of the form

$$F \approx F_l = \sum_{j=0}^l c_j \phi_j, \quad c_j = \langle \phi_j, F \rangle_{H_M}, \quad (22)$$

In particular, because of (14), the expansion coefficients in this approximation are given by

$$c_j = \langle \varphi_j \circ \pi_1, F \rangle_{H_M} = \langle \varphi_j, \iota_{B_1} \rangle_{L^2(B_1, \rho'_1)},$$

where $\iota_{B_1} : B_1 \mapsto \mathbb{R}$ is the canonical inclusion map, mapping $B_1 \subset \mathbb{R}$ to \mathbb{R} . Thus, if it happens that ι_{B_1} lies in $\text{ran } \overline{P'_1}$ (the closure of the range of P'_1), the expansion in (22) asymptotically recovers the full input signal, i.e., $F = \lim_{l \rightarrow \infty} F_l := F_\infty$. As in the case of F , we use an overarrow notation, \vec{F}_l , to represent the unique continuous function in $C(A; C(Y))$ such that $\vec{F}_l(x)(y) = F_l((x, y))$ for all $x \in A$ and $y \in Y$.

While, in general, it is not guaranteed that $\iota_{B_1} \in \text{ran } \overline{P'_1}$, the expansion in (22) is generally expected to provide a good approximation to F , even at small l . In particular, if the first nonconstant eigenfunction φ_1 is monotonic on B_1 , then even the two-term approximation with $l = 1$ will recover the qualitative features of the input signal. For instance, if the measure ρ'_1 has a smooth density γ relative to the Lebesgue measure on B_1 , then it follows from well known results in manifold learning [20, 23, 27, 29] that as the kernel bandwidth parameter ϵ approaches zero, the eigenfunctions of P'_1 approximate the eigenfunctions of a Laplace-Beltrami operator on B_1 for Neumann boundary conditions and a Riemannian metric that depends on γ . It is also straightforward to verify that φ_1 is indeed monotonic in this case, though the fact that the φ_j have vanishing derivative at b_1 and b_2 means that $\iota_{B_1} \notin \text{ran } \overline{P'_1}$, and (22) cannot recover the input signal exactly.

In general, given the spatiotemporal signal $t \mapsto \vec{F}(\Phi^t(x))$ from (2) associated with the observation map, we can interpret the signal $t \mapsto \vec{F}_l(\Phi^t(x))$ as a filtered spatiotemporal signal, which, if (22) indeed provides a good

approximation, is well matched with the source used to compute F_l . If, moreover, the number of basis functions l required for the approximation to be accurate is small, then the filtered signal can be reconstructed with high robustness. It is important to note that besides the strict limit $Q = 1$, these properties can also be expected to hold for the eigenfunctions of P_Q for moderate values of $Q > 1$. This is particularly useful for denoising noisy data since data-driven approximations of P_Q become increasingly robust to noise as Q grows [50].

B. Limit of infinitely many delays

As shown in [50, 51], scalar-valued kernels operating on delay-coordinate mapped data acquire an important property as the number of delays Q tends to infinity; namely, that the associated kernel integral operators (the analogs of K_∞ and P_∞ from Section III operating on the scalar-function space H_A) commute with the Koopman operators U^t of the system, and as a result have common eigenspaces with these operators. Since scalar-valued Koopman eigenfunctions in ergodic systems evolve periodically and with a single Fourier frequency along trajectories of the dynamics [43], this property endows kernel eigenfunctions with high temporal coherence and ability to recover distinct intrinsic dynamical timescales from multiscale input data. Using similar arguments, here we show that an analogous commutation result holds between K_∞ , P_∞ , and the lifted Koopman operators \tilde{U}^t acting on H_M :

Theorem 4. *For any $t \in \mathbb{R}$, the Koopman operators \tilde{U}^t commute with both K_∞ and P_∞ . In particular, since K_∞ and P_∞ are compact, if they have eigenspaces corresponding to nonzero eigenvalues, then these eigenspaces are finite-dimensional and are also eigenspaces of \tilde{U}^t . Moreover, \tilde{U}^t is unitarily diagonalizable on these eigenspaces.*

A proof of this theorem is included in Appendix C.

Before examining the implications of Theorem 4, it is worthwhile pointing out its similarities and differences with Theorem 3 on the commutation between K_Q , P_Q , and the symmetry operators R_M^q . Clearly, there are strong analogies between these two results, as they both establish commutation properties between compact kernel integral operators for data analysis and groups of (not necessarily compact) unitary operators. Therefore, these results allow for identification of finite-dimensional invariant subspaces of important operators associated with the dynamics and spatial symmetries from operators accessible from data (see Section VI).

At the same time, Theorems 3 and 4 have important differences, namely that (1) the commutation between the symmetry and kernel integral operators holds at any Q , whereas the corresponding result for the Koopman operators holds in the asymptotic limit $Q \rightarrow \infty$, and (2) in systems of sufficiently high complexity (mixing), \tilde{U}^t may have no nonconstant eigenfunctions with respect to the

degrees of freedom on A . In particular, it is a standard result from the spectral theory of dynamical systems [42] that a necessary and sufficient condition for (2) to occur is that the system is weak-mixing (a weak form of dynamical chaos which is implied if the system is mixing). In such cases, any finite-dimensional eigenspaces of P_∞ would only recover static spatial patterns with no dependence on the A coordinates.

That being said, many systems encountered in real-world applications are of the mixed-spectrum type, and in such systems, our technique applied with a large number of delays should have high skill in recovering coherent spatiotemporal patterns associated with the point spectrum of the Koopman group. Moreover, even if the system is weak-mixing, in practice we always operate at finitely many delays, and it is reasonable to expect that physically meaningful spatiotemporal patterns could still be detected. For example, the experiments in Section VII are conducted on the KS system in a chaotic regime, yet P_Q is found to recover physically meaningful spatiotemporal patterns. In general, however, our theoretical understanding of the behavior of the techniques presented here and in [50, 51] at large numbers of delays remains limited when the spectrum of the Koopman operator of the system is purely continuous.

For the remainder of this section, we will assume that the system under study has at least some nontrivial Koopman eigenfunctions. In particular, let $\tilde{z} \in H_M$ be an eigenfunction of \tilde{U}^t at eigenvalue Λ^t . Then,

$$\tilde{U}^t P_\infty \tilde{z} = P_\infty \tilde{U}^t \tilde{z} = \Lambda^t P_\infty \tilde{z}, \quad (23)$$

which implies that $P_\infty \tilde{z}$ is also an eigenfunction of \tilde{U}^t at the same eigenvalue. As discussed in Appendix A, the eigenvalues of \tilde{U}^t are identical to those of the Koopman operator U^t on H_A . However, unlike those of U^t , the eigenvalues of \tilde{U}^t are not simple, and we cannot conclude that $P_\infty \tilde{z} = \lambda \tilde{z}$ for some number λ ; i.e., it is not necessarily the case that \tilde{z} is also an eigenfunction of P_∞ (despite the fact that (23) implies that every eigenspace of \tilde{U}^t is invariant under P_∞). In fact, the eigenspaces of \tilde{U}^t are infinite-dimensional, and there is no a priori distinguished set of spatiotemporal patterns in each eigenspace.

To identify a distinguished set of spatiotemporal patterns associated with Koopman eigenfunctions, we take advantage of the fact that P_∞ is a compact operator. This means that the eigenspaces of P_∞ corresponding to nonzero eigenvalues are finite-dimensional, and for each such eigenspace we can find an orthonormal basis consisting of simultaneous eigenfunctions of P_∞ and \tilde{U}^t .

To verify this, let $W_k \subset H_M$ be the eigenspace of P_∞ corresponding to eigenvalue λ_k , and let f be an arbitrary element of W_k . Since

$$P_\infty \tilde{U}^t f = \tilde{U}^t P_\infty f = \lambda_k \tilde{U}^t f,$$

we can conclude that $\tilde{U}^t f \in W_k$; i.e., that W_k is a finite-dimensional invariant subspace of H_M under \tilde{U}^t . Choosing an orthonormal basis $\{\phi_{1k}, \dots, \phi_{m_k k}\}$ for this space,

where $m_k = \dim W_k$, we can expand $f = \sum_{j=1}^{m_k} c_j \phi_{jk}$ with $c_j = \langle \phi_{jk}, f \rangle_{H_M}$, and compute

$$\tilde{U}^t f = \sum_{i,j=1}^{m_k} \phi_{ik} \tilde{U}_{ij} c_j, \quad \tilde{U}_{ij} = \langle \phi_{ik}, \tilde{U}^t \phi_{jk} \rangle_{H_M}. \quad (24)$$

By unitarity of \tilde{U}^t , the $m_k \times m_k$ matrix \mathbf{U} with elements \tilde{U}_{ij} is unitary, and therefore unitarily diagonalizable. Let then $\{v_j\}_{j=1}^{m_k}$ with $v_j = (v_{1j}, \dots, v_{m_k j})^\top$ be an orthonormal basis of \mathbb{C}^{m_k} consisting of eigenvectors of \mathbf{U} , and let $\Lambda_1^t, \dots, \Lambda_{m_k}^t$ be the corresponding eigenvalues. It is a direct consequence of (24) that the set $\{\tilde{z}_1, \dots, \tilde{z}_{m_k}\}$ with $\tilde{z}_j = \sum_{i=1}^{m_k} v_{ij} \phi_{ik}$ is an orthonormal basis of W_k consisting of Koopman eigenfunctions corresponding to the eigenvalues Λ_j^t . Since every element of W_k is an eigenfunction of P_∞ , we conclude that the \tilde{z}_j are simultaneous eigenfunctions of \tilde{U}^t and P_∞ .

Associated with each \tilde{z}_j is a unique continuous vector-valued function $\tilde{z}_j \in C(A; C(Y))$ such that $\tilde{z}_j(x)(y) = \tilde{z}_j((x, y))$ for all $x \in A$ and $y \in Y$. It follows from the Koopman eigenvalue equation (1) that \tilde{z}_j evolves according to

$$\tilde{z}_j(\Phi^t(x)) = e^{i\alpha_j t} \tilde{z}_j(x), \quad x \in A,$$

where $\alpha_j \in \mathbb{R}$ is the eigenfrequency corresponding to the eigenvalue $\Lambda_j^t = e^{i\alpha_j t}$. As a result, the spatiotemporal pattern from (2) associated with \tilde{z}_j is periodic with period $2\pi/\omega_j$.

It should be noted that in DMD and related algorithms applied to ergodic dynamical systems one assigns to every eigenvalue of the Koopman operator U^t acting on scalar functions a single pair (z_j, ψ_j) consisting of a Koopman eigenfunction $z_j \in H_A$ and the corresponding spatial pattern (Koopman mode) $\psi_j \in H_Y$. The associated spatiotemporal pattern in H_M is given by $z_j \otimes \psi_j$. Here, in the setting of vector-valued observables, from every (finite-dimensional) eigenspace of P_∞ we recover as many eigenfunctions of the lifted Koopman operator \tilde{U}^t as the dimension of that space.

C. Infinitely many delays with symmetries

As a final asymptotic limit of interest, we consider the limit $Q \rightarrow \infty$ under the assumption that the measure-preserving symmetry group actions from Section IV B 2 are available. In that case, the commutation relations

$$[R_M^g, \tilde{U}^t] = [R_M^g, P_\infty] = [P_\infty, \tilde{U}^t] = 0,$$

which imply that there exist finite-dimensional subspaces of H_M spanned by simultaneous eigenfunctions of R_M^g , \tilde{U}^t , and P_∞ . We also know from (20) that these eigenfunctions have the form

$$\tilde{z} = z \otimes \psi, \quad (25)$$

where $z \in H_A$ is a simultaneous eigenfunction of the Koopman operator U^t and the symmetry operator R_A^g ,

and ψ an eigenfunction of the symmetry operator R_Y^g on H_Y . Further, it follows from Theorem 3 that \tilde{z} is an eigenfunction of R_M^g at eigenvalue 1. The latter, in conjunction with (21), imposes the constraint

$$\gamma_A^{g*} \gamma_Y^g = 1$$

on the admissible R_A^g eigenvalues of z and R_Y^g eigenvalues of ψ in (25).

To conclude, in this limit, every eigenfunction in (25) is characterized by three eigenvalues; namely (1) a Markov eigenvalue λ associated with P_∞ , (2) a Koopman eigenfrequency α associated with U^t , (3) a spatial symmetry eigenvalue γ_Y^g .

VI. DATA-DRIVEN APPROXIMATION

A. Assumptions

We begin by stating certain assumptions that will be sufficient for our data-driven scheme to converge. First, we assume that the invariant measure μ is physical [52]; this means that there exists a set $\mathcal{B}_\mu \subseteq X$, of positive Lebesgue measure, called the basin of μ , defined as the set of all points $x \in X$ such that, for every continuous function $f : X \mapsto \mathbb{C}$, the property

$$\lim_{N_X \rightarrow \infty} \frac{1}{N_X} \sum_{n=0}^{N_X-1} f(\Phi^{n\tau}(x)) = \int_A f d\mu. \quad (26)$$

holds. Since μ is an ergodic measure, the invariant set A lies in the topological closure $\bar{\mathcal{B}}_\mu$ of \mathcal{B}_μ .

