



# Spatiotemporal Pattern Extraction by Spectral Analysis of Vector-Valued Observables

Dimitrios Giannakis<sup>1</sup> · Abbas Ourmazd<sup>2</sup> · Joanna Slawinska<sup>2</sup> · Zhizhen Zhao<sup>3</sup>

Received: 3 March 2018 / Accepted: 13 April 2019 / Published online: 13 May 2019  
© The Author(s) 2019

## Abstract

We present 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 multitask machine learning with delay-coordinate maps of dynamical systems. In contrast to 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. Moreover, the kernel construction naturally quotients out dynamical symmetries in the data and exhibits an asymptotic commutativity property with the Koopman evolution operator of the system, enabling decomposition of multiscale signals into dynamically intrinsic patterns. Application of VSA to the Kuramoto–Sivashinsky model demonstrates significant performance gains in efficient and meaningful decomposition over eigendecomposition techniques utilizing scalar-valued kernels.

**Keywords** Spatiotemporal patterns · Dynamical systems · Spectral decomposition · Kernel methods · Koopman operators · Dynamical symmetries

**Mathematics Subject Classification** 28B05 · 35B36 · 37C80 · 37E99 · 37M10 · 47B34

---

Communicated by Paul Newton.

✉ Dimitrios Giannakis  
dimitris@cims.nyu.edu

Extended author information available on the last page of the article

## 1 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 (Cross and Hohenberg 1993; Ahlers et al. 2009; Fung et al. 2016). 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 (Constantin et al. 1989). 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) and principal component analysis (PCA) techniques for spatiotemporal pattern extraction have been based on the spectral properties of temporal and spatial covariance operators estimated from snapshot data (Aubry et al. 1991; Holmes et al. 1996). In singular spectrum analysis (SSA) and related algorithms (Broomhead and King 1986; Vautard and Ghil 1989; Ghil et al. 2002), combining this approach with delay-coordinate maps of dynamical systems (Packard et al. 1980; Takens 1981; Sauer et al. 1991; Robinson 2005; Deyle and Sugihara 2011) 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 harmonic analysis (Schölkopf et al. 1998; Belkin and Niyogi 2003; Coifman et al. 2005; Coifman and Lafon 2006; Singer 2006; von Luxburg et al. 2008; Berry and Harlim 2016; Berry and Sauer 2016) have led to techniques for 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) (Giannakis and Majda 2012), and independently in Berry et al. (2013), the diffusion maps algorithm (Coifman and Lafon 2006) 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 (Mezić and Banaszuk 2004; Mezić 2005; Rowley et al. 2009; Giannakis et al. 2015; Williams et al. 2015; Brunton et al. 2017; Das and Giannakis 2019; Giannakis 2017) and Perron–Frobenius (Dellnitz and Junge 1999; Froyland and Dellnitz 2000) operators governing the evolution of observables and probability measures, respectively, in dynamical systems (Budisić et al. 2012; Eisner et al. 2015). Koopman eigenfunction analysis is also related to the dynamic mode decomposition (DMD) algorithm (Schmid 2010) and linear inverse model techniques (Penland 1989). An advantage of these approaches is that they target operators defined intrinsically for the dynamical system generating the data, and thus able, in principle, to recover temporal and spatial patterns of higher physical interpretability and utility in predictive modeling than kernel-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 spectra) than kernel integral operators, hindering the stability and convergence of data-driven approximation techniques. These issues were recently addressed through an approximation scheme for the generator of the Koopman group with rigorous convergence guarantees (Giannakis 2017; Das and Giannakis 2019), utilizing a data-driven orthonormal basis of the  $L^2$  space associated with the invariant measure, acquired through diffusion maps. There, it was also shown that the eigenfunctions of kernel integral operators defined on delay-coordinate mapped data (e.g., the covariance and heat operators in SSA and NLSA, respectively) in fact converge to Koopman eigenfunctions in the limit of infinitely many delays, indicating a deep connection between these two branches of data analysis algorithms.

All of the techniques described above recover from the data a set of temporal patterns and a corresponding set of spatial patterns, sometimes referred to as “chronos” and “topos” modes, respectively (Aubry et al. 1991). In particular, for a dynamical system with a state space  $X$  developing patterns in a physical domain  $Y$ , each chronos mode,  $\varphi_j$ , corresponds to a scalar- (real- or complex-) valued function on  $X$ , and the corresponding topos mode,  $\psi_j$ , 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  $\varphi_j \otimes \psi_j$ , mapping pairs of points  $(x, y)$  in the product space  $\Omega = X \times Y$  to the number  $\varphi_j(x)\psi_j(y)$ . For a dynamical system possessing a compact invariant set  $A \subseteq X$  (e.g., an attractor) supporting an ergodic invariant measure, the chronos modes effectively become scalar-valued functions on  $A$ , which may be of significantly smaller dimension than  $X$ , increasing the robustness of approximation of these modes from finite datasets.

Evidently, for spatiotemporal signals  $F(x, y)$  of high complexity, tensor product patterns, with separable dependence on  $x$  and  $y$ , can be highly inefficient in capturing the properties of the input signal. That is, the number  $l$  of such patterns needed to recover  $F$  at high accuracy via a linear superposition

$$F \approx \sum_{j=0}^{l-1} \varphi_j \otimes \psi_j \quad (1)$$

is generally large, with none of the individual patterns  $\varphi_j \otimes \psi_j$  being 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 is that in the presence of nontrivial 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 (Aubry et al. 1993; Holmes et al. 1996).

Here, we present a framework for spatiotemporal pattern extraction, called vector-valued spectral analysis (VSA), designed to alleviate the shortcomings mentioned above. The fundamental underpinning of VSA is that time-evolving spatial patterns have a natural structure as vector-valued observables on the system’s state space, and

thus data analytical techniques operating on such spaces are likely to offer maximal descriptive efficiency and physical insight. We show that eigenfunctions of kernel integral operators on vector-valued observables, constructed by combining aspects of the theory of operator-valued kernels (Micchelli and Pontil 2005; Caponnetto et al. 2008; Carmeli et al. 2010) with delay-coordinate maps of dynamical systems (Packard et al. 1980; Takens 1981; Sauer et al. 1991; Robinson 2005; Deyle and Sugihara 2011): (a) Are superior to conventional algorithms in capturing signals with intermittency in both space and time; (b) Naturally incorporate any underlying dynamical symmetries, eliminating redundant modes and thus improving physical interpretability of the results; (c) Have a correspondence with Koopman operators, allowing detection of intrinsic dynamical timescales; and, (d) Can be stably approximated via data-driven techniques that provably converge in the asymptotic limit of large data.

The plan of this paper is as follows. Section 2 introduces the class of dynamical systems under study and provides an overview of data analysis techniques based on scalar kernels. In Sect. 3, we present the VSA framework for spatiotemporal pattern extraction using operator-valued kernels, and in Sect. 4 discuss the behavior of the method in the presence of dynamical symmetries, as well as its correspondence with Koopman operators. Section 5 describes the data-driven implementation of VSA. In Sect. 6, we present applications to the Kuramoto–Sivashinsky (KS) PDE model (Kuramoto and Tsuzuki 1976; Sivashinsky 1977) in periodic and chaotic regimes. Our primary conclusions are described in Sect. 7. Technical results, descriptions of basic properties of kernels and Koopman operators, pseudocode, and an overview of NLSA are collected in six appendices.

## 2 Background

### 2.1 Dynamical System and Spaces of Observables

We begin by introducing the dynamical system and the spaces of observables under study. The dynamics evolves by a  $C^1$  flow map  $\Phi^t : X \rightarrow X$ ,  $t \in \mathbb{R}$ , on a manifold  $X$ , possessing an ergodic, invariant, Borel probability measure  $\mu$  with compact support  $A \subseteq X$ . The system develops patterns on a spatial domain  $Y$ , which has the structure of a compact metric space, supporting a finite Borel measure (volume)  $\nu$ . As a natural space of vector-valued observables, we consider the Hilbert space  $H = L^2(X, \mu; H_Y)$  of square-integrable functions with respect to the invariant measure  $\mu$ , taking values in  $H_Y = L^2(Y, \nu)$ . That is, modulo sets of  $\mu$ -measure 0, the elements of  $H$  are functions  $\vec{f} : X \rightarrow H_Y$ , such that for any dynamical state  $x \in X$ ,  $\vec{f}(x)$  is a scalar (complex-valued) field on  $Y$ , square-integrable with respect to  $\nu$ . For every such observable  $\vec{f}$ , the map  $t \mapsto \vec{f}(\Phi^t(x))$  describes a spatiotemporal pattern generated by the dynamics. Given  $\vec{f}, \vec{f}' \in H$  and  $g, g' \in H_Y$ , the corresponding inner products on  $H$  and  $H_Y$  are given by  $\langle \vec{f}, \vec{f}' \rangle_H = \int_X \langle \vec{f}(x), \vec{f}'(x) \rangle_{H_Y} d\mu(x)$  and  $\langle g, g' \rangle_{H_Y} = \int_Y g^*(y)g'(y) d\nu(y)$ , respectively.

An important property of  $H$  is that it exhibits the isomorphisms

$$H \simeq H_X \otimes H_Y \simeq H_\Omega,$$

where  $H_X = L^2(X, \mu)$  and  $H_\Omega = L^2(\Omega, \rho)$  are Hilbert spaces of scalar-valued functions on  $X$  and the product space  $\Omega = X \times Y$ , square-integrable with respect to the invariant measure  $\mu$  and the product measure  $\rho = \mu \times \nu$ , respectively (the inner products of  $H_X$  and  $H_\Omega$  have analogous definitions to the inner product of  $H_Y$ ). That is, every  $\vec{f} \in H$  can be equivalently viewed as an element of the tensor product space  $H_X \otimes H_Y$ , meaning that it can be decomposed as  $\vec{f} = \sum_{j=0}^\infty \varphi_j \otimes \psi_j$  for some  $\varphi_j \in H_X$  and  $\psi_j \in H_Y$ , or it can be represented by a scalar-valued function  $f \in H_\Omega$  such that  $\vec{f}(x)(y) = f(x, y)$ . Of course, not every observable  $\vec{f} \in H$  is of pure tensor product form,  $f = \varphi \otimes \psi$ , for some  $\varphi \in H_X$  and  $\psi \in H_Y$ .

We consider that measurements  $\vec{F}(x_n)$  of the system are taken along a dynamical trajectory  $x_n = \Phi^{n\tau}(x_0)$ ,  $n \in \mathbb{N}$ , starting from a point  $x_0 \in X$  at a fixed sampling interval  $\tau > 0$  through a continuous vector-valued observation map  $\vec{F} \in H$ . We also assume that  $\tau$  is such that  $\mu$  is an ergodic invariant probability measure of the discrete-time map  $\Phi^\tau$ .

### 2.2 Separable Data Decompositions via Scalar Kernel Eigenfunctions

Before describing the operator-valued kernel formalism at the core of VSA, we outline the standard approach to separable decompositions of spatiotemporal data as in (1) via eigenfunctions of kernel integral operators associated with scalar-valued kernels. In this context, a kernel is a continuous bivariate function  $k : X \times X \rightarrow \mathbb{R}$ , which assigns a measure of correlation or similarity to pairs of dynamical states in  $X$ . Sometimes, but not always, we will require that  $k$  be symmetric, i.e.,  $k(x, x') = k(x', x)$  for all  $x, x' \in X$ . Two examples of popular kernels used in applications (both symmetric) are the covariance kernels employed in POD,

$$k(x, x') = \langle \vec{F}(x) - \bar{F}, \vec{F}(x') - \bar{F} \rangle_{H_Y}, \quad \bar{F} = \int_X \vec{F}(x) \, d\mu(x), \tag{2}$$

and radial Gaussian kernels,

$$k(x, x') = \exp\left(-\frac{\|\vec{F}(x) - \vec{F}(x')\|_{H_Y}^2}{\epsilon}\right), \quad \epsilon > 0, \tag{3}$$

which are frequently used in manifold learning applications. Note that in both of the above examples the dependence of  $k(x, x')$  on  $x$  and  $x'$  is through the values of  $\vec{F}$  at these points alone; this allows  $k(x, x')$  to be computable from observed data, without explicit knowledge of the underlying dynamical states  $x$  and  $x'$ . Hereafter, we will always work with such “data-driven” kernels.

Associated with every scalar-valued kernel is an integral operator  $K : H_X \rightarrow H_X$ , acting on  $f \in H_X$  according to the formula

$$Kf(x) = \int_X k(x, x')f(x') \, d\mu(x'). \tag{4}$$

If  $k$  is symmetric, then by compactness of  $A$  and continuity of  $k$ ,  $K$  is a compact, self-adjoint operator with an associated orthonormal basis  $\{\varphi_0, \varphi_1, \dots\}$  of  $H_X$  consisting of its eigenfunctions. Moreover, the eigenfunctions  $\varphi_j$  corresponding to nonzero eigenvalues are continuous. These eigenfunctions are employed as the chronos modes in (1), each inducing a continuous temporal pattern,  $t \mapsto \varphi_j(\Phi^t(x))$ , for every state  $x \in X$ . The spatial pattern  $\psi_j \in H_Y$  corresponding to  $\varphi_j$  is obtained by pointwise projection of the observation map onto  $\varphi_j$ , namely

$$\psi_j(y) = \langle \varphi_j, F_y \rangle_{H_X}, \quad (5)$$

where  $F_y \in H_X$  is the continuous scalar-valued function on  $X$  satisfying  $F_y(x) = \vec{F}(x)(y)$  for all  $x \in X$ .

### 2.3 Delay-Coordinate Maps and Koopman Operators

A potential shortcoming of spatiotemporal pattern extraction via the kernels in (2) and (3) is that the corresponding integral operators depend on the dynamics only indirectly, e.g., through the geometrical structure of the set  $\vec{F}(A) \subset H_Y$  on which the data is concentrated. Indeed, a well-known deficiency of POD, particularly in systems with symmetries, is failure to identify low-variance, yet dynamically important patterns (Aubry et al. 1993). As a way of addressing this issue, it has been found effective (Broomhead and King 1986; Vautard and Ghil 1989; Ghil et al. 2002; Giannakis and Majda 2012; Berry et al. 2013) to first embed the observed data in a higher-dimensional data space through the use of delay-coordinate maps, and then extract spatial and temporal patterns through a kernel operating in delay-coordinate space. For instance, analogs of the covariance and Gaussian kernels in (2) and (3) in delay-coordinate space are given by

$$k_Q(x, x') = \frac{1}{Q} \sum_{q=0}^{Q-1} \langle \vec{F}(\Phi^{-q\tau}(x)) - \bar{F}, \vec{F}(\Phi^{-q\tau}(x')) - \bar{F} \rangle_{H_Y}, \quad (6)$$

and

$$k_Q(x, x') = \exp \left( -\frac{1}{\epsilon Q} \sum_{q=0}^{Q-1} \|\vec{F}(\Phi^{-q\tau}(x)) - \vec{F}(\Phi^{-q\tau}(x'))\|_{H_Y}^2 \right), \quad (7)$$

respectively, here  $Q \in \mathbb{N}$  is the number of delays. The covariance kernel in (6) is essentially equivalent to the kernel employed in multi-channel SSA (Ghil et al. 2002) in an infinite-channel limit, and the Gaussian kernel in (7) is closely related to the kernel utilized in NLSA (though the NLSA kernel employs a state-dependent distance scaling akin to (19) ahead, as well as Markov normalization, and these features lead to certain technical advantages compared to unnormalized radial Gaussian kernels). See Appendix F for a description of NLSA.

As is well known (Packard et al. 1980; Takens 1981; Sauer et al. 1991; Robinson 2005; Deyle and Sugihara 2011), delay-coordinate maps can help recover the topological structure of state space from partial measurements of the system (i.e., non-injective observation maps), but in the context of kernel algorithms they also endow the kernels, and thus the corresponding eigenfunctions, with an explicit dependence on the dynamics. In Giannakis (2017) and Das and Giannakis (2019), it was established that as the number of delays  $Q$  grows, the integral operators  $K_Q$  associated with a family of scalar kernels  $k_Q$  operating in delay-coordinate space converge in operator norm, and thus in spectrum, to a compact kernel integral operator  $K_\infty$  on  $H_X$  commuting with the Koopman evolution operators (Budisić et al. 2012; Eisner et al. 2015) of the dynamical system. The latter are the unitary operators  $U^t : H_X \rightarrow H_X$ ,  $t \in \mathbb{R}$ , acting by composition with the flow map,

$$U^t f = f \circ \Phi^t,$$

thus governing the evolution of observables in  $H_X$  under the dynamics.

In the setting of measure-preserving ergodic systems, associated with  $U^t$  is a distinguished orthonormal set  $\{z_j\}$  of observables  $z_j \in H_X$  consisting of Koopman eigenfunctions (see Appendix A). These observables have the special property of exhibiting time-periodic evolution under the dynamics at a single frequency  $\alpha_j \in \mathbb{R}$  intrinsic to the dynamical system,

$$U^t z_j = e^{i\alpha_j t} z_j,$$

even if the underlying dynamical flow  $\Phi^t$  is aperiodic. Moreover, every Koopman eigenspace is one dimensional by ergodicity. Because commuting operators have common eigenspaces, and the eigenspaces of compact operators corresponding to nonzero eigenvalues are finite-dimensional, it follows that as  $Q$  increases, the eigenfunctions of  $K_Q$  at nonzero eigenvalues acquire increasingly coherent (periodic or quasiperiodic) time evolution associated with a finite number of Koopman eigenfrequencies  $\alpha_j$ . This property significantly enhances the physical interpretability and predictability of these patterns, providing justification for the skill of methods such as SSA and NLSA in extracting dynamically significant patterns from complex systems. Conversely, because kernel integral operators are generally more amenable to approximation from data than Koopman operators (which can have a highly complex spectral behavior), the operators  $K_Q$  provide an effective route for identifying finite-dimensional approximation spaces to stably and efficiently solve the Koopman eigenvalue problem.

## 2.4 Differences Between Covariance and Gaussian Kernels

Before closing this section, it is worthwhile pointing out two differences between covariance and Gaussian kernels, indicating that the latter may be preferable to the former in applications.

First, Gaussian kernels are strictly positive and bounded below on compact sets. That is, for every compact set  $S \subseteq X$  (including  $S = A$ ), there exists a constant  $c_S > 0$

such that  $k(x, x') \geq c_S$  for all  $x, x' \in S$ . This property allows Gaussian kernels to be normalizable to ergodic Markov diffusion kernels (Coifman and Lafon 2006; Berry and Sauer 2016). In a dynamical systems context, an important property of such kernels is that the corresponding integral operators always have an eigenspace at eigenvalue 1 containing constant functions, which turns out to be useful in establishing well-posedness of Galerkin approximation techniques for Koopman eigenfunctions (Das and Giannakis 2019). Markov diffusion operators are also useful for constructing spaces of observables of higher regularity than  $L^2$ , such as Sobolev spaces.

Second, if there exists a finite-dimensional linear subspace of  $H_Y$  containing the image of  $A$  under  $\vec{F}$ , then the integral operator  $K$  associated with the covariance kernel has necessarily finite rank (bounded above by the dimension of that subspace), even if  $\vec{F}$  is an injective map on  $A$ . This effectively limits the richness of observables that can be stably extracted from data-driven approximations of covariance eigenfunctions. In fact, it is a well-known property of covariance kernels that every eigenfunction  $\varphi_j$  at nonzero corresponding eigenvalue depends linearly on the observation map; specifically, up to proportionality constants,  $\varphi_j(x) = \langle \psi_j, \vec{F}(x) \rangle_{H_Y}$  with  $\psi_j$  given by (5), and the number of such patterns is clearly finite if  $\vec{F}(x)$  spans a finite-dimensional linear space as  $x$  is varied. On the other hand, apart from trivial cases, the kernel integral operators associated with Gaussian kernels have infinite rank (even if  $\vec{F}$  is non-injective), and if  $\vec{F}$  is injective they have no zero eigenvalues. In the latter case, data-driven approximations to the eigenfunctions of  $K$  provide an orthonormal basis for the full  $H_X$  space. Similar arguments also motivate the use of Gaussian kernels over polynomial kernels. In effect, by invoking the Taylor series expansion of the exponential function, a Gaussian kernel can be thought of as an “infinite-order” polynomial kernel.