The significance of (26) is that it will allow our algorithms to converge as the number of samples N_X tends to infinity from experimentally accessible data; that is, measurements $F(\Phi^{n\tau}(x))$ of the system starting from any initial condition x in a set, \mathcal{B}_μ , of positive measure (volume). In particular, we do not require that the dynamical states underlying our measurements lie exactly on A , as the latter will be a measure-zero set in many applications (e.g., in dissipative systems). Of course, if it happens that $X = A$, then we also have $\bar{\mathcal{B}}_\mu = X$, and our methods converge by ergodicity without invoking an additional physical-measure assumption. Here, we will additionally assume that $\bar{\mathcal{B}}_\mu$ contains a compact set \mathcal{U} , again of positive Lebesgue measure, such that $A \subseteq \mathcal{U}$. The significance of this assumption is that the space of continuous functions on \mathcal{U} will provide a natural, universal (i.e., N_X -independent) space to establish spectral convergence of data-driven analogs of the kernel integral operators introduced in Section III.

The assumptions stated thus are sufficient to ensure that scalar-valued kernel algorithms with observation maps taking values in finite-dimensional spaces converge as the number of samples tends to infinity [51]. However, in the present setting of interest involving vector-valued observables that take values in infinite-dimensional

spaces, we also need to ensure that our algorithms converge in a practical scenario where measurements of the system are taken at a finite number of spatial points in Y . In particular, we assume that there exists a sequence $\mathcal{Q}_0, \mathcal{Q}_1, \dots$, whose every element is a collection $\mathcal{Q}_{N_Y} = \{(y_{0,N_Y}, w_{0,N_Y}), (y_{1,N_Y}, w_{1,N_Y}), \dots, (y_{N_Y,N_Y}, w_{N_Y,N_Y})\}$ with $y_{r,N_Y} \in Y$ and $w_{r,N_Y} > 0$, such that for every continuous function $f \in C(Y)$, we have

$$\lim_{N_Y \rightarrow \infty} \frac{1}{N_Y} \sum_{r=0}^{N_Y-1} f(y_{r,N_Y}) w_{r,N_Y} = \int_Y f d\nu. \quad (27)$$

The families \mathcal{Q}_{N_Y} could be made available, say, from a quadrature rule, or from an orbit of an ergodic dynamical system on Y (in which case the y_{r,N_Y} would be points along a typical orbit y_0, y_1, \dots of the dynamics and the weights w_{r,N_Y} would all be equal to 1). Together, (26) and (27) imply that for every function $f : \Omega \mapsto \mathbb{C}$, continuous on the compact set $\mathcal{V} = \mathcal{U} \times Y \subseteq \Omega$, and Lebesgue almost every initial condition $x_0 \in \mathcal{U}$, we have

$$\lim_{N_X \rightarrow \infty} \lim_{N_Y \rightarrow \infty} \frac{1}{N_X N_Y} \sum_{n=0}^{N_X-1} \sum_{r=0}^{N_Y-1} f((x_n, y_{r,N_Y})) w_{r,N_Y} = \int_M f d\rho, \quad (28)$$

where $x_n = \Phi^{n\tau}(x_0)$, and the order of the limits is not important.

In what follows, we will consider that a set of measurements $F_{nr} = F((x_n, y_r))$ with $n \in \{0, \dots, N_X - 1\}$ and $r \in \{0, \dots, N_Y - 1\}$ corresponding to an arbitrary state $x_0 \in \mathcal{U}$ and a set of points y_0, y_1, \dots in Y for which (27) holds with $w_{r,N_Y} = 1$ is available to us. The latter assumption is made for notational simplicity—our results hold for non-constant w_{r,N_Y} with obvious modifications. Also, we will arrange the points (x_n, y_r) in an ordered sequence $\omega_0, \omega_1, \dots, \omega_N$ where $N = N_X N_Y$ and $\omega_j = (x_{n_j}, y_{r_j})$ for some ordering (n_j, r_j) of the indices n and r . We will use the single-index notation ω_j whenever convenient, but will revert to (x_n, y_r) whenever we wish to emphasize the dependence of functions on Ω on their X and Y coordinates. In single-index notation, limits $N \rightarrow \infty$, will always be understood as limits induced by the double limit $N_X, N_Y \rightarrow \infty$ so that (28) holds. We will also denote the finite set of sampled points in Ω by $\Omega_N = \{\omega_0, \dots, \omega_{N-1}\}$. We emphasize again that we do not assume that we have knowledge of the underlying states x_n and/or the dynamical system (X, Φ^t) .

B. Data-driven Hilbert spaces and kernel integral operators

Following [51], in data-driven applications we approximate the Hilbert space of observables $H_M = L^2(M, \rho)$ by the finite-dimensional Hilbert space $H_{\Omega, N} = L^2(\Omega, \rho_N)$

associated with the sampling measure

$$\rho_N = \frac{1}{N} \sum_{j=0}^{N-1} \delta_{\omega_j},$$

where δ_{ω_j} are Dirac δ -measures centered at $\omega_j \in \Omega$. This Hilbert space is equipped with the inner product given by the finite sum

$$\langle f, g \rangle_{H_{\Omega, N}} = \int_{\Omega} f^* g d\rho_N = \frac{1}{N} \sum_{j=0}^{N-1} f^*(\omega_j) g(\omega_j).$$

Note that the elements of $H_{\Omega, N}$ are equivalence classes of complex-valued functions on Ω with equal values at the sampled points $\omega_0, \dots, \omega_N$; that is, the values of $f \in H_{M, N}$ away from those points are arbitrary, much like the values of elements H_M are arbitrary on sets of ρ -measure zero. This, in conjunction with the fact that the ω_j will in general not lie in M (which is the support of the invariant measure ρ), means that the spaces H_M and $H_{\Omega, N}$ are generally incompatible. In particular, the fact that ω_j may not lie in M raises the question of how to establish convergence of results obtained in $H_{\Omega, N}$ to the results of Section III involving functions in H_M . This is precisely where the assumptions on physical measures stated in Section VIA become important. In particular, it follows from (28) that if f and g are continuous functions on the compact set $\mathcal{V} \subseteq \Omega$ (and thus are representatives of equivalence classes in both $H_{\Omega, N}$ and H_M), then as $N \rightarrow \infty$, the inner products $\langle f, g \rangle_{H_{\Omega, N}}$ converge to $\langle f, g \rangle_{H_M}$.

By construction, $H_{\Omega, N}$ is isomorphic as a vector space to \mathbb{C}^N . As a result we can represent every $f \in H_{\Omega, N}$ by a column vector $\underline{f} = (f_0, \dots, f_{N-1})^T \in \mathbb{C}^N$, where f_j is equal to the value of any representative in the equivalence class $f \in H_{\Omega, N}$ evaluated at ω_j . Similarly, we can represent any linear operator $T : H_{\Omega, N} \mapsto H_{\Omega, N}$ by a matrix $\mathbf{T} \in \mathbb{C}^{N \times N}$ defined uniquely by the requirement that $g = Tf$ is equivalent to $\underline{g} = \mathbf{T}\underline{f}$ for all $f, g \in H_{\Omega, N}$. The inner product on H_N can also be represented by the canonical Euclidean inner product on \mathbb{C}^N through the relationship $\langle f, g \rangle_{H_{\Omega, N}} = \underline{f} \cdot \underline{g}/N$. By virtue of these relationships, we can carry out all linear operations on $H_{\Omega, N}$ by means of standard linear algebra. In particular, the eigenvalue problem for T is equivalent to an $N \times N$ matrix eigenvalue problem for \mathbf{T} .

As in Section (III A), we will consider kernel integral operators acting on $H_{\Omega, N}$. Since these spaces contain equivalence classes of functions defined on points outside of M , we will strengthen the continuity assumptions on kernels stated in Section III A (which were required to be continuous on M), and require that kernels be continuous on the compact set $\mathcal{V} \supseteq M$. With these assumptions, we will approximate the kernel k_Q in (8) with a data-driven kernel $k_{Q, N} : \mathcal{V} \times \mathcal{V} \mapsto \mathbb{R}_+$ such that

$$k_{Q, N}(\omega, \omega') = e^{-\tilde{d}_{Q, N}^2(\omega, \omega')/\epsilon},$$

where $\tilde{d}_{Q, N} : \mathcal{V} \times \mathcal{V} \mapsto \mathbb{R}$ is a pseudometric

$$\tilde{d}_{Q, N}(\omega, \omega') = d_Q(\omega, \omega') \sqrt{s_{Q, N}(\omega) s_{Q, N}(\omega')}$$

defined in terms of data-driven scaling functions $s_{Q, N} \in C(\mathcal{V})$. In particular, we will require that as $N \rightarrow \infty$, $s_{Q, N}$ converges uniformly to s_Q . Explicit definitions for $s_{Q, N}$ and a Lemma establishing uniform convergence of these functions to s_Q are included in Appendices B and D, respectively.

Next, we associate to $k_{Q, N}$ a linear operator $K_{Q, N} : H_{\Omega, N} \mapsto H_{\Omega, N}$, defined analogously to (3) by

$$\begin{aligned} K_{Q, N} f &= \int_{\Omega} k_{Q, N}(\cdot, \omega) f(\omega) d\rho_N(\omega) \\ &= \frac{1}{N} \sum_{j=0}^{N-1} k_{Q, N}(\cdot, \omega_j) f(\omega_j) \end{aligned} \quad (29)$$

with $f \in H_{\Omega, N}$. Note that $K_{Q, N}$ is trivially compact since $H_{\Omega, N}$ is finite-dimensional. We also construct a compact Markov operator $P_{Q, N} : H_{\Omega, N} \mapsto H_{\Omega, N}$ with kernel $p_{Q, N} : \Omega \times \Omega \mapsto \mathbb{R}_+$ by applying diffusion maps normalization to k as in Section III C, viz.

$$\begin{aligned} p_{Q, N}(\omega, \omega') &= \frac{k_{Q, N}(\omega, \omega')}{l_N(\omega) r_N(\omega')}, \\ r_N &= K_N 1_{\Omega}, \quad l_N = K_N(1_{\Omega}/r_N), \end{aligned} \quad (30)$$

where r_N and l_N are positive, continuous functions on \mathcal{V} , bounded away from zero. By construction, $p_{Q, N}$ has the Markov property

$$\int_{\Omega} p_{Q, N}(\omega, \cdot) d\rho_N = \frac{1}{N} \sum_{j=0}^{N-1} p_{Q, N}(\omega, \omega_j) = 1$$

for all $\omega \in \mathcal{V}$. It also follows from (28) that as $N \rightarrow \infty$, $p_{Q, N}(\omega, \omega')$ converges to the value of the Markov kernel $p(\omega, \omega')$ from (9) for all $\omega, \omega' \in \mathcal{V}$.

With these definitions, $P_{Q, N}$ is represented by the $N \times N$ Markov matrix $\mathbf{P}_N = [p_{Q, N}(\omega_i, \omega_j)/N]_{ij}$, and we can compute the eigenvalues, $\lambda_{j, N}$, and corresponding eigenvectors, $\phi_{j, N}$, of \mathbf{P}_N with $j \in \{0, 1, \dots, N-1\}$ by solving the matrix eigenvalue problem for \mathbf{P}_N . In particular, it follows by analogous arguments to those used in Section III C for the Markov operator P that the eigenvalues of \mathbf{P}_N are real and admit the ordering $1 = \lambda_{N, 0} > \lambda_{N, 1} \leq \lambda_{N, 2} \geq \dots$. Moreover, while the eigenvectors $\phi_{j, N}$ are defined only at the finite dataset $\Omega_N \subset \Omega$, those corresponding to nonzero eigenvalues $\lambda_{j, N}$ can be extended to continuous functions $\tilde{\phi}_{j, N}$ on \mathcal{V} given by

$$\tilde{\phi}_{j, N}(\omega) = \frac{1}{\lambda_{j, N}} \int_{\Omega} p_N(\omega, \omega') \phi_{j, N}(\omega') d\rho_N(\omega'), \quad \omega \in \mathcal{V}. \quad (31)$$

Note that these functions are consistent with $\phi_{j, N}$ on Ω_N ; i.e., $\tilde{\phi}_{j, N}(\omega_n) = \phi_{j, N}(\omega_n)$ for all $\omega_n \in \Omega_N$. Furthermore, $\tilde{\phi}_{j, N} \in C(\mathcal{V})$ corresponds to a vector-valued

function $\vec{\phi}_{j,N} \in C(\mathcal{U}; C(Y))$ such that (cf. (12))

$$\vec{\phi}_{j,N}(x)(y) = \tilde{\phi}_{j,N}(x, y), \quad x \in \mathcal{U}, \quad y \in Y,$$

and, for every initial condition $x \in \mathcal{U}$, $\vec{\phi}_{j,N}$ induces a continuous spatiotemporal pattern via (2).

In VSA, we compute data-driven spatiotemporal patterns of the class (31) associated with operators $P_{Q,N} : H_{\Omega,N} \mapsto H_{\Omega,N}$ constructed via the procedure described above applied to the class of kernels k_Q utilizing delay coordinate maps (see Section III B). In what follows, we will establish convergence of these data-driven patterns to those recovered from P_Q as described in Section III.