### 3 Vector-Valued Spectral Analysis (VSA) Formalism

The main goal of VSA is to construct a decomposition of the observation map  $\vec{F}$  via an expansion of the form

$$\vec{F} \approx \sum_{j=0}^{l-1} \vec{F}_j, \quad \vec{F}_j = c_j \vec{\phi}_j, \quad (8)$$

where the  $c_j$  and  $\vec{\phi}_j$  are real-valued coefficients and vector-valued observables in  $H$ , respectively. Along a dynamical trajectory starting at  $x \in X$ , every such  $\vec{\phi}_j$  gives rise to a spatiotemporal pattern  $t \mapsto \vec{\phi}_j(\Phi^t(x))$ , generalizing the time series  $t \mapsto \varphi_j(\Phi^t(x))$  from Sect. 2.2. A key consideration in the VSA construction is that the recovered patterns should not necessarily be of the form  $\vec{\phi}_j = \varphi_j \otimes \psi_j$  for some  $\varphi_j \in H_X$  and  $\psi_j \in H_Y$ , as would be the case in the conventional decomposition in (1). To that end, we will determine the  $\vec{\phi}_j$  through the vector-valued eigenfunctions of an integral operator acting on  $H$  directly, as opposed to first identifying scalar-valued eigenfunctions in  $H_X$ , and then forming tensor products with the corresponding projection-based spatial patterns, as in Sect. 2.2. As will be described in detail below, the integral operator nominally employed by VSA is constructed using the theory of



operator-valued kernels (Micchelli and Pontil 2005; Caponnetto et al. 2008; Carmeli et al. 2010) for multitask machine learning, combined with delay-coordinate maps and Markov normalization as in NLSA.

### 3.1 Operator-Valued Kernel and Vector-Valued Eigenfunctions

Let  $B(H_Y)$  be the Banach space of bounded linear maps on  $H_Y$ , equipped with the operator norm. For our purposes, an operator-valued kernel is a continuous map  $l : X \times X \rightarrow B(H_Y)$ , mapping pairs of dynamical states in  $X$  to a bounded operator on  $H_Y$ . Every such kernel has an associated integral operator  $L : H \rightarrow H$ , acting on vector-valued observables according to the formula [cf. (4)]

$$L\vec{f}(x) = \int_X l(x, x')\vec{f}(x') \, d\mu(x'),$$

where the integral above is a Bochner integral (a vector-valued generalization of the Lebesgue integral). Note that operator-valued kernels and their corresponding integral operators can be viewed as generalizations of their scalar-valued counterparts from Sect. 2.2, in the sense that if  $Y$  only contains a single point, then  $H_Y$  is isomorphic to the vector space of complex numbers (equipped with the standard operations of addition and scalar multiplication and the inner product  $\langle w, z \rangle_{\mathbb{C}} = w^*z$ ), and  $B(H_Y)$  is isomorphic to the space of multiplication operators on  $\mathbb{C}$  by complex numbers. In that case, the action  $l(x, x')\vec{f}(x')$  of the linear map  $l(x, x') \in B(H_Y)$  on the function  $\vec{f}(x') \in H_Y$  becomes equivalent to multiplication of the complex number  $f(x)$ , where  $f$  is a complex-valued observable in  $H_X$ , by the value  $k(x, x') \in \mathbb{C}$  of a scalar-valued kernel  $k$  on  $X$ .

Consider now an operator-valued kernel  $l : X \times X \rightarrow B(H_Y)$ , such that for every pair  $(x, x')$  of states in  $X$ ,  $l(x, x') = L_{xx'}$  is a kernel integral operator on  $H_Y$  associated with a continuous kernel  $l_{xx'} : Y \times Y \rightarrow \mathbb{R}$  with the symmetry property

$$l_{xx'}(y, y') = l_{x'x}(y', y), \quad \forall x, x' \in X, \quad \forall y, y' \in Y. \tag{9}$$

This operator acts on a scalar-valued function  $g \in H_Y$  on the spatial domain via an integral formula analogous to (4), viz.

$$L_{xx'}g(y) = \int_Y l_{xx'}(y, y')g(y') \, d\nu(y').$$

Moreover, it follows from (9) that the corresponding operator  $L$  on vector-valued observables is self-adjoint and compact, and thus there exists an orthonormal basis  $\{\vec{\phi}_j\}$  of  $H$  consisting of its eigenfunctions,

$$L\vec{\phi}_j = \lambda_j\vec{\phi}_j, \quad \lambda_j \in \mathbb{R}.$$

Hereafter, we will always order the eigenvalues  $\lambda_j$  of integral operators in decreasing order starting at  $j = 0$ . By continuity of  $l$  and  $l_{xx'}$ , and compactness of  $A$  and  $Y$ , every eigenfunction  $\vec{\phi}_j$  at nonzero corresponding eigenvalue is a continuous function on  $X$ ,

taking values in the space of continuous functions on  $Y$ . Such eigenfunctions can be employed in the VSA decomposition in (8) with the expansion coefficients

$$c_j = \langle \vec{\phi}_j, \vec{F} \rangle_H = \int_X \langle \vec{\phi}_j(x), \vec{F}(x) \rangle_{H_Y} d\mu(x). \quad (10)$$

Note that, as with scalar kernel techniques, the decomposition in (8) does not include eigenfunctions at zero corresponding eigenvalue, for, to our knowledge, no data-driven approximation schemes are available for such eigenfunctions. See Sect. 5 and Appendix D for further details.

Because  $H$  is isomorphic as Hilbert space to the space  $H_\Omega$  of scalar-valued observables on the product space  $\Omega = X \times Y$  (see Sect. 2.1), every operator-valued kernel satisfying (9) can be constructed from a symmetric scalar kernel  $k : \Omega \times \Omega \rightarrow \mathbb{R}$  by defining  $l(x, x') = L_{xx'}$  as the integral operator associated with the kernel

$$l_{xx'}(y, y') = k(\omega, \omega'), \quad \omega = (x, y), \quad \omega' = (x', y'). \quad (11)$$

In particular, the vector-valued eigenfunctions of  $L$  are in one-to-one correspondence with the scalar-valued eigenfunctions of the integral operator  $K : H_\Omega \rightarrow H_\Omega$  associated with  $k$ , where

$$Kf(\omega) = \int_\Omega k(\omega, \omega') f(\omega') d\rho(\omega'). \quad (12)$$

That is, the eigenvalues and eigenvectors of  $K$  satisfy the equation  $K\phi_j = \lambda_j\phi_j$  for the same eigenvalues as those of  $L$ , and we also have

$$\vec{\phi}_j(x)(y) = \phi_j((x, y)), \quad \forall x \in X, \quad \forall y \in Y. \quad (13)$$

It is important to note that unless  $k$  is separable as a product of kernels on  $X$  and  $Y$ , i.e.,  $k((x, y), (x', y')) = k^{(X)}(x, x')k^{(Y)}(y, y')$  for some  $k^{(X)} : X \times X \rightarrow \mathbb{R}$  and  $k^{(Y)} : Y \times Y \rightarrow \mathbb{R}$ , the  $\vec{\phi}_j$  will not be of pure tensor product form,  $\vec{\phi}_j = \varphi_j \otimes \psi_j$  with  $\varphi_j \in H_X$  and  $\psi_j \in H_Y$ . Thus, passing to an operator-valued kernel formalism allows one to perform decompositions of significantly higher generality than the conventional approach in (1).

### 3.2 Operator-Valued Kernels with Delay-Coordinate Maps

While the framework described in Sect. 3.1 can be implemented with a broad range of kernels, VSA employs kernels leveraging the insights gained from SSA, NLSA, and related techniques on the use of kernels operating in delay-coordinate space. That is, analogously to the kernels employed by these methods that depend on the values  $\vec{F}((x)), \vec{F}(\Phi^{-\tau}(x)), \dots, \vec{F}(\Phi^{-(Q-1)\tau}(x))$  of the observation map on dynamical trajectories, VSA is based on kernels on the product space  $\Omega$  that also depend on data observed on dynamical trajectories, but with the key difference that this dependence is through the *local* values  $F_y(x), F_y(\Phi^{-\tau}(x)), \dots, F_y(\Phi^{-(Q-1)\tau}(x))$  of the observation map at each point  $y$  in the spatial domain  $Y$ . Specifically, defining the family of pointwise delay-embedding maps  $\vec{F}_Q : \Omega \rightarrow \mathbb{R}^Q$  with  $Q \in \mathbb{N}$  and

$$\tilde{F}_Q((x, y)) = \left( F_y(x), F_y(\Phi^{-\tau}(x)), \dots, F_y(\Phi^{-(Q-1)\tau}(x)) \right), \tag{14}$$

we require that the kernels  $k_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  utilized in VSA have the following properties:

1. For every  $Q \in \mathbb{N}$ ,  $k_Q$  is the pullback under  $\tilde{F}_Q$  of a continuous kernel  $\tilde{k}_Q : \mathbb{R}^Q \times \mathbb{R}^Q \rightarrow \mathbb{R}$ , i.e.,

$$k_Q(\omega, \omega') = \tilde{k}_Q(\tilde{F}_Q(\omega), \tilde{F}_Q(\omega')), \quad \forall \omega, \omega' \in \Omega. \tag{15}$$

2. The sequence of kernels  $k_1, k_2, \dots$  converges in  $H_\Omega \otimes H_\Omega$  norm to a kernel  $k_\infty \in H_\Omega \otimes H_\Omega$ .
3. The limit kernel  $k_\infty$  is invariant under the dynamics, in the sense that for all  $t \in \mathbb{R}$  and  $(\rho \times \rho)$ -a.e.  $(\omega, \omega') \in \Omega \times \Omega$ , where  $\omega = (x, y)$  and  $\omega' = (x', y')$ ,

$$k_\infty((\Phi^t(x), y), (\Phi^t(x'), y')) = k_\infty(\omega, \omega'). \tag{16}$$

We denote the corresponding integral operator on vector-valued observables in  $H$  corresponding to  $K_Q$ , determined through (11), by  $L_Q$ . As we will see below, operators of this class can be highly advantageous for the analysis of signals with an intermittent spatiotemporal character, as well as signals generated in the presence of dynamical symmetries. In addition, the family  $L_Q$  exhibits a commutativity with Koopman operators in the infinite-delay limit as in the case of SSA and NLSA.

Let  $\omega = (x, y)$  and  $\omega' = (x', y')$  with  $x, x' \in X$  and  $y, y' \in Y$  be arbitrary points in  $\Omega$ . As concrete examples of kernels satisfying the conditions listed above,

$$k_Q(\omega, \omega') = \frac{1}{Q} \sum_{q=0}^{Q-1} [F_y(\Phi^{-q\tau}(x)) - \bar{F}_y] [F_{y'}(\Phi^{-q\tau}(x')) - \bar{F}_{y'}],$$

$$\bar{F}_y = \int_X F_y(x) d\mu(x), \tag{17}$$

and

$$k_Q(\omega, \omega') = \exp \left( -\frac{1}{\epsilon Q} \sum_{q=0}^{Q-1} |F_y(\Phi^{-q\tau}(x)) - F_{y'}(\Phi^{-q\tau}(x'))|^2 \right), \quad \epsilon > 0, \tag{18}$$

are analogs of the covariance and Gaussian kernels in (2) and (3), respectively, defined on  $\Omega$ . For the reasons stated in Sect. 2.4, in practice we generally prefer working with Gaussian kernels than covariance kernels. Moreover, following the approach employed in NLSA and in Berry and Harlim (2016), and Giannakis (2017), we consider a more general class of Gaussian kernels than (18), namely

$$k_Q(\omega, \omega') = \exp \left( -\frac{a_Q(\omega)a_Q(\omega')}{\epsilon Q} \sum_{q=0}^{Q-1} |F_y(\Phi^{-q\tau}(x)) - F_{y'}(\Phi^{-q\tau}(x'))|^2 \right), \quad \epsilon > 0, \tag{19}$$

where  $a_Q : \Omega \rightarrow \mathbb{R}$  is a continuous nonnegative scaling function. Intuitively, the role of  $a_Q$  is to adjust the bandwidth (variance) of the Gaussian kernel in order to account for variations in the sampling density and time tendency of the data. The explicit construction of this function is described in Appendix C.1. For the purposes of the present discussion, it suffices to note that  $a_Q(\omega)$  can be evaluated given the values of  $F_y$  on the lagged trajectory  $\Phi^{-q\tau}(x)$ , so that, as with the covariance and radial Gaussian kernels, the class of kernels in (19) also satisfy (15). The existence of the limit  $k_\infty$  for this family of kernels, as well as the covariance kernels in (17), satisfying the conditions listed above is established in Appendix C.3.

### 3.3 Markov Normalization

As a final kernel construction step, when working with a strictly positive, symmetric kernel  $k_Q$ , such as (18) and (19), we normalize it to a continuous Markov kernel  $p_Q : \Omega \times \Omega \rightarrow \mathbb{R}$ , satisfying  $\int_\Omega p_Q(\omega, \cdot) d\rho = 1$  for all  $\omega \in \Omega$ , using the normalization procedure introduced in the diffusion maps algorithm (Coifman and Lafon 2006) and in Berry and Sauer (2016); see Appendix C.2 for a description. Due to this normalization, the corresponding integral operator  $P_Q : H_\Omega \rightarrow H_\Omega$  is an ergodic Markov operator having a simple eigenvalue  $\lambda_0 = 1$  and a corresponding constant eigenfunction  $\phi_0$ . Moreover, the range of  $P_Q$  is included in the space of continuous functions of  $\Omega$ . While this operator is not necessarily self-adjoint (since the kernel  $p_Q$  resulting from diffusion maps normalization is generally non-symmetric), it can be shown that it is related to a self-adjoint, compact operator by a similarity transformation. As a result, all eigenvalues of  $P_Q$  are real and admit the ordering  $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \cdots$ . Moreover, there exists a (non-orthogonal) basis  $\{\phi_0, \phi_1, \dots\}$  of  $H_\Omega$  consisting of eigenfunctions corresponding to these eigenvalues, as well as a dual basis  $\{\phi'_0, \phi'_1, \dots\}$  consisting of eigenfunctions of  $P_Q^*$  satisfying  $\langle \phi'_i, \phi_j \rangle_{H_\Omega} = \delta_{ij}$ . As with their unnormalized counterparts  $k_Q$ , the sequence of Markov kernels  $p_Q$  has a well-defined, shift-invariant limit  $p_\infty \in H_\Omega \otimes H_\Omega$  as  $Q \rightarrow \infty$ ; see Appendix C.2 for further details.

The eigenfunctions  $\phi_j$  induce vector-valued observables  $\vec{\phi}_j \in H$  through (13), which are in turn eigenfunctions of an integral operator  $\mathcal{P}_Q : H \rightarrow H$  associated with the operator-valued kernel determined via (11), applied to the Markov kernel  $p_Q$ . Similarly, the dual eigenfunctions  $\phi'_j$  induce vector-valued observables  $\vec{\phi}'_j \in H$ , which are eigenfunctions of  $\mathcal{P}_Q^*$  satisfying  $\langle \vec{\phi}'_i, \vec{\phi}_j \rangle_H = \delta_{ij}$ . Equipped with these observables, we perform the VSA decomposition in (8) with the expansion coefficients  $c_j = \langle \vec{\phi}'_j, \vec{F} \rangle_H$ . The latter expression can be viewed as a generalization of (10), applicable for non-orthonormal eigenbases.

## 4 Properties of the VSA Decomposition

In this section, we study the properties of the operators  $K_Q$  employed in VSA and their eigenfunctions in two relevant scenarios in spatiotemporal data analysis, namely data generated by systems with (i) dynamical symmetries, and (ii) nontrivial Koopman eigenfunctions. These topics will be discussed in Sects. 4.2 and 4.3, respectively.

We begin in Sect. 4.1 with some general observations on the topological structure of spatiotemporal data in delay-coordinate space, and the properties this structure imparts on the recovered eigenfunctions.

### 4.1 Bundle Structure of Spatiotemporal Data

In order to gain insight on the behavior of VSA, it is useful to consider the triplet  $(\Omega, B_Q, \pi_Q)$ , where  $B_Q = \tilde{F}_Q(\Omega)$  is the image of the product space  $\Omega$  under the delay-coordinate observation map, and  $\pi_Q : \Omega \rightarrow B_Q$  is the continuous surjective map defined as  $\pi_Q(\omega) = \tilde{F}_Q(\omega)$  for any  $\omega \in \Omega$ . Such a triplet forms a topological bundle with  $\Omega$ ,  $B_Q$ , and  $\pi_Q$  playing the role of the total space, base space, and projection map, respectively. In particular,  $\pi_Q$  partitions  $\Omega$  into equivalence classes

$$[\omega]_Q = \pi_Q^{-1}(x) \subseteq \Omega, \tag{20}$$

called fibers, on which  $\pi_Q(\omega)$  attains a fixed value (i.e.,  $\tilde{\omega}$  lies in  $[\omega]_Q$  if  $\pi_Q(\tilde{\omega}) = \pi_Q(\omega)$ ).

By virtue of (15), the kernel  $k_Q$  is a continuous function, constant on the  $[\cdot]_Q$  equivalence classes, i.e., for all  $\omega, \omega' \in \Omega$ ,  $\tilde{\omega} \in [\omega]_Q$ , and  $\tilde{\omega}' \in [\omega']_Q$ ,

$$k_Q(\omega, \omega') = k_Q(\tilde{\omega}, \tilde{\omega}'). \tag{21}$$

Therefore, since for any  $f \in H_\Omega$  and  $\tilde{\omega} \in [\omega]_Q$ ,

$$K_Q f(\omega) = \int_\Omega k_Q(\omega, \omega') f(\omega') \, d\rho(\omega') = \int_\Omega k_Q(\tilde{\omega}, \omega') f(\omega') \, d\rho(\omega') = K_Q f(\tilde{\omega}),$$

the range of the integral operator  $K_Q$  is a subspace of the continuous functions on  $\Omega$ , containing functions that are constant on the  $[\cdot]_Q$  equivalence classes. Correspondingly, the eigenfunctions  $\phi_j$  corresponding to nonzero eigenvalues (which lie in  $\text{ran } K_Q$ ) have the form  $\phi_j = \eta_j \circ \pi_Q$ , where  $\eta_j$  are continuous functions in the Hilbert space  $L^2(B_Q, \pi_{Q*}\rho)$  of scalar-valued functions on  $B_Q$ , square-integrable with respect to the pushforward of the measure  $\rho$  under  $\pi_Q$ . We can thus conclude that, if all eigenvalues  $\lambda_j$  with  $j \leq l - 1$  are nonzero, the VSA-reconstructed signal from (8) (viewed as a scalar-valued function on  $\Omega$ ), lies in the closed subspace  $\overline{\text{ran } K_Q} = \overline{\text{span}\{\phi_j : \lambda_j > 0\}}$  of  $H_\Omega$  spanned by functions that are constant on the  $[\cdot]_Q$  equivalence classes. Note that  $\overline{\text{ran } K_Q}$  is not necessarily decomposable as a tensor product of  $H_X$  and  $H_Y$  subspaces.