C. Spectral convergence

We begin by noting that every eigenfunction ϕ_j of P_Q with nonzero corresponding eigenvalue λ_j can be extended to a continuous function $\tilde{\phi}_j$ on \mathcal{V} via a formula analogous to (31), namely,

$$\tilde{\phi}_j(\omega) = \frac{1}{\lambda_j} \int_{\Omega} p_Q(\omega, \omega') \phi_j(\omega') d\rho(\omega'). \quad (32)$$

We also note that while P_Q is defined as an operator on H , we can also define an analogous operator $\tilde{P}_Q : C(\mathcal{V}) \mapsto C(\mathcal{V})$ via the same integral formula as in (3), but with the operand f and its image $P_Q f$ understood as continuous functions on \mathcal{V} . Similarly, we can define $\tilde{P}_{Q,N} : C(\mathcal{V}) \mapsto C(\mathcal{V})$ analogously to (29). Our goal in this section is to establish a type of spectral convergence of $P_{Q,N}$ to P_Q in the limit of large data, $N \rightarrow \infty$, that involves the continuous functions $\tilde{\phi}_{j,N}$ and $\tilde{\phi}_j$ and the operators $\tilde{P}_{Q,N}$ and \tilde{P}_Q . In essence, we will use $C(\mathcal{V})$ as a universal space to compare results derived from $P_{Q,N}$ and P_Q , even though these operators act on different spaces.

For that, we need the important notion of *compact convergence* of operators [23, 64].

Definition 5. A sequence of bounded operators $T_n : E \mapsto E$ on a Banach space E is said to converge compactly to a bounded operator $T : E \mapsto E$ if T_n converges to T pointwise (i.e., $T_n f \mapsto T f$ for all $f \in E$), and for every bounded sequence of vectors $f_n \in E$, the sequence $g_n = (T_n - T)f_n$ has compact closure (equivalently, g_n has a convergent subsequence).

While compact convergence is weaker than convergence in operator norm, it is nevertheless sufficient to imply convergence of isolated eigenvalues of bounded operators [64], and hence convergence of nonzero eigenvalues of compact operators and their corresponding eigenspaces. In the setting of interest here, these results imply that data-driven eigenvalues and eigenfunctions computed from $P_{Q,N}$ converge as $N \rightarrow \infty$ to their counterparts from P_Q , as follows.

Theorem 6. *The following hold:*

- (a) $\tilde{P}_{Q,N}$ and \tilde{P}_Q are both compact operators on $C(\mathcal{V})$ with respect to the uniform norm. As a result, their nonzero eigenvalues have finite multiplicities, and accumulate only at zero.
- (b) As $N \rightarrow \infty$, $\tilde{P}_{Q,N}$ converges compactly to \tilde{P}_Q .
- (c) Every nonzero eigenvalue λ_j of \tilde{P}_Q is an eigenvalue of P_Q . Moreover, if ϕ_j is an eigenfunction of P_Q corresponding to that eigenvalue, then $\tilde{\phi}_j$ from (32) is an eigenfunction of \tilde{P}_Q corresponding to the same eigenvalue. Analogous results hold for $\lambda_{j,N} > 0$, $\tilde{P}_{Q,N}$, $P_{Q,N}$, $\phi_{j,N}$, and $\tilde{\phi}_{j,N}$ from (31).

A proof of Theorem 6 can be found in Appendix D. This theorem leads to our following main convergence result:

Corollary 7. *For every nonzero eigenvalue λ_j of P_Q , there exists a positive integer N_0 such that the sequence of eigenvalues $\lambda_{j,N}$ of $P_{Q,N}$ with $N \geq N_0$ converges, as $N \rightarrow \infty$, to λ_j . Moreover, for every eigenfunction $\phi_j \in H_{\Omega}$ of P_Q corresponding to that eigenvalue, there exists a sequence of eigenfunctions $\phi_{j,N}$ of $P_{Q,N}$ with $N \geq N_0$, corresponding to $\lambda_{j,N}$, whose continuous extensions $\tilde{\phi}_{j,N}$ from (31) converge uniformly to $\tilde{\phi}_j$ from (32).*

Proof. Since λ_j is nonzero, it follows from Theorem 6(c) that it is an eigenvalue of \tilde{P}_Q and that $\tilde{\phi}_j$ from (32) is a corresponding eigenfunction. Moreover, since, by Theorem 6(b), $\tilde{P}_{Q,N}$ converges to \tilde{P}_Q compactly (and thus in spectrum for nonzero eigenvalues [64]), there exists $N_0 \in \mathbb{N}$ such that the j -th eigenvalues $\lambda_{j,N}$ of $\tilde{P}_{Q,N}$ are all nonzero for $N \geq N_0$, and thus, by Theorem 6(c), they are eigenvalues of $P_{Q,N}$. This establishes convergence of $\lambda_{j,N}$ to λ_j . The existence of eigenfunctions $\phi_{j,N}$ of $P_{Q,N}$ corresponding to λ_j , such that $\tilde{\phi}_{j,N}$ converges uniformly to $\tilde{\phi}_j$ is shown in [51]. \square

VII. APPLICATION TO THE KURAMOTO-SIVASHINSKY MODEL

A. Overview of the Kuramoto Sivashinsky model

The KS model, originally introduced as a model for wave propagation in a dissipative medium [59], or laminar flame propagation [60], is one of the most widely studied dissipative PDEs displaying spatiotemporal chaos. On a one-dimensional spatial domain $Y = [0, L]$, $L \geq 0$, the governing evolution equation for the real-valued scalar field $u(t, \cdot) : Y \mapsto \mathbb{R}$, $t \geq 0$, is given by

$$\dot{u} = -u \nabla u + \Delta u - \Delta^2 u, \quad (33)$$

where ∇ and $\Delta = -\nabla^2$ are the derivative and (positive definite) Laplace operators on Y , respectively. In what follows, we always work with periodic boundary conditions, $u(t, 0) = u(t, L)$, $\nabla u(t, 0) = \nabla u(t, L)$, \dots , for all $t \geq 0$.

The domain size parameter L controls the dynamical complexity of the system. At small values of this parameter, the trivial solution $u = 0$ is globally asymptotically stable, but as L increases, the system undergoes a sequence of bifurcations, marked by the appearance of steady spatially periodic modes (fixed points), then traveling waves (periodic orbits), and progressively more complicated solutions leading to chaotic behavior at $L \gtrsim 4 \times 2\pi$ [65–69].

A fundamental property of the KS system is that it possesses a global compact attractor, embedded within a finite-dimensional inertial manifold with an associated Lipschitz-continuous dynamical vector field [6, 7, 70–72]. That is, there exists a submanifold \mathcal{X} of the Hilbert space $H_Y = L^2(Y, \nu)$ with ν set to the Lebesgue measure, which is forward-invariant under the dynamics, and to which the solutions $u(t, \cdot)$ are exponentially attracted. This means that after the decay of initial transients, the effective degrees of freedom of the KS system, bounded above by the dimension of \mathcal{X} , is finite. Moreover, by Lipschitz continuity, the dynamics on \mathcal{X} have unique global (in time) solutions. Dimension estimates of inertial manifolds [72, 73] and attractors [74] of the KS system as a function of L indicate that the system exhibits extensive chaos, i.e., unbounded growth of the attractor dimension with L . As is well known, analogous results to those outlined above are not available for many other important models of complex spatiotemporal dynamics such as the Navier-Stokes equations.

For our purposes, the availability of strong theoretical results and rich spatiotemporal dynamics makes the KS model particularly well-suited to test the VSA framework presented in Sections III–VI. In our notation, an inertial manifold \mathcal{X} of the KS system will act as the state space manifold X , which is embedded in this case in H_Y . Moreover, the compact invariant set A will be a subset of the global attractor supporting an ergodic probability measure, μ . As stated above, on X the dynamical vector field, v , is Lipschitz continuous, which implies that the dynamics restricted on that set is described by a flow map $\Phi^t : X \mapsto X$, $t \in \mathbb{R}$, as in Section II. In particular, for every initial condition $x \in X$, the orbit $t \mapsto x(t) = \Phi^t(x_0)$ with $t \geq 0$ is the unique solution to (33) with initial conditions x . While in practice the initial data will likely not lie on X , the dynamics possesses an exponential tracking property that ensures that for any admissible initial condition $u \in H_Y$ there exists a trajectory $x(t)$ on X to which the evolution starting from u converges exponentially fast.

As stated in Section VIA, for data-driven approximation purposes, we will formally assume that the measure μ is physical. While, to our knowledge there are no results in the literature addressing the existence of physical measures (with appropriate modifications to account for the infinite state space dimension) specifically for the KS system, recent results [75, 76] on infinite-dimensional dynamical systems that include the class of dissipative systems in which the KS system belongs, indicate that

analogous of the assumptions made in Section VIA should hold.

Another important feature of the KS system is that it admits nontrivial symmetry group actions on the spatial domain Y , which have played a central role in bifurcation studies of this system [65–68]. In particular, it is a direct consequence of the structure of the governing equation (33) and the periodic boundary conditions that if $u(x, t)$ is a solution, then so are $u(x + \alpha, t)$ and $u(-x, t)$, where $\alpha \in \mathbb{R}$. As discussed in Section IV B, this implies that the vector field v on the inertial manifold is invariant under the actions induced by the orthogonal group $O(2)$ and the reflection group on the circle. In particular, under $O(2)$ symmetry, the theoretical spatial patterns recovered by POD and comparable eigendecomposition techniques would be Fourier modes [9], which are arguably non-representative of the complex spatiotemporal patterns generated by the KS system.

We emphasize again that the existence of symmetries does not necessarily imply that they are inherited by data-driven operators for extracting spatial and temporal patterns constructed from a single orbit of the dynamics, since, e.g., the ergodic measure sampled by that orbit may not be invariant under the symmetry group action. While studies have determined that this type of symmetry breaking indeed occurs at certain dynamical regimes of the KS system [53], the presence of symmetries still dominates the leading spatial patterns recovered by POD and comparable eigendecomposition techniques utilizing scalar-valued kernels.

B. Results and discussion

We apply VSA to data generated by the KS model at the chaotic regimes $L = 22$ and $L = 94$; these “standard” regimes have been investigated extensively in the literature (e.g., [68, 69]). We have integrated the model using the publicly available Matlab code accompanying Ref. [77]. This code is based on a Fourier pseudospectral discretization, and utilizes a fourth-order exponential time-differencing Runge-Kutta integrator appropriate for stiff problems. Throughout, we use 65 Fourier modes (which is equivalent to a uniform grid on Y with $N_Y = 65$ gridpoints), and a timestep of $\tau = 0.25$ natural time units.

Each of the experiments described below starts from initial conditions given by setting the first four Fourier coefficients to 0.6 and the remaining 61 to zero. Before collecting data for analysis, we let the system equilibrate near its attractor for a time interval of 2500 natural time units. We compute spatiotemporal patterns using the eigenfunctions of the data-driven Markov operator $P_{Q,N}$ as described in Section VI. In one of our $L = 22$ experiments, we also compare our results with spatiotemporal patterns computed via the NLSA algorithm [31] which utilizes a scalar-valued kernel operating on delay-coordinate mapped data. This method is applied to the

same KS data as the operator-valued kernel approach using the same number of delays.

We begin by presenting results obtained by applying VSA to a dataset of $N_X = 1000$ samples at the $L = 22$ regime, using a small number of delays, $Q = 15$. As shown in Fig. 1(a), in this regime the system supports two large-scale (wavenumber $\sim L/2$) quasi-stationary patterns that interact chaotically with one another. Viewed over longer time intervals (see Fig. 2) these patterns are seen to intermittently undergo left or right translations in the spatial domain, giving rise to an apparent traveling-wave signal. From a state-space perspective, this behavior can be described in terms of unstable equilibria (forming a continuous $O(2)$ family due to translation invariance) connected by homoclinic orbits [67, 68]. The wavenumber $L/2$ structures also act as envelopes for smaller-scale intermittent traveling wave disturbances, propagating either to the left or right.

As discussed in Section V A, at the small- Q limit VSA is expected to yield eigenfunctions $\phi_{k,N}$ which are approximately constant on the level sets of the input signal, and, with increasing k , capture increasingly smaller scales in the directions transverse to the level sets. As is evident in Fig. 1, the leading three nonconstant eigenfunctions, $\phi_{1,N}$, $\phi_{2,N}$, and $\phi_{3,N}$, indeed display this behavior, featuring wavenumbers 2, 3, and 4, respectively, in the directions transverse to the level sets. This behavior continues for eigenfunctions $\phi_{k,N}$ with higher k . To assess the efficacy of these patterns in reconstructing the input signal, we compute the fractional explained variances (cf. (22))

$$\frac{|\langle F, \phi_{j,N} \rangle_{H_{\Omega,N}}|^2}{\|\phi_{j,N}\|_{H_{\Omega,N}}^2 \|F\|_{H_{\Omega,N}}^2},$$

where $\|\phi_j\|_{H_{\Omega,N}} = 1$ since we use normalized eigenfunctions. For the $\phi_{1,N}$, $\phi_{2,N}$, and $\phi_{3,N}$ eigenfunctions in Fig. 1, these quantities are 0.91, 5.2×10^{-4} , and 0.016, respectively, which demonstrates that even the one-term ($l = 1$) reconstruction via (22) captures most of the signal variance. As noted in Section V A, the limit of small Q is not the intended parameter regime for our method to produce a decomposition of the signal into simpler components, but can nevertheless be useful in a number of scenarios such as level set estimation and matched filtering of noisy spatiotemporal data.