Observe now that with the definition of the kernel in (19), the  $[\cdot]_Q$  equivalence classes consist of pairs of dynamical states  $x \in \Omega$  and spatial points  $y \in Y$  for which the evolution of the observable  $F_y$  is identical over  $Q$  delays. While one can certainly envision scenarios where these equivalence classes each contain only one point, in a number of cases of interest, including the presence of dynamical symmetries examined below, the  $[\cdot]_Q$  equivalence classes will be nontrivial, and as a result  $\overline{\text{ran } K_Q}$  will be a strict subspace of  $H_\Omega$ . In such cases, the patterns recovered by VSA naturally

factor out data redundancies, which generally enhances both robustness and physical interpretability of the results. Besides spatiotemporal data, the bundle construction described above may be useful in other scenarios, e.g., analysis of data generated by dynamical systems with varying parameters (Yair et al. 2017).

## 4.2 Dynamical Symmetries

An important class of spatiotemporal systems exhibiting nontrivial  $[\cdot]_{\mathcal{Q}}$  equivalence classes is PDE models with equivariant dynamics under the action of symmetry groups on the spatial domain (Holmes et al. 1996). As a concrete example, we consider a PDE for a scalar field in  $H_Y$ , possessing a  $C^1$  inertial manifold, i.e., a finite-dimensional, forward-invariant submanifold of  $H_Y$  containing the attractor of the system, onto which every trajectory is exponentially attracted (Constantin et al. 1989). In this setting, the inertial manifold plays the role of the state space manifold  $X$ . Moreover, we assume that the full system state is observed, so that the observation map  $\vec{F}$  reduces to the inclusion  $X \hookrightarrow H_Y$ .

Consider now a topological group  $G$  (the symmetry group) with a continuous left action  $\Gamma_Y^g : Y \rightarrow Y$ ,  $g \in G$ , on the spatial domain, preserving null sets with respect to  $\nu$ . Suppose also that the dynamics is equivariant under the corresponding induced action  $\Gamma_X^g : X \rightarrow X$ ,  $\Gamma_X^g(x) = x \circ \Gamma_Y^{g^{-1}}$ , on the state space manifold. This means that the dynamical flow map and the symmetry group action commute,

$$\Gamma_X^g \circ \Phi^t = \Phi^t \circ \Gamma_X^g, \quad \forall t \in \mathbb{R}, \quad \forall g \in G, \quad (22)$$

or, in other words, if  $t \mapsto \Phi^t(x)$  is a solution starting at  $x \in X$ , then  $t \mapsto \Phi^t(\Gamma_X^g(x))$  is a solution starting at  $\Gamma_X^g(x)$ . Additional aspects of symmetry group actions and equivariance are outlined in Appendix B. Our goal for this section is to examine the implications of (22) to the properties of the operators  $K_{\mathcal{Q}}$  employed in VSA and their eigenfunctions.

### 4.2.1 Dynamical Symmetries and VSA Eigenfunctions

We begin by considering the induced action  $\Gamma_{\Omega}^g : \Omega \rightarrow \Omega$  of  $G$  on the product space  $\Omega$ , defined as

$$\Gamma_{\Omega}^g = \Gamma_X^g \otimes \Gamma_Y^g.$$

This group action partitions  $\Omega$  into orbits, defined for every  $\omega \in \Omega$  as the subsets  $\Gamma_{\Omega}(\omega) \subseteq \Omega$  with

$$\Gamma_{\Omega}(\omega) = \{\Gamma_{\Omega}^g(\omega) \mid g \in G\}.$$

As with the subsets  $[\omega]_{\mathcal{Q}} \subseteq \Omega$  from (20) associated with delay-coordinate maps, the  $G$ -orbits on  $\Omega$  form equivalence classes, consisting of points connected by symmetry group actions (as opposed to having common values under delay-coordinate maps).

In general, these two sets of equivalence classes are unrelated, but in the presence of dynamical symmetries, they are, in fact, compatible, as follows:

**Proposition 1** *If the equivariance property in (22) holds, then for every  $\omega \in \Omega$ , the  $G$ -orbit  $\Gamma_\Omega(\omega)$  is a subset of the  $[\omega]_Q$  equivalence class. As a result, the following diagram commutes:*

$$\begin{array}{ccc} \Omega & \xrightarrow{\Gamma_\Omega^g} & \Omega \\ & \downarrow \pi_Q \swarrow \pi_Q & \\ B_Q & & \end{array}$$

**Proof** Let  $x(y)$  denote the value of the dynamical state  $x \in X \subset H_Y$  at  $y \in Y$ . It follows from (22) that for every  $t \in \mathbb{R}$ ,  $g \in G$ ,  $x \in X$ , and  $y \in G$ ,

$$\Phi^t(\Gamma_X^g(x))(\Gamma_Y^g(y)) = \Gamma_X^g(\Phi^t(x))(\Gamma_Y^g(y)) = \Gamma_X^{g^{-1}}(\Gamma_X^g(\Phi^t(x)))(y) = \Phi^t(x)(y).$$

Therefore, since  $\vec{F}$  is an inclusion (i.e.,  $F_y(x) = x(y)$ ), setting  $\omega = (x, y) \in \Omega$ , we obtain

$$\begin{aligned} \pi_Q(\omega) &= \vec{F}_Q(\omega) = \left( F_y(x), F_y(\Phi^{-\tau}(x)), \dots, F_y(\Phi^{-(Q-1)\tau}(x)) \right) \\ &= \left( x(y), \Phi^{-\tau}(x)(y), \dots, \Phi^{-(Q-1)\tau}(x)(y) \right) \\ &= \left( \Gamma_X^g(x)(\Gamma_Y^g(y)), \Phi^{-\tau}(\Gamma_X^g(x))(\Gamma_Y^g(y)), \dots, \Phi^{-(Q-1)\tau}(\Gamma_X^g(x))(\Gamma_Y^g(y)) \right) \\ &= \vec{F}_Q(\Gamma_\Omega^g(\omega)) = \pi_Q(\Gamma_\Omega^g(\omega)). \end{aligned} \quad \square$$

We thus conclude from Proposition 1 and (21) that the kernel  $k_Q$  is constant on  $G$ -orbits,

$$k_Q(\Gamma_\Omega^g(\omega), \Gamma_\Omega^{g'}(\omega')) = k_Q(\omega, \omega'), \quad \forall \omega, \omega' \in \Omega, \quad \forall g, g' \in G, \quad (23)$$

and therefore the eigenfunctions  $\phi_j$  corresponding to nonzero eigenvalues of  $K_Q$  are continuous functions with the invariance property

$$\phi_j \circ \Gamma_\Omega^g = \phi_j, \quad \forall g \in G.$$

This is one of the key properties of VSA, which we interpret as factoring the symmetry group from the recovered spatiotemporal patterns.

### 4.2.2 Spectral Characterization

In order to be able to say more about the implication of the results in Sect. 4.2.1 at the level of operators, we now assume that the group action  $\Gamma_\Omega^g$  preserves the measure  $\rho$ . Then, there exists a unitary representation of  $G$  on  $H_\Omega$ , whose representatives are

unitary operators  $R_\Omega^g : H_\Omega \rightarrow H_\Omega$  acting on functions  $f \in H_\Omega$  by composition with  $\Gamma_\Omega^g$ , i.e.,  $R_\Omega^g f = f \circ \Gamma_\Omega^g$ . Another group of unitary operators acting on  $H_\Omega$  consists of the Koopman operators,  $\tilde{U}^t : H_\Omega \rightarrow H_\Omega$ , which we define here via a trivial lift of the Koopman operators  $U^t$  on  $H_X$ , namely  $\tilde{U}^t = U^t \otimes I_{H_Y}$ , where  $I_{H_Y}$  is the identity operator on  $H_Y$ ; see Appendix A for further details. In fact, the map  $t \mapsto \tilde{U}^t$  constitutes a unitary representation of the Abelian group of real numbers (playing the role of time), equipped with addition as the group operation, much like  $g \mapsto R_\Omega^g$  is a unitary representation of the symmetry group  $G$ . The following theorem summarizes the relationship between the symmetry group representatives and the Koopman and kernel integral operators on  $H_\Omega$ .

**Theorem 2** *For every  $g \in G$  and  $t \in \mathbb{R}$ , the operator  $R_\Omega^g$  commutes with  $K_Q$  and  $\tilde{U}^t$ . Moreover, every function in the range of  $K_Q$  is invariant under  $R_\Omega^g$ , i.e.,  $R_\Omega^g K_Q = K_Q$ .*

**Proof** The commutativity between  $R_\Omega^g$  and  $\tilde{U}^t$  is a direct consequence of (22). To verify the claims involving  $K_Q$ , we use (23) and the fact that  $\Gamma_\Omega^g$  preserves  $\rho$  to compute

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

where  $g$  and  $g'$  are arbitrary, and the equalities hold for  $\rho$ -a.e.  $\omega \in \Omega$ . Setting  $g' = g^{-1}$  in the above, and acting on both sides by  $R_\Omega^g$ , leads to  $R_\Omega^g K_Q = K_Q R_\Omega^g$ , i.e.,  $[R_\Omega^g, K_Q] = 0$ , as claimed. On the other hand, setting  $g$  to the identity element of  $G$  leads to  $K_Q = R_\Omega^{g'} K_Q$ , completing the proof of the theorem.  $\square$

Because commuting operators have common eigenspaces, Theorem 2 establishes the existence of two sets of common eigenspaces associated with the symmetry group, namely common eigenspaces between  $R_\Omega^g$  and  $K_Q$  and those between  $R_\Omega^g$  and  $\tilde{U}^t$ . In general, these two families of eigenspaces are not compatible since  $\tilde{U}^t$  and  $K_Q$  many not commute, so for now we will focus on the common eigenspaces between  $R_\Omega^g$  and  $K_Q$  which are accessible via VSA with finitely many delays. In particular, because  $R_\Omega^g K_Q = K_Q$ , and every eigenspace  $W_l$  of  $K_Q$  at nonzero corresponding eigenvalue  $\lambda_l$  is finite-dimensional (by compactness of that operator), we can conclude that the  $W_l$  are finite-dimensional subspaces onto which the action of  $R_\Omega^g$  reduces to the identity. In other words, the eigenspaces of  $K_Q$  at nonzero corresponding eigenvalues are finite-dimensional trivial representation spaces of  $G$ , and every VSA eigenfunction  $\phi_j$  is also an eigenfunction of  $R_\Omega^g$  at eigenvalue 1.



At this point, one might naturally ask to what extent these properties are shared in common between VSA and conventional eigendecomposition techniques based on scalar kernels on  $X$ . In particular, in the measure-preserving setting for the product measure  $\rho = \mu \times \nu$  examined above, it must necessarily be the case that the group actions  $\Gamma_X^g$  and  $\Gamma_Y^g$  separately preserve  $\mu$  and  $\nu$ , respectively, thus inducing unitary operators  $R_X^g : H_X \rightarrow H_X$  and  $R_Y^g : H_Y \rightarrow H_Y$ , defined analogously to  $R_\Omega^g$ . For a variety of kernels  $k_Q^{(X)} : X \times X \rightarrow \mathbb{R}$  that only depend on observed data through inner products and norms on  $H_Y$  (e.g., the covariance and Gaussian kernels in Sect. 2.2), the unitarity of  $R_X^g$  and  $R_Y^g$  implies that the invariance property

$$k_Q^{(X)}(\Gamma_X^g(x), \Gamma_X^g(x')) = k_Q^{(X)}(x, x') \tag{24}$$

holds for all  $g \in G$  and  $x, x' \in X$ . Moreover, proceeding analogously to the proof of Theorem 2, one can show that  $R_X^g$  and the integral operator  $K_Q^{(X)} : H_X \rightarrow H_X$  associated with  $k_Q^{(X)}$  commute, and thus have common eigenspaces  $W_l^{(X)}, \lambda_l^{(X)} \neq 0$ , which are finite-dimensional invariant subspaces under  $R_X^g$ . Projecting the observation map  $\vec{F}$  onto  $W_l^{(X)}$  as in (5), then yields a finite-dimensional subspace  $W_l^{(Y)} \subset H_Y$ , which is invariant under  $R_Y^g$ , and thus  $W_l^{(X)} \otimes W_l^{(Y)} \subset H_\Omega$  is invariant under  $R_\Omega^g$ . The fundamental difference between the representation of  $G$  on  $W_l^{(X)} \otimes W_l^{(Y)}$  and that on the  $W_l$  subspaces recovered by VSA, is that the former is generally *not* trivial, i.e., in general,  $R_\Omega^g$  does not reduce to the identity map on  $W_l^{(X)} \otimes W_l^{(Y)}$ . A well-known consequence of this is that the corresponding spatiotemporal patterns  $\varphi_j \otimes \psi_j$  from (1) become pure symmetry modes (e.g., Fourier modes in dynamical systems with translation invariance), hampering their physical interpretability.

This difference between VSA and conventional eigendecomposition techniques can be traced back to the fact that on  $X$  there is no analog of Proposition 1, relating equivalence classes of points with respect to delay-coordinate maps and group orbits on that space. Indeed, Proposition 1 plays an essential role in establishing the kernel invariance property in (23), which is stronger than (24) as it allows action by two independent group elements. Equation 23 is in turn necessary to determine that  $K_Q R_\Omega^g = K_Q$  in Theorem 2. In summary, these considerations highlight the importance of taking into account the bundle structure of spatiotemporal data when dealing with systems with dynamical symmetries.

### 4.3 Connection with Koopman Operators

#### 4.3.1 Behavior of Kernel Integral Operators in the Infinite-Delay Limit

As discussed in Sect. 4.2, in general, the kernel integral operators  $K_Q$  do not commute with the Koopman operators  $\tilde{U}^t$ , and thus these families of operators do not share common eigenspaces. Nevertheless, as we establish in this section, under the conditions on kernels stated in Sect. 3.2, the sequence of operators  $K_Q$  has an asymptotic

commutativity property with  $\tilde{U}^t$  as  $Q \rightarrow \infty$ , allowing the kernel integral operators from VSA to approximate eigenspaces of Koopman operators.

In order to place our results in context, we begin by noting that an immediate consequence of the bundle construction described in Sect. 4.1 is that if the support of the measure  $\rho$ , denoted  $M \subseteq \Omega$ , is connected as a topological space, then in the limit of no delays,  $Q = 1$ , the image of  $M$  under the delay-coordinate map  $\tilde{F}_1$  is a closed interval in  $\mathbb{R}$ , and correspondingly the eigenfunctions  $\phi_j$  are pullbacks of orthogonal functions on that interval under  $\tilde{F}_1$ . In particular, because  $\tilde{F}_1$  is equivalent to the vector-valued observation map, in the sense that  $\tilde{F}(x)(y) = \tilde{F}_1((x, y))$ , the eigenfunctions  $\phi_j$  of the  $Q = 1$  operator corresponding to nonzero eigenvalues are continuous functions, constant on the level sets of the input signal. Therefore, in this limit, the recovered eigenfunctions will generally have comparable complexity to the input data, and thus be of limited utility for the purpose of decomposing complex signals into simpler patterns. Nevertheless, besides the strict  $Q = 1$  limit, the  $\phi_j$  should remain approximately constant on the level sets of the input signal for moderately small values  $Q > 1$ , and this property should be useful in a number of applications, such as signal denoising and level set estimation [note that data-driven approximations to  $\phi_j$  become increasingly robust to noise with increasing  $Q$  (Giannakis 2017)]. Mathematically, in this small- $Q$  regime VSA has some common aspects with nonlocal averaging techniques in image processing (Buades et al. 2005).

We now focus on the behavior of VSA in the infinite-delay limit, where the following is found to hold.

**Theorem 3** *Under the conditions on the kernels  $k_Q$  stated in Sect. 3.2, the associated integral operators  $K_Q$  converge as  $Q \rightarrow \infty$  in operator norm, and thus in spectrum, to the integral operator  $K_\infty$  associated with the kernel  $k_\infty$ . Moreover,  $K_\infty$  commutes with the Koopman operator  $\tilde{U}^t$  for all  $t \in \mathbb{R}$ .*

**Proof** Since  $k_Q$  and  $k_\infty$  all lie in  $H_\Omega \otimes H_\Omega$ ,  $K_Q$  and  $K_\infty$  are Hilbert–Schmidt integral operators. As a result, the operator norm  $\|K_Q - K_\infty\|$  is bounded above by  $\|k_Q - k_\infty\|_{H_\Omega \otimes H_\Omega}$ , and the convergence of  $\|K_Q - K_\infty\|$  to zero follows from the fact that  $\lim_{Q \rightarrow \infty} \|k_Q - k_\infty\|_{H_\Omega \otimes H_\Omega} = 0$ , as stated in the conditions in Sect. 3.2. To verify that  $K_\infty$  and  $\tilde{U}^t$  commute, we proceed analogously to the proof of Theorem 2, using the shift invariance of  $k_\infty$  in (16) and the fact that  $\tilde{\Phi}^t = \Phi^t \otimes I_Y$  preserves the measure  $\rho$  to compute

$$\begin{aligned} K_\infty f(\omega) &= \int_{\Omega} k_\infty(\omega, \omega') f(\omega') \, d\rho(\omega') \\ &= \int_{\Omega} k_\infty(\omega, \tilde{\Phi}^t(\omega')) f(\tilde{\Phi}^t(\omega')) \, d\rho(\omega') \\ &= \int_{\Omega} k_\infty(\tilde{\Phi}^{-t}(\omega), \omega') f(\tilde{\Phi}^t(\omega')) \, d\rho(\omega') \\ &= \tilde{U}^{-t} K_\infty \tilde{U}^t f(\omega), \end{aligned}$$

where the equalities hold for  $\rho$ -a.e.  $\omega \in \Omega$ . Pre-multiplying these expressions by  $\tilde{U}^t$  leads to

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

as claimed.  $\square$

Theorem 3 generalizes the results in Giannakis (2017) and Das and Giannakis (2019), where analogous commutativity properties between Koopman and kernel integral operators were established for scalar-valued observables in  $H_X$ . By virtue of the commutativity between  $K_\infty$  and  $\tilde{U}^t$ , at large numbers of delays  $Q$ , VSA decomposes the signal into patterns with a coherent temporal evolution associated with intrinsic frequencies of the dynamical system. In particular, being a compact operator,  $K_\infty$  has finite-dimensional eigenspaces,  $W_l$ , corresponding to nonzero eigenvalues, whereas the eigenspaces of  $\tilde{U}^t$  are infinite-dimensional, yet are spanned by eigenfunctions with a highly coherent (periodic) time evolution at the corresponding eigenfrequencies  $\alpha_j \in \mathbb{R}$ ,

$$\tilde{U}^t \tilde{z}_j = e^{i\alpha_j t} \tilde{z}_j, \quad \tilde{z}_j \in H_\Omega;$$

see Appendix A.1 for further details. The commutativity between  $K_\infty$  and  $\tilde{U}^t$  allows us to identify finite-dimensional subspaces  $W_l$  of  $H_\Omega$  containing distinguished observables which are simultaneous eigenfunctions of  $K_\infty$  and  $\tilde{U}^t$ . As shown in Appendix A.2, these eigenfunctions have the form

$$\tilde{z}_{jl} = z_{jl} \otimes \psi_{jl}, \quad j \in \{1, \dots, \dim W_l\}, \quad (25)$$

where  $z_{jl}$  is an eigenfunction of the Koopman operator  $U^t$  on  $H_X$  at eigenfrequency  $\alpha_{jl}$ , and  $\psi_{jl}$  a spatial pattern in  $H_Y$ . Note that here we use a two-index notation,  $z_{jl}$  and  $\alpha_{jl}$ , for Koopman eigenvalues and eigenfrequencies, respectively, to indicate the fact that they are associated with the  $W_l$  eigenspace of  $K_\infty$ . We therefore deduce from (25) that in the infinite-delay limit, the spatiotemporal patterns recovered by VSA can be factored into a separable, tensor product form similar to the conventional decomposition in (1) based on scalar kernel algorithms. It is important to note, however, that unlike (1), the spatial patterns  $\psi_{jl}$  in (25) are not necessarily given by linear projections of the observation map onto the corresponding scalar Koopman eigenfunctions  $z_{jl} \in H_X$  [called Koopman modes in the Koopman operator literature (Mezić 2005)]. In effect, taking into account the intrinsic structure of spatiotemporal data as vector-valued observables allows VSA to recover more general spatial patterns than those associated with linear projections of observed data.