Next, we consider a longer dataset with $N_X = 10,000$ samples (2500 natural time units), again at $L = 22$, analyzed using $Q = 500$ delays. The raw data and representative eigenfunctions from this analysis, obtained via either VSA or NLSA, are displayed in Fig. 2. Figure 3 highlights a portion of the raw data and vector-valued eigenfunctions over an interval spanning 1000 time units. As discussed in Section V B, at large Q we expect these eigenfunctions to approximately lie in finite-dimensional subspaces of the Hilbert space H of vector-valued observables associated with the point spectrum of the Koopman operator, thus acquiring timescale separation. This is clearly the case in the eigenfunctions in

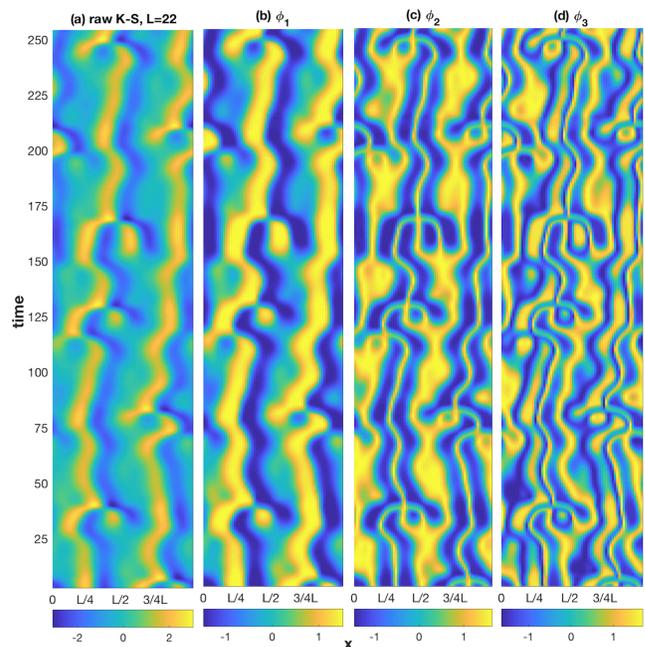


FIG. 1. Raw data (a) and representative VSA spatiotemporal patterns (b–d), obtained through the eigenfunctions $\phi_{k,N}$ of the Markov operator $P_{Q,N}$ associated with the Markov kernel in (30) for the KS model with $L = 22$ using $Q = 15$ delays. The spatiotemporal patterns in (b–d) were computed from eigenfunctions $\phi_{2,N}$, $\phi_{3,N}$, and $\phi_{4,N}$ of this operator by applying (2); i.e., they correspond to evaluations of the vector-valued observables $\vec{\phi}_{k,N}$ along the orbit of the dynamics sampled in the training phase. Notice that the patterns are approximately constant on the level sets of the raw data.

Figs. 2 and 3, where $\phi_{1,N}$ is seen to capture the evolution of the wavenumber $L/2$ structures, whereas $\phi_{10,N}$ and $\phi_{15,N}$ recover smaller-scale traveling waves embedded within the large-scale structures with a general direction of propagation either to the right ($\phi_{10,N}$) or left ($\phi_{15,N}$). The fractional explained variances associated with eigenfunctions $\phi_{1,N}$, $\phi_{10,N}$, and $\phi_{15,N}$ are 0.23, 0.047, and 0.017, respectively. As expected, these values are smaller than the 0.91 value due to eigenfunction $\phi_{1,N}$ for $Q = 15$, but are still fairly high despite the intermittent nature of the input signal. Ranked with respect to fractional explained variance, $\phi_{1,N}$, $\phi_{10,N}$, and $\phi_{15,N}$ are the first, fourth, and fifth among the $Q = 200$ VSA eigenfunctions. The second and third eigenfunctions in this ordering are $\phi_{3,N}$ and $\phi_{5,N}$ (not shown here), with 0.052 and 0.065 fractional explained variance, respectively. These and other eigenfunctions not discussed here exhibit similarly high skill to the VSA patterns in Figs. 2 and 3 in capturing coherent patterns generated by the KS system with intermittency in both space and time.

In contrast, while the patterns from NLSA are seen to exhibit strong timescale separation (as expected theoretically at large Q [50, 51]), they are clearly less efficient in

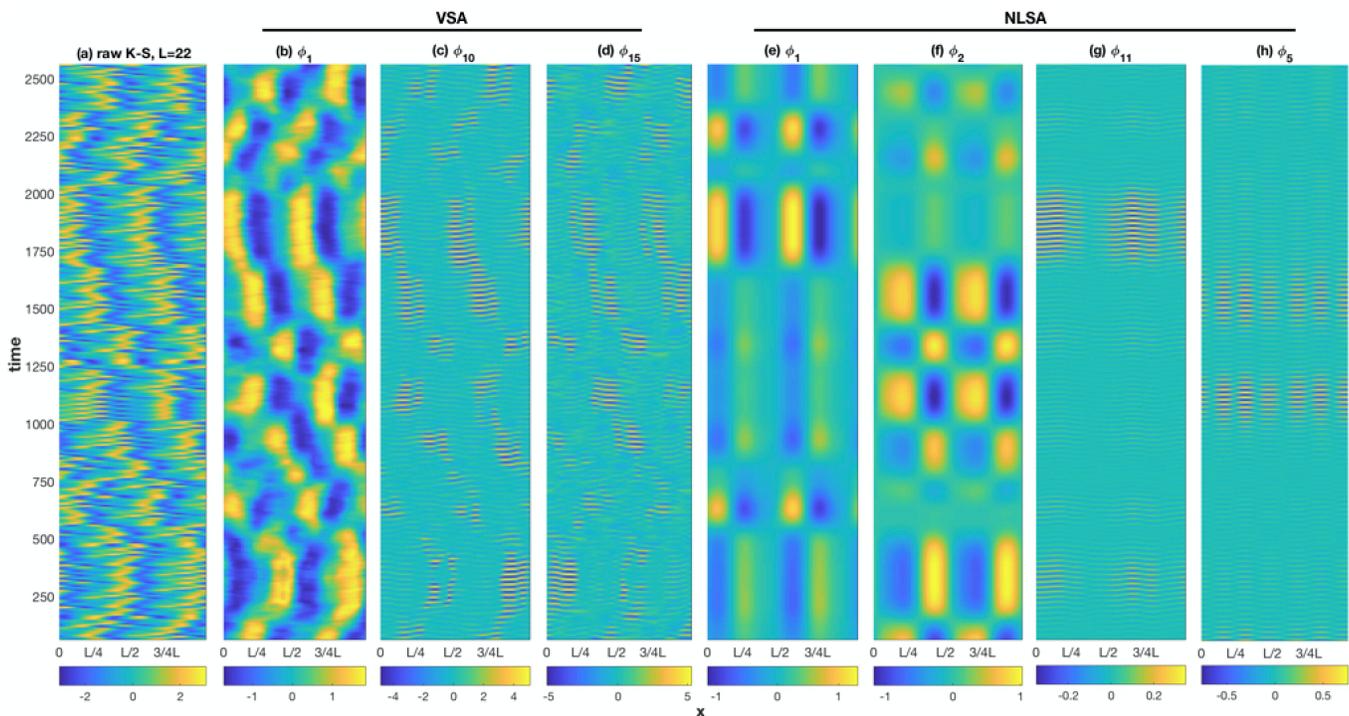


FIG. 2. Raw data (a) and representative spatiotemporal patterns obtained through VSA (b–d) and NLSA (e–h) from the KS model with $L = 22$ using $Q = 500$ delays. The VSA patterns (b–d) are vector-valued observables $\vec{\phi}_{k,N}$ determined from the eigenfunctions of the data-driven Markov operator $P_{Q,N}$ associated with the Markov kernel in (30). The NLSA patterns are constructed from temporal patterns determined from the eigenfunctions of a Markov operator analogous to $P_{Q,N}$, but acting on the Hilbert space H_A of scalar-valued observables, together with the corresponding spatial patterns as described in [31]. The VSA pattern in (b) captures the $O(2)$ family of the wavenumber $L/2$ structures that the KS system develops at this dynamical regime through a single eigenfunction. Notice the manifestly non-separable character of this pattern with respect to the spatial and temporal coordinates. The VSA patterns in (c, d) capture smaller-scale waves embedded within the structures in (b). The NLSA patterns in (e–h) exhibit a low-rank, separable behavior in space and time, and while they appear to capture the characteristic timescales of the large- (e, f) and small-scale waves (g, h), they are not representative of the intermittent character of the signal in space.

capturing the salient spatial features of the input signal. Consider, for example, the leading two NLSA patterns shown in Fig. 2(e,f). These patterns are clearly associated with the $O(2)$ family of wavenumber $L/2$ structures in the raw data, but because they have low-rank by construction, they are unable to represent the intermittent spatial translations of these patterns produced by chaotic dynamics in this regime. Their fractional explained variance is 0.13 and 0.15, respectively. Qualitatively, it appears that the NLSA patterns in Fig. 2(e, f) isolate periods during which the wavenumber $L/2$ structures are quasistationary, and translated relative to each other by $L/4$. In other words, it appears that NLSA captures the unstable equilibria that the system visits in the analysis time period through individual patterns, but does not adequately represent the transitory behavior associated with homoclinic orbits connecting this family of equilibria. Moreover, due to the presence of a continuous symmetry, a complete description of the spatiotemporal signal associated with the wavenumber $L/2$ structures would require a large number of modes. In contrast, VSA

effectively captures this dynamics through a small set of leading eigenfunctions.

As a final example, we extract spatiotemporal patterns from a KS system with a larger domain size, $L = 94$, using $N_X = 10,000$ samples and $Q = 500$ delays as in the $L = 22$ experiments. The raw data and representative VSA patterns recovered in this regime are displayed in Fig. 4. As is evident from Fig. 4(a), the dynamical complexity in this regime is markedly higher than for $L = 22$, as multiple traveling and quasistationary waves can now be accommodated in the domain, resulting in a spatiotemporal signal with high intermittency in both space and time. Despite this higher dynamical complexity, the recovered eigenfunctions, examples of which are shown in Fig. 4(b–f), decompose the signal into a pattern $\phi_{1,N}$ that captures the evolution of unstable fixed points and the homoclinic connections between them, and other patterns, $\phi_{3,N}$, $\phi_{5,N}$, $\phi_{8,N}$, and $\phi_{11,N}$, that are dominated by traveling waves. The fractional variances associated with these patterns are 6.0×10^{-3} ($\phi_{1,N}$), 0.020 ($\phi_{3,N}$), 0.040 ($\phi_{5,N}$), 0.067

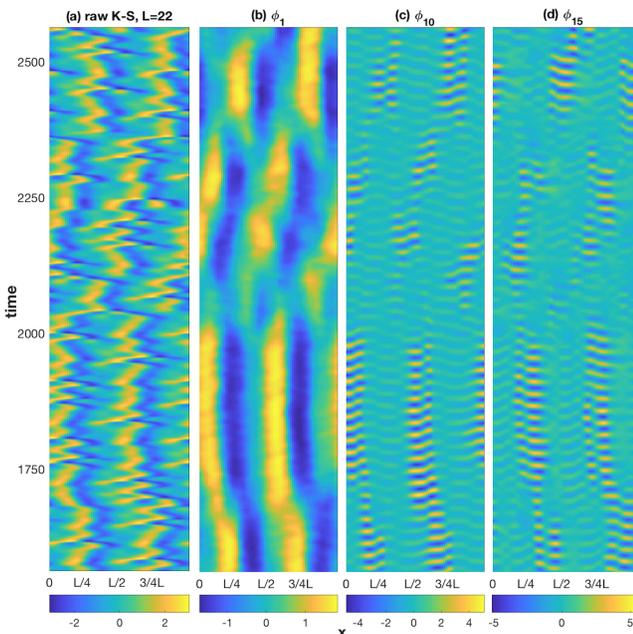


FIG. 3. As in Fig. 2(a-d), but for the time interval [1625, 2625] in the training dataset, highlighting the features of the small-scale waves in (c, d).

($\phi_{8,N}$), and 0.066 ($\phi_{11,N}$); that is, in this regime the traveling wave patterns are dominant in terms of explained variance. In general, the variance explained by individual eigenfunctions at $L = 94$, is smaller than those identified for $L = 22$, consistent with the higher dynamical complexity of the former regime.

It is worthwhile noting that $L = 94$ eigenfunction $\phi_{1,N}$ bears some qualitative similarities with the covariant Lyapunov vector (CLV) patterns identified at a nearby ($L = 96$) regime of the KS system in [69] (see Fig. 2 of that reference). Other VSA patterns, not shown here, also display qualitatively similar features to $\phi_{1,N}$ and to CLVs. However, presently we have not established a theoretical correspondence between the two techniques, so these similarities should be interpreted with caution.

In separate calculations, we have confirmed that NLSA applied to the $L = 94$ dataset behaves qualitatively similar to the $L = 22$ example described above. We have also verified that the VSA patterns from the $L = 22$ regime are robust under corruption of the input data by i.i.d. Gaussian noise of variance up to 30% and 40% of the raw signal variance, respectively for the small-scale waves in Fig. 2(c-d) and the large-scale structures in Fig. 2(b). On the basis of these results, we expect the method to be useful in recovering intermittent spatiotemporal patterns from noisy observational data, such as satellite brightness temperature data for atmospheric convection [78–80].

VIII. CONCLUSIONS

In this paper, we have presented a method for extracting spatiotemporal patterns from complex dynamical systems, called Vector-valued Spectral Analysis (VSA), which combines aspects of the theory of operator-valued kernels for machine learning with delay-coordinate maps of dynamical systems. A key element of this approach is that it operates directly on spaces of vector-valued observables appropriate for dynamical systems generating spatially extended patterns. This allows extraction of spatiotemporal patterns through eigenfunctions of kernel integral operators with far more general structure than those captured by pairs of temporal and spatial modes in conventional eigendecomposition techniques utilizing scalar-valued kernels. In particular, VSA enables efficient and physically meaningful decomposition of signals with intermittency in both space and time, while naturally factoring out symmetries present in the data. In addition, through the incorporation of delay-coordinate maps, the recovered patterns lie, in an asymptotic limit of infinitely many delays, in finite-dimensional invariant subspaces of observables associated with the point spectrum of the Koopman operator of the system. This endows these patterns with high dynamical significance and ability to decompose multiscale signals into distinct coherent modes. We demonstrated with applications to the Kuramoto-Sivashinsky model in chaotic regimes that VSA recovers intermittent patterns such as homoclinic orbits associated with translation families of unstable fixed points and traveling waves, with significantly higher skill than comparable eigendecomposition techniques operating on spaces of scalar-valued observables. It is expected that this framework will have applicability across disciplines dealing with complex spatiotemporal data.

Appendix A: Basic properties of Koopman operators and their eigenfunctions

In this appendix, we outline some of the basic properties of the Koopman operator U^t acting on scalar-valued observables in H_A and its lift \tilde{U}^t acting on scalar-valued observables in H_M (and, by the isomorphism $H \simeq H_M$, on vector-valued observables in H). Additional details on these topics can be found in one of the many references in the literature on ergodic theory, e.g., [42–44].

We begin by noting that for the class of measure-preserving dynamical systems generated by Lipschitz-continuous vector fields studied here (see Section II), the group $U = \{U^t\}_{t \in \mathbb{R}}$ of Koopman operators is a strongly continuous unitary group. This means that for every $f \in H_A$, the map $t \mapsto U^t f$ is continuous with respect to the H_A norm at every $t \in \mathbb{R}$. This in turn implies that there exists an unbounded, skew-adjoint operator $V : D(V) \mapsto H_A$ with dense domain $D(V) \subset H_A$, called the generator of U , such that $U^t = e^{tV}$. This operator completely characterizes the group U . Its action on

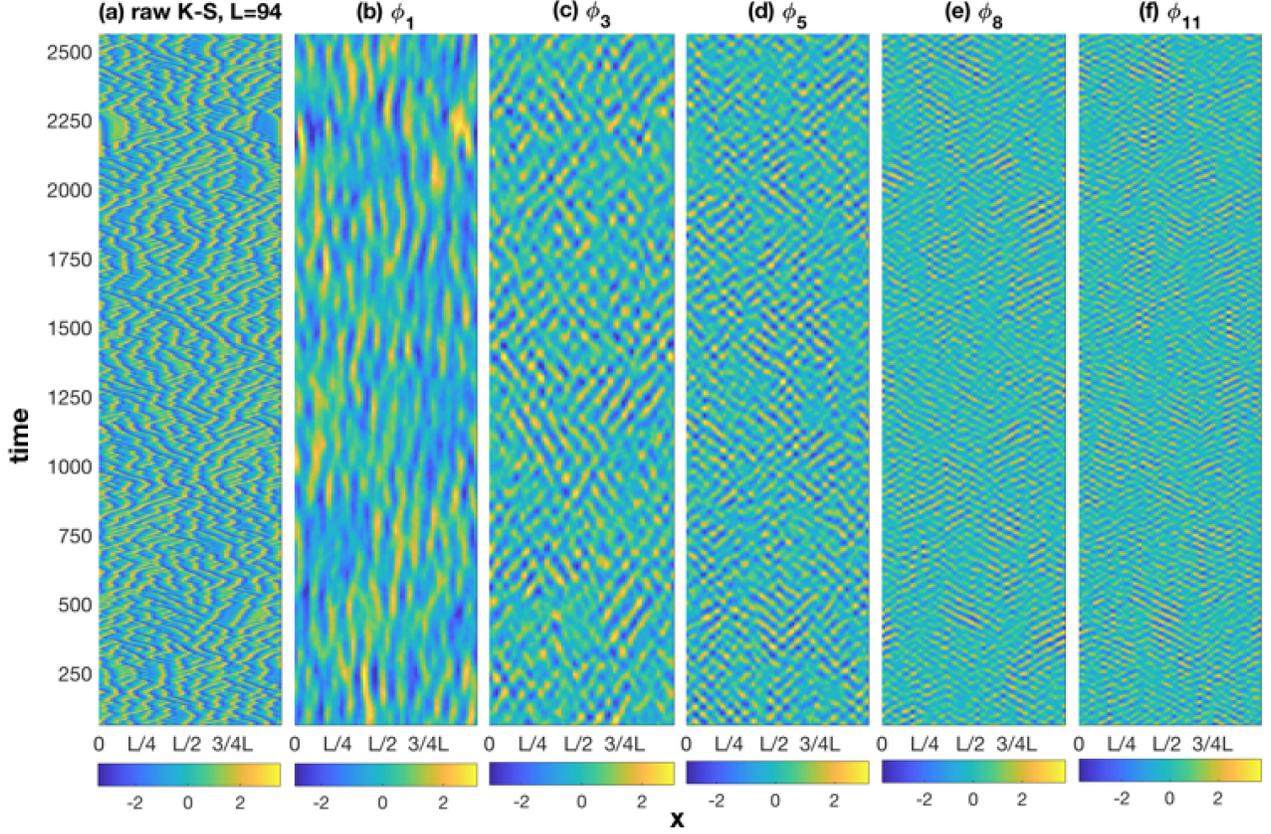


FIG. 4. As in Fig. 2(a–d), but for the KS system with $L = 94$.

observables $f \in D(V)$, is given by

$$Vf = \lim_{t \rightarrow 0} \frac{f \circ \Phi^t - f}{t},$$

where the limit is taken with respect to the H_A norm. If f is a differentiable function in $C^1(X)$, then $Vf = v(f)$, where v is the dynamical vector field.

A distinguished class of observables in H_A are the eigenfunctions of the generator of the Koopman group. Every such eigenfunction, z_j , satisfies the equation

$$Vz_j = i\alpha_j z_j,$$

where α_j is a real frequency, intrinsic to the dynamical system. In the presence of ergodicity (assumed here), all eigenvalues of V are simple, and eigenfunctions corresponding to distinct eigenvalues are orthogonal. Moreover, the eigenfunctions can be normalized so that $|z_j(x)| = 1$ for μ -almost every $x \in X$. That is, Koopman eigenfunctions of ergodic dynamical systems can be normalized to take values on the unit circle in the complex plane, much like the functions $e^{i\omega t}$ in Fourier analysis.

Every eigenfunction z_j of V at eigenvalue $i\alpha_j$ is also an eigenfunction of U^t , corresponding to the eigenvalue $\Lambda_j^t = e^{i\alpha_j t}$. This means that along an orbit of the dynamical

system, z_j evolves purely by multiplication by a periodic phase factor, viz.

$$U^t z_j(x) = z_j(\Phi^t(x)) = e^{i\alpha_j t} z_j(x),$$

where the equality holds for μ -almost every $x \in X$. This property makes Koopman eigenfunctions highly predictable observables, which warrant identification from data. In general, the evolution of any observable f lying in the closed subspace $\mathcal{D}_A = \overline{\text{span}} z_j$ of H_A spanned by Koopman eigenfunctions has the closed-form expansion

$$U^t f = \sum_j e^{i\omega_j t} c_j z_j, \quad c_j = \langle z_j, f \rangle_{H_A}. \quad (\text{A1})$$

This shows that the evolution of observables in \mathcal{D}_A can be characterized as a countable sum of Koopman eigenfunctions with time-periodic phase factors.

Koopman eigenvalues and eigenfunctions of ergodic systems also have an important group property; namely, if z_j and z_k are eigenfunctions of V at eigenvalue $i\omega_j$ and $i\omega_k$, respectively, then the product $z_j z_k$ is also an eigenfunction, corresponding to the eigenvalue $i(\alpha_j + \alpha_k)$. Thus, the eigenvalues and eigenfunctions of the Koopman generator form groups, with addition of complex numbers and multiplication of complex-valued functions acting as the group operations, respectively. If, in addition, the eigenfunctions are continuous (which is assumed

here), these groups are finitely generated. Specifically, in that case there exists a finite collection of rationally-independent eigenfrequencies $\tilde{\alpha}_1, \dots, \tilde{\alpha}_l$, such that every eigenfrequency has the form $\alpha_j = \sum_{i=1}^l j_i \tilde{\alpha}_i$, where $j = (j_1, \dots, j_l)$ is a vector of integers. Moreover, the Koopman eigenfunction corresponding to eigenfrequency α_j is given by $z_j = \tilde{z}_1^{j_1} \dots \tilde{z}_l^{j_l}$, where $\tilde{z}_1, \dots, \tilde{z}_l$ are the eigenfunctions corresponding to $\tilde{\alpha}_1, \dots, \tilde{\alpha}_l$, respectively. It follows from these facts in conjunction with (A1) that the evolution of every observable in \mathcal{D}_A can be uniquely determined given knowledge of finitely many Koopman eigenfunctions and their corresponding eigenfrequencies.

Yet, despite these attractive properties, in typical systems, not every observable will admit a Koopman eigenfunction expansion as in (A1); that is, \mathcal{D}_A will generally be a strict subspace of H_A . In such cases, we have the orthogonal decomposition

$$H_A = \mathcal{D}_A \oplus \mathcal{D}_A^\perp, \quad (\text{A2})$$

which is invariant under the action of U^t for all $t \in \mathbb{R}$. For observables in the orthogonal complement \mathcal{D}_A^\perp of \mathcal{D}_A , dynamical evolution is not determined by (A1), but rather by a spectral expansion involving a continuous spectral density (intuitively, an uncountable set of frequencies). This evolution can exhibit the characteristic behaviors associated with chaotic dynamics, such as decay of temporal correlations. In particular, it can be shown that for any $f \in \mathcal{D}_A$ and any $g \in \mathcal{D}_A^\perp$, the cross-correlation function, defined by

$$C_{fg}(x, x', t) = \frac{1}{t} \int_0^t U^s f^*(x) U^s g(x') ds,$$

vanishes as $|t| \rightarrow \infty$ for μ -almost every $x, x' \in X$. Note that this relationship does not necessarily hold for $f \in \mathcal{D}_A^\perp$.

We now turn to the unitary group $\tilde{U} = \{\tilde{U}^t\}_{t \in \mathbb{R}}$ associated with the Koopman operators \tilde{U}^t on H_M . As discussed in Section II, these operators are obtained by a trivial lift $\tilde{U}^t = U^t \otimes I_{H_Y}$ of the Koopman operators on H_A . Correspondingly, the group \tilde{U} is generated by the densely-defined, skew-adjoint operator $\tilde{V} : D(\tilde{V}) \mapsto H_M$, such that

$$\tilde{V} = V \otimes I_{H_Y}, \quad \tilde{V}f = \lim_{t \rightarrow 0} \frac{f \circ \tilde{\Phi}^t - f}{t}, \quad \tilde{\Phi}^t = \Phi^t \otimes I_Y. \quad (\text{A3})$$

Similarly, the decomposition in (A2) extends to the orthogonal decomposition

$$H_M = \mathcal{D}_M \oplus \mathcal{D}_M^\perp, \quad \mathcal{D}_M = \mathcal{D}_A \otimes H_Y, \quad \mathcal{D}_M^\perp = \mathcal{D}_A^\perp \otimes H_Y,$$

which is invariant under \tilde{U}^t for all $t \in \mathbb{R}$.

Consider now the eigenvalue problem for \tilde{U}^t ,

$$\tilde{U}^t \tilde{z} = \tilde{\Lambda}^t \tilde{z}, \quad \tilde{\Lambda}^t \in \mathbb{C}, \quad \tilde{z} \in H_M. \quad (\text{A4})$$

As stated in Section II, the eigenvalues of \tilde{U}^t are identical to those of U^t . In addition, it can be shown that every

eigenfunction \tilde{z} of \tilde{U}^t at eigenvalue Λ^t has the form

$$\tilde{z} = z \otimes f, \quad (\text{A5})$$

where $z \in H_A$ is an eigenfunction of U^t at the same eigenvalue (unique up to normalization), and f an arbitrary spatial pattern in H_Y . To see this, expand \tilde{z} in a tensor product basis $\{\varphi_k \otimes \psi_l\}$ of H_M ,

$$\tilde{z} = \sum_{kl} \tilde{c}_{kl} \varphi_k \otimes \psi_l, \quad \tilde{c}_{kl} \in \mathbb{C},$$

where $\{\varphi_k\}$ and $\{\psi_l\}$ are orthonormal bases of H_A and H_Y , respectively. Substituting this expansion in (A4), and taking inner products with $\varphi_i \otimes \varphi_j$, shows that the expansion coefficients \tilde{c}_{kl} must satisfy the equations

$$\sum_k U_{ik} \tilde{c}_{kj} = \tilde{\Lambda}^t \tilde{c}_{ij}, \quad (\text{A6})$$

where $U_{ik} = \langle \varphi_i, U^t \varphi_j \rangle_{H_A}$ are the ‘‘matrix elements’’ of the Koopman operator U^t with respect to the $\{\varphi_k\}$ basis. Comparing this to the analogous representation of the eigenvalue problem for U^t , i.e.,

$$U^t z = \Lambda^t z \iff \sum_k U_{ik} c_k = \Lambda^t c_k, \quad (\text{A7})$$

where $z = \sum_k c_k \varphi_k$, it follows that for each j , the \tilde{c}_{ij} in (A6) solve the same eigenvalue problem as the c_i in (A7). This implies that $\tilde{\Lambda}^t = \Lambda^t$, and (assuming all solutions are normalized to unit L^2 norm) the \tilde{c}_{kl} can be expressed as $\tilde{c}_{kl} = c_k e^{i\theta_l}$, where $\theta_l \in [0, 2\pi)$ are arbitrary phase angles. We therefore have

$$\tilde{z} = \sum_{kl} c_k e^{i\theta_l} \varphi_k \otimes \psi_l = z \otimes \left(\sum_l e^{i\theta_l} \psi_l \right),$$

which is consistent with (A5).