Another consideration to keep in mind (which applies for many techniques utilizing delay-coordinate maps besides VSA) is that  $K_\infty$  can only recover patterns in a subspace  $\mathcal{D}_\Omega$  of  $H_\Omega$  associated with the point spectrum of the dynamical system generating the data (i.e., the Koopman eigenfrequencies; see Appendix A.1). Dynamical systems of sufficient complexity will exhibit a nontrivial subspace  $\mathcal{D}_\Omega^\perp$  associated

with the continuous spectrum, which does not admit a basis associated with Koopman eigenfunctions. One can show via analogous arguments to Das and Giannakis (2019) that  $\mathcal{D}_\Omega^\perp$  is, in fact, contained in the nullspace of  $K_\infty$ , which is a potentially infinite-dimensional space not accessible from data. Of course, in practice, one always works with finitely many delays  $Q$ , which in principle allows recovery of patterns in  $\mathcal{D}_\Omega^\perp$  through eigenfunctions of  $K_Q$ , and these patterns will not have an asymptotically separable behavior as  $Q \rightarrow \infty$  analogous to (25).

In light of the above, we can conclude that increasing  $Q$  from small values will impart changes to the topology of the base space  $B_Q$ , and in particular the image of the support  $M$  of  $\rho$  under  $\pi_Q$ , but also the spectral properties of the operators  $K_Q$ . On the basis of classical delay-embedding theorems (Sauer et al. 1991), one would expect the topology of  $\pi_Q(M)$  to eventually stabilize, in the sense that for every spatial point  $y \in Y$  the set  $A_y = A \times \{y\} \subseteq M$  will map homeomorphically under  $\pi_Q$  for  $Q$  greater than a finite number (that is, topologically,  $\pi_Q(A_y)$  will be a “copy” of  $A$ ). However, apart from special cases,  $K_Q$  will continue changing all the way to the asymptotic limit  $Q \rightarrow \infty$  where Theorem 3 holds.

Before closing this section, we also note that while VSA does not directly provide estimates of Koopman eigenfrequencies, such estimates could be computed through Galerkin approximation techniques utilizing the eigenspaces of  $K_Q$  at large  $Q$  as trial and test spaces, as done elsewhere (Giannakis et al. 2015; Giannakis 2017; Das and Giannakis 2019) for scalar-valued Koopman eigenfunctions. A study of such techniques in the context of vector-valued Koopman eigenfunctions (equivalently, eigenfunctions in  $H_\Omega$ ) is beyond the scope of this work, though it is expected that their well-posedness and convergence properties should follow from fairly straightforward modification of the approach in the references cited above.

### 4.3.2 Infinitely Many Delays with Dynamical Symmetries

As a final asymptotic limit of interest, we consider the limit  $Q \rightarrow \infty$  under the assumption that a symmetry group  $G$  acts on  $H_\Omega$  via unitary operators  $R_\Omega^g$ , as described in Sect. 4.2. In that case, the commutation relations

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

imply that there exist finite-dimensional subspaces of  $H_\Omega$  spanned by simultaneous eigenfunctions of  $R_\Omega^g$ ,  $\tilde{U}^t$ , and  $K_\infty$ . We know from (25) that these eigenfunctions,  $\tilde{z}_{jl}$ , are given by a tensor product between a Koopman eigenfunction  $z_{jl} \in H_X$  and a spatial pattern  $\psi_{jl} \in H_Y$ . It can further be shown (see Appendix B.2) that  $z_{jl}$  and  $\psi_{jl}$  are eigenfunctions of the unitary operators  $R_X^g$  and  $R_Y^g$ , i.e.,

$$R_X^g z_{jl} = \gamma_{X,jl}^g z_{jl}, \quad R_Y^g \psi_{jl} = \gamma_{Y,jl}^g \psi_{jl}, \quad |\gamma_{X,jl}^g| = |\gamma_{Y,jl}^g| = 1,$$

and moreover the eigenvalues  $\gamma_{X,jl}^g$  and  $\gamma_{Y,jl}^g$  satisfy  $\gamma_{X,jl}^g \gamma_{Y,jl}^g = 1$ . In particular, we have  $R_\Omega^g = R_X^g \otimes R_Y^g$ , and the quantity  $\gamma_{\Omega,jl}^g = \gamma_{X,jl}^g \gamma_{Y,jl}^g$  is equal to the eigenvalue of  $R_\Omega^g$  corresponding to  $\tilde{z}_{jl}$ , which is equal to 1 by Theorem 2.

In summary, every simultaneous eigenfunction  $\tilde{z}_{jl}$  of  $K_\infty$ ,  $\tilde{U}^t$ , and  $R_\Omega^g$  is characterized by three eigenvalues, namely (i) a kernel eigenvalue  $\lambda_l$  associated with  $K_\infty$ ; (ii) a Koopman eigenfrequency  $\alpha_{jl}$  associated with  $\tilde{U}^t$ ; and (iii) a spatial symmetry eigenvalue  $\gamma_{Y,jl}^g$  (which can be thought of as a “wavenumber” on  $Y$ ).

## 5 Data-Driven Approximation

In this section, we consider the problem of approximating the eigenvalues and eigenfunctions of the kernel integral operators employed in VSA from a finite dataset consisting of time-ordered measurements of the vector-valued observable  $\vec{F}$ . Specifically, we assume that available to us are measurements  $\vec{F}(x_0), \vec{F}(x_1), \dots, \vec{F}(x_{N-1})$  taken along an (unknown) orbit  $x_n = \Phi^{n\tau}(x_0)$  of the dynamics at the sampling interval  $\tau$ , starting from an initial state  $x_0 \in X$ . We also consider that each scalar field  $\vec{F}(x_n) \in H_Y$  is sampled at a finite collection of distinct points  $y_0, y_1, \dots, y_{S-1}$  in  $Y$ . We will exclude the trivial case that the support  $A$  of the invariant measure  $\mu$  is a fixed point by assumption. Given such data, and without assuming knowledge of the underlying dynamical flow and/or state space geometry, our goal is to construct a family of operators, whose eigenvalues and eigenfunctions converge, in a suitable sense, to those of  $K_Q$ , in an asymptotic limit of large data,  $N, S \rightarrow \infty$ . In essence, we seek to address a problem on spectral approximation of kernel integral operators from an unstructured grid of points  $(x_n, y_s)$  in  $\Omega$ .

### 5.1 Data-Driven Hilbert Spaces and Kernel Integral Operators

An immediate consequence of the fact that the dynamics is unknown is that the invariant measure  $\mu$  defining the Hilbert space  $H_X = L^2(X, \mu)$  is also unknown (arguably, apart from special cases,  $\mu$  would be difficult to explicitly determine even if  $\Phi^t$  were known). This means that instead of  $H_X$  we only have access to a finite-dimensional Hilbert space  $H_{X,N} = L^2(X, \mu_N)$  associated with the sampling measure  $\mu_N = \sum_{n=0}^{N-1} \delta_{x_n}/N$  on the trajectory  $X_N = \{x_0, \dots, x_{N-1}\}$ , where  $\delta_{x_n}$  is the Dirac probability measure supported at  $x_n \in X$ . Since  $\mu$  is not supported at a fixed point, it follows by ergodicity of  $\Phi^\tau$  and continuity of  $t \mapsto \Phi^t$  that all points in  $X_N$  are distinct for  $\mu$ -a.e. starting state  $x_0$ . The analysis that follows will thus only treat the case of distinct sampled states  $x_n$ . In that case,  $H_{X,N}$  consists of equivalence classes of functions on  $X$  having common values on the finite set  $X_N \subset X$ , and is equipped with the inner product  $\langle f, g \rangle_{H_{X,N}} = \sum_{n=0}^{N-1} f^*(x_n)g(x_n)/N$ . Because every such equivalence class  $f$  is uniquely characterized by  $N$  complex numbers,  $f(x_0), \dots, f(x_{N-1})$ , corresponding to the values of one of its representatives on  $X_N$ ,  $H_{X,N}$  is isomorphic to  $\mathbb{C}^N$ , equipped with a normalized Euclidean inner product. Thus, we can represent every  $f \in H_{X,N}$  by an  $N$ -dimensional column vector  $\underline{f} = (f(x_0), \dots, f(x_{N-1}))^\top \in \mathbb{C}^N$ , and every linear operator  $T : H_{X,N} \rightarrow H_{X,N}$  by an  $N \times N$  matrix  $\mathbf{T}$  such that  $\mathbf{T}\underline{f}$  is equal to the column-vector representation of  $Tf$ . In particular, associated with every scalar kernel  $k : X \times X \rightarrow \mathbb{R}$  is a kernel integral operator  $K_N : H_{X,N} \rightarrow H_{X,N}$ , acting on  $f \in H_{X,N}$  according to the formula (cf. (4))

$$K_N f(x_m) = \int_X k(x_m, x_n) d\mu_N(x_n) = \frac{1}{N} \sum_{n=0}^{N-1} k(x_m, x_n) f(x_n). \quad (26)$$

This operator is represented by an  $N \times N$  kernel matrix  $\mathbf{K} = [k(x_m, x_n)/N]$ .