Appendix B: Choice of scaling function

As stated in Section III B, our distance scaling function s_Q is based on the corresponding function introduced in [50], which has dependencies on both the local sampling density and time tendency of the data. We will first introduce these dependencies, and subsequently construct s_Q and its data-driven approximation $s_{Q,N}$.

1. Local density function

We begin by introducing the kernel integral operator $\bar{K}_Q : C(\Omega) \mapsto C(\Omega)$ associated with the un-scaled kernel $\bar{k}_Q : \Omega \times \Omega \mapsto \mathbb{R}_+$ with $k_Q(\omega, \omega') = e^{-d_Q^2(\omega, \omega')/\epsilon}$, $\epsilon > 0$, viz.

$$\bar{K}_Q f = \int_{\Omega} k_Q(\cdot, \omega) f(\omega) d\rho(\omega).$$

Following [27], we define the function

$$\sigma_Q = \bar{K}_Q 1_\Omega = \int_M \bar{k}_Q(\cdot, \omega) d\rho(\omega), \quad (\text{B1})$$

which is continuous, strictly positive, and restricted to M (or the compact set $\mathcal{V} \supseteq M$ employed in the data-driven approximations in Section VI and Appendix D), is bounded away from zero. It is a standard result from the theory of kernel density estimation that if M is a smooth, m -dimensional Riemannian manifold and the delay-coordinate observation map $F_Q : M \mapsto \mathbb{R}^Q$ (see Section IV A) is an embedding, then, as $\epsilon \rightarrow 0$, the quantity $\bar{\sigma}_Q(\omega) = \sigma_Q(\omega)/(2\pi\epsilon^{m/2})$ converges for every $\omega \in M$ to the density $\frac{d\rho}{d\text{vol}}(\omega)$ of the measure ρ with respect to the Riemannian measure vol on M . Moreover, $\bar{\sigma}_Q$ has physical dimension (units) of length $^{-m}$, and as a result $\bar{\sigma}_Q^{-1/m}$ assigns a characteristic length at each point in M . Here, we do not assume that M has the structure of a smooth manifold, so we will not be taking $\epsilon \rightarrow 0$ limits. However, due to the exponential decay of $\bar{k}_Q(\omega, \omega')$ with respect to $d_Q(\omega, \omega')$, we can still interpret σ_Q from (B1) as a local density-like function.

In the data-driven setting, we introduce the finite-rank operator $\bar{K}_{Q,N} : C(\Omega) \mapsto C(\Omega)$ with

$$\bar{K}_{Q,N} f = \int_\Omega \bar{k}_Q(\cdot, \omega) f(\omega) d\rho_N(\omega),$$

and approximate σ_Q by the continuous function

$$\sigma_{Q,N} = \bar{K}_{Q,N} 1_\Omega = \frac{1}{N} \sum_{n=0}^{N-1} \bar{k}_Q(\cdot, \omega_n). \quad (\text{B2})$$

Uniform convergence of $\sigma_{Q,N}$ to σ_Q will be established in Appendix D 2.

2. Phase space velocity

Throughout this section, we will assume that the point-wise observation map $F_y : X \mapsto \mathbb{R}$ is continuously differentiable for every $y \in Y$, which is equivalent to assuming that the vector-valued observation map F lies in the space $C^1(X; C(Y))$. This assumption is natural for a wide class of observation maps encountered in applications. In particular, it implies that the “energy” of the signal, expressed in terms of the Koopman generator \tilde{V} from Appendix A as $\int_M |\tilde{V}(F)|^2 d\rho$ is finite. Under this condition, the phase space speed function $\xi_Q : \Omega \mapsto \mathbb{C}$, defined as

$$\xi_Q^2((x, y)) = \frac{1}{Q} \sum_{q=0}^{Q-1} \zeta^2((\Phi^{-q\tau}(x), y)), \quad (\text{B3})$$

where

$$\begin{aligned} \zeta((x, y)) &= |\tilde{V}F((x, y))| \\ &= \left| \lim_{t \rightarrow 0} \frac{F((\Phi^t(x), y)) - F((x, y))}{t} \right|, \end{aligned}$$

is continuous. Note that $\xi_Q((x, y))$ may vanish, e.g., as stated in Section III B, if y lies in the boundary of Y and F obeys time-independent boundary conditions. In the special case $Q = 1$, ξ_Q will also vanish at local maxima/minima of the signal with respect to time.

Phase space speed functions analogous to ξ_Q were previously employed in NLSA [31] and related kernel [26] and Koopman operator techniques [50]. In what follows, we will adopt the approach introduced in [50, Section 6], which utilizes ξ_Q in such a way so that if $\xi_Q(\omega)$ is zero then $s_Q(\omega)$ vanishes too. In [50], the motivation for constructing s_Q in this manner was to “remove” the effects of fixed points of the dynamics which were found to adversely impact the conditioning of data-driven kernel techniques. The role of ξ_Q is analogous in the setting of vector-valued observables of interest here, where this function can vanish not only at fixed points of the dynamics, but also at local extrema and/or boundary points as stated above.

Following [26, 31, 50], in the data-driven setting we approximate ζ by a continuous function $\zeta_\tau : \Omega \mapsto \mathbb{R}$ that provides a finite-difference approximation of ζ with respect to the sampling interval τ . As a concrete example, we consider a central difference scheme,

$$\zeta_\tau((x, y)) = \frac{|F((\Phi^\tau(x), y)) - F((\Phi^{-\tau}(x), y))|}{2\tau},$$

which, under the assumed differentiability properties of F , converges as $\tau \rightarrow 0$ to ζ , uniformly on compact subsets of Ω (including M and \mathcal{V}). We will also consider that the sampling interval is specified as a function $\tau(N)$ such that $\tau(N) \rightarrow 0$ and $N\tau(N) \rightarrow \infty$ as $N \rightarrow \infty$. With this assumption, the limit $N \rightarrow \infty$ corresponds to infinitely short sampling interval (required for convergence of finite difference schemes) and infinitely long total sampling time (required for convergence of ergodic averages). Hereafter, we will use the notation $\xi_{Q,N}$ for the function resulting by substituting ζ by $\zeta_{\tau(N)}$ in (B3).

3. Scaling function

With the local density and phase space speed functions from Appendices B 1 and B 2, we define the continuous scaling functions

$$s_Q = \left(\frac{\sigma_Q \xi_Q}{\bar{\xi}_Q} \right)^\gamma, \quad s_{Q,N} = \left(\frac{\sigma_{Q,N} \xi_{Q,N}}{\bar{\xi}_{Q,N}} \right)^\gamma, \quad (\text{B4})$$

where γ is a positive parameter, and $\bar{\xi}_Q = \int_M \xi_Q d\rho$ and $\bar{\xi}_{Q,N} = \int_\Omega \xi_{Q,N} d\rho_N$ are the mean values of ξ_Q and $\xi_{Q,N}$, respectively. This definition is motivated by [50], where it was shown that an analogous scaling function employed in scalar-valued kernels for ergodic dynamical systems on compact Riemannian manifolds can be interpreted, for an appropriate choice of γ and in a suitable limit of vanishing kernel bandwidth parameter ϵ , as a conformal

change of Riemannian metric that depends on the vector field of the dynamics.

More specifically, a Markov operator analogous to P_Q from Section III C was shown to approximate the heat semigroup generated by a Laplace-Beltrami operator associated with this conformally transformed metric. In [50], this change of geometry was associated with a rescaling of the vector field of the dynamics (i.e., a time change of the dynamical system [81]) that was found to significantly improve the conditioning of kernel algorithms if the system has fixed points. In particular, for a dataset consisting of finitely many samples, the sampling of the phase space manifold near a fixed point will become highly anisotropic, as most of the near neighbors of datapoints close to the fixed point will lie along the sampled orbit of the dynamics (which is a one-dimensional set), and the directions transverse to the orbit will be comparatively undersampled. The latter is because the phase speed of the system becomes arbitrarily small near a fixed point, meaning that most geometrical nearest neighbors of a data point in its vicinity will lie on a single orbit. Choosing the scaling function (analogous to s_Q) such that it vanishes at the fixed point is tantamount to increasing the bandwidth $1/s_Q$ of the kernel by arbitrarily large amounts (see (7)), thus improving sampling in directions transverse to the orbit.

While the arguments above are strictly valid only in the smooth manifold case (as they rely on $\epsilon \rightarrow 0$ limits), s_Q in (B4) should behave similarly in regions of the product space Ω where the rate-of-change of the observed data (measured by ξ_Q in (B3)) is small. As stated in Section III C, ξ_Q can vanish or be small not only at fixed points of the dynamics on X , but also at points $y \in Y$ where the observable F_y is constant or nearly constant (e.g., near domain boundaries).

What remains for a complete specification of s_Q is to set the exponent parameter γ . According to the discussion in Appendix B 1, if M is an m -dimensional smooth manifold, setting $\gamma = 1/m$ has the effect of normaliz-

ing the distance function d_Q to a dimensionless quantity \tilde{d}_Q . The conformal change of metric in [50] was also performed with that choice of γ . In practice, M will not be a smooth manifold, but we can still assign to it an effective dimension by examining the dependence of the kernel integrals $\kappa = \int_{M \times M} \bar{k}_Q d\rho \otimes d\rho$ (or the corresponding data-driven quantity $\kappa_N = \int_{\Omega \times \Omega} \bar{k}_Q d\rho_N \otimes d\rho_N$) as a function of the bandwidth parameter ϵ . As shown in [27, 82], $d \log \kappa / d \log \epsilon$ can be interpreted as an effective dimension for M at the scale associated by the bandwidth parameter ϵ . This motivates an automatic bandwidth tuning procedure where ϵ is chosen as the maximizer of that function, and the corresponding maximum value \hat{m} provides an estimate of M 's dimension.

Here, we nominally set $\gamma = 1/\hat{m}$ with \hat{m} determined via the method just described. As discussed in Section VII, our results are not too sensitive with respect to changes of γ around that value. In fact, for the systems studied, our results remain qualitatively robust even if the velocity-dependent terms are not included in s_Q and $s_{Q,N}$; that is, qualitatively similar results can also be obtained using the scaling functions

$$s_Q = \sigma_Q^\gamma, \quad s_{Q,N} = \sigma_{Q,N}^\gamma.$$

Scaling functions of this class are continuous even if F is not continuously differentiable with respect to x .

4. Proof of Proposition 2

To prove the claim, it suffices to show that σ_Q and ξ_Q are both invariant functions under the group action Γ_Ω^g . That σ_Q is invariant follows from (B1) in conjunction with the fact that $\bar{k}_Q(\omega, \omega') = \bar{k}_Q(\omega, \Gamma_\Omega^\gamma(\omega'))$ for all $\omega, \omega' \in \Omega$ and $g \in G$; the latter is a direct consequence of Proposition 1. Next, to verify invariance of ξ_Q , fix any $x \in X$, $y \in Y$, and $g \in G$, and use (B3) to compute

$$\begin{aligned} \xi_Q^2(\Gamma_\Omega^g((x, y))) &= \frac{1}{Q} \sum_{q=0}^{Q-1} \left| \lim_{t \rightarrow 0} \frac{\Phi^{t-q\tau}(\Gamma_X^g(x))(\Gamma_Y^g(y)) - \Phi^{-q\tau}(\Gamma_X^g(x))(\Gamma_Y^g(y))}{t} \right|^2 = \frac{1}{Q} \sum_{q=0}^{Q-1} \left| \lim_{t \rightarrow 0} \frac{\Phi^{t-q\tau}(x)(y) - \Phi^{-q\tau}(x)(y)}{t} \right|^2 \\ &= \xi_Q^2((x, y)). \end{aligned}$$

This completes the proof of the Proposition.

Appendix C: Behavior of kernel integral operators in the limit of infinitely-many delays

In this appendix, we describe some of the properties of the d_∞ and \tilde{d}_∞ pseudometrics in (4) and (6), respectively, and the associated kernel integral operators K_∞ and P_∞ .

These properties follow by a similar analysis to that performed in [51] for pseudometrics on the state space X computed from delay-coordinate mapped snapshot data (i.e., without taking into account spatial dependencies on Y).

First, for any two points $y, y' \in Y$ the function $d_{Q,yy'} : X \times X \mapsto \mathbb{R}$ defined by

$$d_{Q,yy'}(x, x') = d_Q((x, y), (x', y'))$$

is continuous. Therefore, by the Birkhoff pointwise ergodic theorem, the limit $\lim_{Q \rightarrow \infty} d_{Q,yy'}(x, x') = d_\infty((x, y), (x', y'))$ exists for almost every $(x, x') \in X \times X$ with respect to the product measure $\mu \times \mu$. This implies that d_∞ is a well defined function in $L^2(M \times M, \rho \times \rho)$. Note that the latter is isomorphic as a Hilbert space to $H_M \otimes H_M$. Next, it follows by invariance of Birkhoff averages, that for every $(y, y') \in Y \times Y$, the function $d_{\infty,yy'} \in L^2(X \times X, \mu \times \mu)$ is invariant under the Koopman operator $U^t \otimes U^t$ of the product dynamical system on $X \times X$; i.e., $(U^t \otimes U^t)d_{\infty,yy'} = d_{\infty,yy'}$. Correspondingly, d_∞ is invariant under the Koopman operator $\tilde{U}^t \otimes \tilde{U}^t$; that is, $(\tilde{U}^t \otimes \tilde{U}^t)d_\infty = d_\infty$. Following [51], it can further be shown that for any $y, y' \in Y$, $d_{\infty,yy'}$ lies in the subspace $\mathcal{D}_A \otimes \mathcal{D}_A$ of $H_A \otimes H_A$ associated with the point spectrum of U^t (see Appendix A). This implies that d_∞ lies in $\mathcal{D}_M \otimes \mathcal{D}_M$.

Analogous results hold in the case of the rescaled pseudometric \tilde{d}_∞ by virtue of the fact that the scaling function $s_\infty = \lim_{Q \rightarrow \infty} s_Q$ from Appendix B is invariant under \tilde{U}^t . This is because (1) the density function $\sigma_\infty = \lim_{Q \rightarrow \infty} \sigma_Q$ from Appendix B1 is based on the kernel \tilde{k}_∞ , which is invariant under $\tilde{U}^t \otimes \tilde{U}^t$ by the invariance of d_∞ just established, and (2) the phase space speed function $\xi_\infty^2 = \lim_{Q \rightarrow \infty} \xi_Q^2$ is a Birkhoff average. It therefore follows that \tilde{d}_∞ and k_∞ are both invariant under $\tilde{U}^t \otimes \tilde{U}^t$, and lie in the $\mathcal{D}_M \otimes \mathcal{D}_M$ subspace of $H_M \otimes H_M$.

We now use these results to prove Theorem 4. Our calculation mimics that in the proof of Theorem 3 on the commutation between kernel integral operators and unitary symmetry operators. In particular, the invariance of d_∞ under $\tilde{U}^t \otimes \tilde{U}^t$, in conjunction with the fact that ρ is invariant under $\tilde{\Phi}^t$, implies that

$$\begin{aligned} K_\infty f(\omega) &= \int_M k_\infty(\omega, \omega') f(\omega') d\rho(\omega') \\ &= \int_M k_\infty(\omega, \tilde{\Phi}^t(\omega')) f(\tilde{\Phi}^t(\omega')) d\rho(\omega') \\ &= \int_M k_\infty(\tilde{\Phi}^{-t}(\omega), \omega') f(\tilde{\Phi}^t(\omega')) d\rho(\omega') \\ &= \tilde{U}^{-t} K_\infty \tilde{U}^t f(\omega). \end{aligned} \quad (\text{C1})$$

Therefore, for every $t \in \mathbb{R}$, K_∞ and \tilde{U}^t are commuting operators with

$$[\tilde{U}^t, K_\infty] = \tilde{U}^t K_\infty - K_\infty \tilde{U}^t = 0.$$

It also follows from the invariance of k_∞ under $\tilde{U}^t \otimes \tilde{U}^t$ that the right and left normalization functions r and l , respectively, are invariant under \tilde{U}^t , and thus that the Markov kernel p_∞ is again invariant under $\tilde{U}^t \otimes \tilde{U}^t$. Therefore, by an analogous calculation to that used to deduce (C1) we can deduce that $[\tilde{U}^t, P_\infty] = 0$. This concludes our proof of Theorem 4.

Appendix D: Proof of Theorem 6

Our proof of Theorem D draws heavily on the spectral convergence results on data-driven kernel integral operators established in [23, 51], though it requires certain modifications appropriate for the class of variable bandwidth kernels in (5), which, to our knowledge, have not been previously discussed. In what follows, we provide explicit proofs of Claims (a) and (b) of the theorem; Claim (c) is a direct consequence of the definition of $\tilde{\phi}_{j,N}$ in (31). Throughout this appendix, all operators will act on the Banach space of continuous functions on \mathcal{V} equipped with the uniform norm, $\|\cdot\|_{C(\mathcal{V})}$. Therefore, for notational simplicity, we will drop the tildes from our notation for \tilde{P}_Q and $\tilde{P}_{Q,N}$.

1. Proof of Claim (a)

First, that $P_{Q,N}$ is compact follows immediately from the fact that it has finite rank. Showing that P_Q is compact is equivalent to showing that for any bounded sequence $f_n \in C(\mathcal{V})$, the sequence $g_n = P_Q f_n$ has a limit point in the uniform norm topology. Since \mathcal{V} is compact, it suffices to show that g_n is equicontinuous and bounded; in that case, the existence of a limit point of g_n is a consequence of the Arzelà-Ascoli theorem. Indeed, for any $\omega \in \mathcal{V}$, we have

$$\begin{aligned} |g_n(\omega)| &= \left| \int_M p_Q(\omega, \omega') f_n(\omega') d\rho(\omega') \right| \\ &\leq \int_M |p_Q(\omega, \omega') f_n(\omega')| d\rho(\omega') \\ &\leq \|p_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f_n\|_{C(\mathcal{V})} \\ &\leq \|p_Q\|_{C(\mathcal{V} \times \mathcal{V})} B, \end{aligned}$$

where $B = \sup_n \|f_n\|_{C(\mathcal{V})}$. This shows that g_n is uniformly bounded. Similarly, we have

$$\begin{aligned} |g_n(\omega) - g_n(\omega')| &\leq \|p_Q(\omega, \cdot) - p_Q(\omega', \cdot)\|_{C(\mathcal{V})} \|f_n\|_{C(\mathcal{V})} \\ &\leq \|p_Q\|_{C(\mathcal{V} \times \mathcal{V})} B \bar{d}_\Omega(\omega, \omega'), \end{aligned}$$

where \bar{d}_Ω is any metric on Ω compatible with its induced topology from X and Y (see Section II). This implies that the family $\{g_n\}$ is equicontinuous, and it follows from the Arzelà-Ascoli theorem that it has a limit point. We therefore conclude that P_Q is compact, as claimed.

2. Proof of Claim (b)

According to Definition 5, we must first show that for every $f \in C(\mathcal{V})$, $P_{Q,N} f$ converges to $P_Q f$ in the uniform norm; that is, we must show that $\lim_{N \rightarrow \infty} \eta_N = 0$, where

$$\eta_N = \|P_{Q,N} f - P_Q f\|_{C(\mathcal{V})}.$$

Defining the Markov kernel $\hat{p}_{Q,N} : C(\mathcal{V}) \times C(\mathcal{V}) \mapsto \mathbb{R}_+$,

$$\hat{p}_{Q,N} = \frac{k_Q}{l_{Q,N}r_{Q,N}}, \quad (\text{D1})$$

and the operators $\tilde{P}_{Q,N} : C(\mathcal{V}) \mapsto C(\mathcal{V})$ with

$$\begin{aligned} \tilde{P}_{Q,N}f &= \int_M p_Q(\cdot, \omega)f(\omega) d\rho_N(\omega), \\ \hat{P}_{Q,N}f &= \int_M \hat{p}_{Q,N}(\cdot, \omega)f(\omega) d\rho_N(\omega), \end{aligned}$$

we have

$$\begin{aligned} \eta_N &\leq \|P_Qf - \tilde{P}_{Q,N}f\|_{C(\mathcal{V})} \\ &\quad + \|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} \\ &\quad + \|\hat{P}_{Q,N}f - P_{Q,N}f\|_{C(\mathcal{V})}; \end{aligned} \quad (\text{D2})$$

that is, we can bound η_N by the sum of contributions due to (1) errors in approximating integrals with respect to the invariant measure ρ by the sampling measure ρ_N (the first term in the right-hand side), (2) errors in approximating the left and right normalization functions l_Q and r_Q by their data-driven counterparts, $l_{Q,N}$ and $r_{Q,N}$, respectively, and (3) errors in approximating the kernel k_Q by the data-driven kernel $k_{Q,N}$ (the third term in the right-hand side).

We first consider the first term,

$$\|P_Qf - \tilde{P}_{Q,N}f\|_{C(\mathcal{V})} = \max_{\omega \in \mathcal{V}} |\tilde{P}_{Q,N}f(\omega) - P_Qf(\omega)|.$$

By the assumptions stated in Section VIA (in particular (28)) in conjunction with the continuity of p_Q , it follows that $\tilde{P}_{Q,N}f(\omega)$ converges to $P_Nf(\omega)$, pointwise with respect to $\omega \in \mathcal{V}$; however, it is not necessarily the case that the convergence is uniform. For the latter, we need the stronger notion of a *Glivenko-Cantelli class*.

Definition 8. Let $\mathbb{E} : C(\mathcal{V}) \mapsto \mathbb{C}$ and $\mathbb{E}_N : C(\mathcal{V}) \mapsto \mathbb{C}$, be the expectation functions with respect to the measures ρ_N and ρ , respectively, i.e.,

$$\mathbb{E}f = \int_M f d\rho, \quad \mathbb{E}_Nf = \int_M f d\rho_N, \quad f \in C(\mathcal{V}).$$

Then, a set of functions $\mathcal{F} \in C(\mathcal{V})$ is said to be a Glivenko-Cantelli class if

$$\lim_{N \rightarrow \infty} \sup_{f \in \mathcal{F}} |\mathbb{E}f - \mathbb{E}_Nf| = 0.$$

Note, in particular, that if the set

$$\mathcal{F}_1 = \{p_Q(\omega, \cdot)f(\cdot) \mid \omega \in \mathcal{V}\}$$

can be shown to be a Glivenko-Cantelli class, then it will follow that $\|\tilde{P}_{Q,N}f - P_Qf\|_{C(\mathcal{V})}$ vanishes as $N \rightarrow \infty$. That this is indeed the case follows from Proposition 11 in [23].

Next, we turn to the second and third terms in (D2). To bound these terms, we first establish convergence of the data-driven distance scaling functions $s_{Q,N}$ to s_Q .

Lemma 9. *The scaling functions $s_{Q,N}$ from Appendix B converge uniformly as $N \rightarrow \infty$ to s_Q .*

Proof. It follows from the definition of s_Q and $s_{Q,N}$ in (B4) that for all $\omega \in \mathcal{V}$,

$$\begin{aligned} |s_{Q,N}(\omega) - s_Q(\omega)| &= |\sigma_{Q,N}(\omega)\xi_{Q,N}(\omega) - \sigma_Q(\omega)\xi_Q(\omega)|^\gamma \\ &\leq (|\sigma_{Q,N}(\omega) - \sigma_Q(\omega)||\xi_{Q,N}(\omega)| \\ &\quad + |\sigma_Q(\omega)||\xi_{Q,N}(\omega) - \xi_Q(\omega)|)^\gamma. \end{aligned}$$

Thus, since $\xi_{Q,N}$ converges uniformly to ξ_Q by continuous differentiability of the observation map F on the compact set \mathcal{V} (see Appendix B 2), $s_{Q,N}$ will converge uniformly to s_Q if $\sigma_{Q,N}$ converges uniformly to σ_Q . Indeed, because

$$|\sigma_{Q,N}(\omega) - \sigma_Q(\omega)| = |\mathbb{E}\bar{k}_Q(\omega, \cdot) - \mathbb{E}_N\bar{k}_Q(\omega, \cdot)|,$$

this will be the case if the set

$$\mathcal{F}_2 = \{\bar{k}_Q(\omega, \cdot) \mid \omega \in \mathcal{V}\}$$

is a Glivenko-Cantelli class. This follows from similar arguments as those used to establish that \mathcal{F}_1 is Glivenko-Cantelli. \square

Lemma 9, in conjunction with the continuity of the kernel shape function used throughout this work (see Section III B) implies in the following:

Corollary 10. *The data-driven kernel $k_{Q,N}$ converges uniformly to k_Q ; that is,*

$$\lim_{N \rightarrow \infty} \|k_{Q,N} - k_Q\|_{C(\mathcal{V} \times \mathcal{V})} = 0.$$

We now proceed to bound the second term in (D2), $\|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})}$. It follows from the definition of the kernels $p_{Q,N}$ and $\hat{p}_{Q,N}$ via (30) and (D1), respectively, that

$$\begin{aligned} \|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} &\leq \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \\ &\quad \times \left\| \frac{1}{l_{Q,N} \otimes r_{Q,N}} - \frac{1}{l_Q \otimes r_Q} \right\|_{C(\mathcal{V} \times \mathcal{V})}. \end{aligned}$$

By our assumptions on kernels stated in Section VIA, the functions l_Q , r_Q , $l_{Q,N}$ and $r_{Q,N}$ are bounded away from zero; therefore, there exists a constant $c > 0$, independent of N , such that

$$\begin{aligned} \|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} &\leq c \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \\ &\quad \times \|l_Q \otimes r_Q - l_{Q,N} \otimes r_{Q,N}\|_{C(\mathcal{V} \times \mathcal{V})} \\ &= c \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \\ &\quad \times \|l_Q - l_{Q,N}\|_{C(\mathcal{V})} \|r_Q - r_{Q,N}\|_{C(\mathcal{V})}. \end{aligned}$$

Observe now that

$$\begin{aligned}
\|r_Q - r_{Q,N}\|_{C(\mathcal{V})} &= \max_{\omega \in \mathcal{V}} |r_Q(\omega) - r_{Q,N}(\omega)| \\
&= \max_{\omega \in \mathcal{V}} |\mathbb{E}k_Q(\omega, \cdot) - \mathbb{E}_N k_{Q,N}(\omega, \cdot)| \\
&\leq \max_{\omega \in \mathcal{V}} |\mathbb{E}k_Q(\omega, \cdot) - \mathbb{E}_N k_Q(\omega, \cdot)| \\
&\quad + \max_{\omega \in \mathcal{V}} |\mathbb{E}_N (k_{Q,N}(\omega, \cdot) - k_Q(\omega, \cdot))| \\
&\leq \max_{\omega \in \mathcal{V}} |\mathbb{E}k_Q(\omega, \cdot) - \mathbb{E}_N k_Q(\omega, \cdot)| \\
&\quad + \|k_{Q,N} - k_Q\|_{C(\mathcal{V} \times \mathcal{V})}.
\end{aligned}$$

Since $\|k_{Q,N} - k_Q\|_{C(\mathcal{V} \times \mathcal{V})}$ converges to zero by Corollary 10, it follows that

$$\lim_{N \rightarrow \infty} \|r_Q - r_{Q,N}\|_{C(\mathcal{V})} = 0 \quad (\text{D3})$$

if it can be shown that

$$\mathcal{F}_3 = \{k_Q(\omega, \cdot) \mid \omega \in \mathcal{V}\}$$

is a Glivenko-Cantelli class. The latter can be verified by means of similar arguments as those used to establish that \mathcal{F}_1 is Glivenko-Cantelli. Equation (D3), in conjunction with the fact that $\|l_Q - l_{Q,N}\|_{C(\mathcal{V})}$ is bounded, is sufficient to deduce that $\lim_{N \rightarrow \infty} \|\hat{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} = 0$.

We now turn to the third term in (D2), $\|\hat{P}_{Q,N}f - P_{Q,N}f\|$. We have

$$\|\hat{P}_{Q,N}f - P_{Q,N}f\|_{C(\mathcal{V} \times \mathcal{V})} \leq \|\hat{p}_{Q,N} - p_{Q,N}\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})},$$

and it follows from the definitions of $\hat{p}_{Q,N}$ and $p_{Q,N}$, in conjunction with the fact that the normalization functions $l_{Q,N}$ and $r_{Q,N}$ are both bounded away from zero, that there exists a strictly positive constant c such that

$$\|\hat{P}_{Q,N}f - P_{Q,N}f\|_{C(\mathcal{V} \times \mathcal{V})} \leq c \|k_Q - k_{Q,N}\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})}.$$

Thus, convergence of $\|\hat{P}_{Q,N}f - P_{Q,N}f\|$ follows Corollary 10.

In summary, we have shown that $\|P_Nf - \tilde{P}_{Q,N}f\|_{C(\mathcal{V})}$, $\|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})}$, and $\|\hat{P}_{Q,N}f - P_{Q,N}f\|_{C(\mathcal{V})}$ all converge to zero, which is sufficient to conclude that $\lim_{N \rightarrow \infty} \eta_N = 0$, and that that $P_{Q,N}f$ converges to P_Qf .

According to Definition 5, it remains to show that for any bounded sequence $f_N \in C(\mathcal{V})$, the sequence $g_N = (P_{Q,N} - P_Q)f_N$ has a limit point. This can be proved by an Arzelà-Ascoli argument as in the proof of claim (a) in conjunction with Glivenko-Cantelli arguments as in the proof of pointwise convergence above. We refer the reader to Proposition 13 in [23] for more details. This completes our proof of claim (b).

ACKNOWLEDGMENTS

D. G. acknowledges support from NSF EAGER grant 1551489, ONR YIP grant N00014-16-1-2649, NSF grant DMS-1521775, and DARPA grant HR0011-16-C-0116. J. S. and A. O. acknowledge support from NSF EAGER grant 1551489. Z. Z. received support from NSF grant DMS-1521775. We thank Shuddho Das for stimulating conversations.

-
- [1] M. P. Cross and P. C. Hohenberg, *Rev. Mod. Phys.* **65**, 851 (1993).
 - [2] G. Ahlers, S. Grossmann, and D. Loshe, *Rev. Mod. Phys.* **81**, 503 (2009).
 - [3] J. Slawinska, O. Pauluis, A. J. Majda, and W. W. Grabowski, *Mon. Wea. Rev.* **143**, 563 (2015).
 - [4] R. Fung, A. M. Hanna, O. Vendrell, S. Ramakrishna, T. Seideman, R. Santra, and A. Ourmazd, *Nature* **532**, 471 (2016).
 - [5] J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
 - [6] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, *Phys. Lett. A* **131**, 433 (1988).
 - [7] P. Constantin, C. Foias, B. Nicolaenko, and R. Témam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations* (Springer, New York, 1989).
 - [8] N. Aubry, R. Guyonnet, and R. Lima, *J. Stat. Phys.* **64**, 683 (1991).
 - [9] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, 1996) p. 440.
 - [10] D. S. Broomhead and G. P. King, *Phys. D* **20**, 217 (1986).
 - [11] R. Vautard and M. Ghil, *Phys. D* **35**, 395 (1989).
 - [12] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
 - [13] F. Takens, in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898 (Springer, Berlin, 1981) pp. 366–381.
 - [14] T. Sauer, J. A. Yorke, and M. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).
 - [15] J. C. Robinson, *Nonlinearity* **18**, 2135 (2005).
 - [16] E. R. Deyle and G. Sugihara, *PLoS ONE* **6**, e18295 (2011).
 - [17] B. Schölkopf, A. Smola, and K. Müller, *Neural Comput.* **10**, 1299 (1998).
 - [18] M. Belkin and P. Niyogi, *Neural Comput.* **15**, 1373 (2003).
 - [19] R. R. Coifman *et al.*, *Proc. Natl. Acad. Sci.* **102**, 7426 (2005).
 - [20] R. R. Coifman and S. Lafon, *Appl. Comput. Harmon. Anal.* **21**, 5 (2006).
 - [21] P. W. Jones, M. Maggioni, and R. Schul, *Proc. Natl. Acad. Sci.* **105**, 1803 (2008).
 - [22] A. Singer and R. R. Coifman, *J. Appl. Comput. Harmon. Anal.* **25**, 226 (2008).
 - [23] U. von Luxburg, M. Belkin, and O. Bousquet, *Ann. Stat.* **26**, 555 (2008).

- [24] R. Talmon and R. R. Coifman, Proc. Natl. Acad. Sci. **110**, 12535 (2013).
- [25] J. Portegies, “Embeddings of Riemannian manifolds with heat kernels and eigenfunctions,” (2014), 1311.7568.
- [26] D. Giannakis, SIAM J. Appl. Dyn. Sys. **14**, 556 (2015).
- [27] T. Berry and J. Harlim, J. Appl. Comput. Harmon. Anal. **40**, 68 (2015).
- [28] T. Berry and T. Sauer, “Consistent manifold representation for topological data analysis,” (2016), 1606.02353.
- [29] T. Berry and T. Sauer, J. Appl. Comput. Harmon. Anal. **40**, 439 (2016).
- [30] D. Giannakis and A. J. Majda, in *Conference on Intelligent Data Understanding 2011* (Mountain View, California, 2011).
- [31] D. Giannakis and A. J. Majda, Proc. Natl. Acad. Sci. **109**, 2222 (2012).
- [32] D. Giannakis and A. J. Majda, Stat. Anal. Data Min. **6**, 180 (2013).
- [33] T. Berry, R. Cressman, Z. Gregurić-Ferenček, and T. Sauer, SIAM J. Appl. Dyn. Sys. **12**, 618 (2013).
- [34] I. Mezić and A. Banaszuk, Phys. D. **197**, 101 (2004).
- [35] I. Mezić, Nonlinear Dyn. **41**, 309 (2005).
- [36] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, J. Fluid Mech. **641**, 115 (2009).
- [37] J. H. Tu, C. W. Rowley, C. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, J. Comput. Dyn. **1**, 391 (2014).
- [38] H. Arbabi and I. Mezić, “Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator,” (2016), 1611.06664.
- [39] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, Nat. Commun. **8** (2017), 10.1038/s41467-017-00030-8.
- [40] M. Dellnitz and O. Junge, SIAM J. Numer. Anal. **36**, 491 (1999).
- [41] G. Froyland and M. Dellnitz, SIAM J. Sci. Comput. **24**, 1839 (2003).
- [42] Y. G. Sinai, ed., *Dynamical Systems, Ergodic Theory and Applications*, 2nd ed., Encyclopedia of Mathematical Sciences, Vol. 100 (Springer, Berlin, 2000).
- [43] M. Budisić, R. Mohr, and I. Mezić, Chaos **22**, 047510 (2012).
- [44] T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, Graduate Texts in Mathematics, Vol. 272 (Springer, 2015).
- [45] P. J. Schmid and J. L. Sesterhenn, in *Bull. Amer. Phys. Soc., 61st APS meeting* (San Antonio, 2008) p. 208.
- [46] P. J. Schmid, J. Fluid Mech. **656**, 5 (2010).
- [47] C. Penland, Mon. Weather Rev. **117**, 2165 (1989).
- [48] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, “A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition,” J. Nonlinear Sci. (2015).
- [49] D. Giannakis, J. Slawinska, and Z. Zhao, J. Mach. Learn. Res. Proceedings **44**, 103 (2015).
- [50] D. Giannakis, Appl. Comput. Harmon. Anal. (2017), 10.1016/j.acha.2017.09.001, in press.
- [51] S. Das and D. Giannakis, Nonlinearity (2017), in review, 1706.08544.
- [52] L.-S. Young, J. Stat. Phys. **108**, 733 (2002).
- [53] N. Aubry, W.-Y. Lian, and E. S. Titi, SIAM J. Sci. Comput. **14**, 483 (1993).
- [54] C. A. Micchelli and M. Pontil, Neural Comput. **17**, 177 (2005).
- [55] A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying, J. Mach. Learn. Res. **9**, 1615 (2008).
- [56] C. Carmeli, E. De Vito, A. Toigo, and V. Umanità, Anal. Appl. **08**, 19 (2010).
- [57] G. Carlsson, Bull. Amer. Math. Soc. **46**, 255 (2009).
- [58] O. Yair, R. Talmon, R. R. Coifman, and I. G. Kevrekidis, “No equations, no parameters, no variables: Data, and the reconstruction of normal forms by learning informed observation geometries,” (2017).
- [59] Y. Kuramoto and T. Tsuzuki, Progr. Theor. Phys. **55**, 356 (1976).
- [60] G. I. Sivashinsky, Acta Astronaut. **4**, 1177 (1977).
- [61] D. V. Anosov and A. B. Katok, Trans. Moscow Math. Soc. **23**, 1 (1970).
- [62] L. Zelnik-Manor and P. Perona, in *Advances in Neural Information Processing Systems*, Vol. 17 (2004) pp. 1601–1608.
- [63] I. M. James, ed., *Handbook of Algebraic Topology* (North Holland, Amsterdam, 1995).
- [64] F. Chatelin, *Spectral Approximation of Linear Operators*, Classics in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 2011).
- [65] J. M. Greene and J. S. Kim, Phys. D **33**, 99 (1988).
- [66] D. Arbruster, J. Guckenheimer, and P. Holmes, SIAM J. Appl. Math. **49**, 676 (1989).
- [67] I. G. Kevrekidis, B. Nicolaenko, and J. C. Scovel, SIAM J. Appl. Math. **50**, 760 (1990).
- [68] P. Cvitanović, R. L. Davidchack, and E. Siminos, SIAM J. Appl. Dyn. Sys. **9**, 1 (2009).
- [69] K. A. Takeuchi, H.-L. Yang, F. Ginelli, G. Radons, and H. Chaté, Phys. Rev. E **84**, 046214 (2011).
- [70] C. Foias, B. Nicolaenko, G. R. Sell, and R. Témam, in *IMA Preprints Series*, 279 (University of Minnesota Digital Conservancy, 1986).
- [71] M. S. Jolly, I. G. Kevrekidis, and E. S. Titi, Phys. D **44**, 39 (1990).
- [72] J. C. Robinson, Phys. Lett. A **184**, 190 (1994).
- [73] M. S. Jolly, R. Rosa, and R. Temam, Adv. Differential Equations **5**, 33 (2000).
- [74] S. Tajima and H. S. Greenside, Phys. Rev. E **66**, 017205 (2002).
- [75] K. Lu, Q. Wang, and L.-S. Young, Memoirs of the American Mathematical Society **224**, 1 (2013).
- [76] Z. Lian, P. Liu, and K. Lu, J. Differential Equ. **261**, 1532 (2016).
- [77] P. Cvitanovic, R. Artuso, R. Mainieri, and G. Tanner, *Chaos: Classical and Quantum* (Niels Bohr Institute, Copenhagen, 2016).
- [78] J. Slawinska, O. Pauluis, A. J. Majda, and W. W. Grabowski, J. Atmos. Sci. **71**, 953 (2014).
- [79] J. Slawinska and D. Giannakis, in *Proceedings of the 6th International Workshop on Climate Informatics*, edited by A. Banerjee, W. Ding, J. Dy, V. Lyubchich, and A. Rhines (Boulder, Colorado, 2016) pp. 49–52.
- [80] J. Slawinska, E. Székely, and D. Giannakis, in *Proceedings of the Workshop on Mining Big Data in Climate and Environment, 17th SIAM International Conference on Data Mining* (Houston, Texas, April 27–29, 2017, 2017) 1711.02526.
- [81] A. Katok and J.-P. Thouvenot, in *Handbook of Dynamical Systems*, Vol. 1B, edited by B. Hasselblatt and A. Katok (North-Holland, Amsterdam, 2006) Chap. 11, pp. 649–743.
- [82] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, IEEE Trans. Image Process. **17**, 1891 (2008).