In the setting of spatiotemporal data analysis, one has to also take into account the finite sampling of the spatial domain, replacing  $H_Y = L^2(Y, \nu)$  by the  $S$ -dimensional Hilbert space  $H_{Y,S} = L^2(Y, \nu_S)$  associated with a discrete measure  $\nu_S = \sum_{s=0}^{S-1} \beta_{s,S} \delta_{y_s}$ . Here, the  $\beta_{s,S}$  are positive quadrature weights such that given any continuous function  $f : Y \rightarrow \mathbb{C}$ , the quantity  $\sum_{s=0}^{S-1} \beta_{s,S} f(y_s)$  approximates  $\int_Y f d\nu$ . For instance, if  $\nu$  is a probability measure, and the sampling points  $y_s$  are equidistributed with respect to  $\nu$ , a natural choice is uniform weights,  $\beta_{s,S} = 1/S$ . The space  $H_{Y,S}$  is constructed analogously to  $H_{X,N}$ , and similarly we replace  $H_\Omega = L^2(\Omega, \rho)$  by the  $NS$ -dimensional Hilbert space  $H_{\Omega,NS} = L^2(\Omega, \rho_{NS})$ , where  $\rho_{NS} = \mu_N \times \nu_S = \sum_{n=0}^{N-1} \sum_{s=0}^{S-1} \beta_{s,S} \delta_{\omega_{ns}}/N$  and  $\omega_{ns} = (x_n, y_s)$ . As a Hilbert space,  $H_{\Omega,NS}$  is isomorphic to the space  $H_{NS} = L^2(X, \mu_N; H_{Y,N})$  of vector-valued observables, which is the data-driven analog of  $H$ , as well as the tensor product space  $H_{X,N} \otimes H_{Y,S}$  (cf. Sect. 2.1). Given a kernel  $k_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  satisfying the conditions in Sect. 3.2, there is an associated integral operator  $K_{Q,NS} : H_{\Omega,NS} \rightarrow H_{\Omega,NS}$ , defined analogously to (26) by

$$K_{Q,NS} f(\omega_{mr}) = \int_\Omega k_Q(\omega_{mr}, \omega_{ns}) d\rho_{NS}(\omega_{ns}), \quad (27)$$

and represented by the  $(NS) \times (NS)$  matrix  $\mathbf{K}$  with elements

$$K_{mr,ns} = k_Q(\omega_{mr}, \omega_{ns})/(NS).$$

Solving the eigenvalue problem for  $K_{Q,NS}$  (which is equivalent to the matrix eigenvalue problem for  $\mathbf{K}$ ) leads to eigenvalues  $\lambda_{NS,j} \in \mathbb{R}$  and eigenfunctions  $\phi_{NS,j} \in H_{\Omega,NS}$ , the latter, represented by column vectors  $\underline{\phi}_j \in \mathbb{R}^{NS}$  with elements equal to  $\phi_{NS,j}(\omega_{ns})$ . We consider  $\lambda_{NS,j}$  and  $\phi_{NS,j}$  as data-driven approximations to the eigenvalues and eigenfunctions  $\lambda_j$  and  $\phi_j$ , respectively, of the integral operator in (12) associated with the kernel  $k_Q$ . The convergence properties of this approximation will be made precise in Sect. 5.2.

A similar data-driven approximation can be performed for operators based on the Markov kernels  $p_Q$  from Sect. 3.3, which is our preferred class of kernels for VSA. However, in this case the kernels  $p_{Q,NS} : \Omega \times \Omega \rightarrow \mathbb{R}$  associated with the approximating operators  $P_{Q,NS}$  on  $H_{\Omega,NS}$  are Markov-normalized with respect to the measure  $\rho_{NS}$ , i.e.,  $\int_\Omega p_{Q,NS}(\omega, \cdot) d\rho_{NS} = 1$ , so they acquire a dependence on  $N$  and  $S$ . As with the eigenvalues of  $P_Q$ , the eigenvalues  $\lambda_{NS,j}$  of  $P_{Q,NS}$  are real, and admit the ordering  $1 = \lambda_{NS,0} > \lambda_{NS,1} \geq \lambda_{NS,2} \geq \dots \geq \lambda_{NS,NS-1}$ . Moreover, there exists a basis of  $H_{\Omega,NS}$  consisting of corresponding eigenvectors,  $\phi_{NS,j}$ , as well as a dual basis with elements  $\phi'_{NS,j}$  such that  $\langle \phi'_{NS,i}, \phi_{NS,j} \rangle_{H_{\Omega,NS}} = \delta_{ij}$ . Further details on this construction and its convergence properties can be found in Appendix D.

The  $\phi_{NS,j}$  and  $\phi'_{NS,j}$  have associated vector-valued functions  $\vec{\phi}_{NS,j}$  and  $\vec{\phi}'_{NS,j}$ , respectively, in  $H_{NS}$ , which we employ to perform a decomposition of the observation map  $\vec{F}$  analogous to (8), viz.

$$\vec{F} \approx \sum_{j=0}^{l-1} \vec{F}_{NS,j}, \quad \vec{F}_{NS,j} = c_{NS,j} \vec{\phi}_{NS,j}, \quad c_{NS,j} = \langle \vec{\phi}'_{NS,j}, \vec{F} \rangle_{H_{NS}}. \quad (28)$$

Here, the reconstructed signal  $\sum_{j=0}^{l-1} \vec{F}_{NS,j}$  converges to  $\vec{F}$  in the limit of  $l = NS - 1$  in  $H_{NS}$  norm; this is equivalent to pointwise convergence on the sampled dynamical states  $x_n \in X$  and spatial points in  $y_s \in Y$ . Moreover, as we will see in Sect. 5.2, if all eigenvalues  $\lambda_{NS,j}$  with  $j \leq l - 1$  are nonzero,  $\vec{F}_{NS,l}$  has a continuous representative, which can be evaluated at arbitrary  $x \in X$  and  $y \in Y$ . A pseudocode implementation of the full VSA pipeline for the class of Markov kernels  $p_{Q,NS}$  is included in Appendix E.

### 5.2 Spectral Convergence

For a spectrally consistent data-driven approximation scheme, we would like to be able to establish that, as  $N$  and  $S$  increase, the sequence of eigenvalues  $\lambda_{NS,j}$  of  $K_{NS}$  converges to eigenvalue  $\lambda_j$  of  $K$ , and for an eigenfunction  $\phi_j$  of  $K$  corresponding to  $\lambda_j$  there exists a sequence of eigenfunctions  $\phi_{NS,j}$  of  $K_{NS}$  converging to it. While convergence of eigenvalues can be unambiguously understood in terms of convergence of real numbers, in the setting of interest here a suitable notion of convergence of eigenfunctions (or, more generally, eigenspaces) is not obvious, since  $\phi_{NS,j}$  and  $\phi_j$  lie in fundamentally different spaces. That is, there is no natural way of mapping equivalence classes of functions with respect to  $\rho_{NS}$  (i.e., elements of  $H_{\Omega,NS}$ ) to equivalence classes of functions with respect to  $\rho$  (i.e., elements of  $H_{\Omega}$ ), allowing one, e.g., to establish convergence of eigenfunctions in  $H_{\Omega}$  norm. This issue is further complicated by the fact, that in many cases of interest, the support  $A$  of the invariant measure  $\mu$  is a non-smooth subset of  $X$  of zero Lebesgue measure (e.g., a fractal attractor), and the sampled states  $x_n$  do not lie exactly on  $A$  (as that would require starting states  $x_0$  drawn from a measure zero subset of  $X$ , which is not feasible experimentally). In fact, the issues outlined above are common to many other data-driven techniques for analysis of dynamical systems besides VSA (e.g., POD and DMD), yet are oftentimes not explicitly addressed in the literature.

Here, following Das and Giannakis (2019), we take advantage of the fact that, by the assumed continuity of VSA kernels, every kernel integral operator  $K_Q : H_{\Omega} \rightarrow H_{\Omega}$  from Sect. 3.2 can be also be viewed as an integral operator on the space  $C(\mathcal{V})$  of continuous functions on any compact subset  $\mathcal{V} \subset \Omega$  containing the support of  $\rho$ . This integral operator, denoted by  $\tilde{K}_Q : C(\mathcal{V}) \rightarrow C(\mathcal{V})$ , acts on continuous functions through the same integral formula as (12), although the domains and codomains of  $K_Q$  and  $\tilde{K}_Q$  are different. It is straightforward to verify that every eigenfunction  $\phi_j \in H_{\Omega}$  of  $K_Q$  at nonzero eigenvalue  $\lambda_j$  has a unique continuous representative  $\tilde{\phi}_j \in C(\mathcal{V})$ , given by

$$\tilde{\phi}_j(\omega) = \frac{1}{\lambda_j} \int_{\Omega} k_Q(\omega, \omega') \phi_j(\omega') \, d\rho(\omega'), \quad (29)$$

and  $\tilde{\phi}_j$  is an eigenfunction of  $\tilde{K}_Q$  at the same eigenvalue  $\lambda_j$ . Assuming further that  $\mathcal{V}$  also contains the supports of the measures  $\rho_{NS}$  for all  $N, S \geq 1$ , we can define  $\tilde{K}_{Q,NS} : C(\mathcal{V}) \rightarrow C(\mathcal{V})$  analogously to (27). Then, every eigenfunction  $\phi_{NS,j} \in H_{\Omega,NS}$  of  $K_{Q,NS}$  at nonzero corresponding eigenvalue  $\lambda_{NS,j}$  has a continuous representative  $\tilde{\phi}_{NS,j}$ , with

$$\tilde{\phi}_{NS,j}(\omega) = \frac{1}{\lambda_j} \int_{\Omega} k_Q(\omega, \omega') \phi_{NS,j}(\omega') \, d\rho_{NS}(\omega'), \quad (30)$$

which is an eigenfunction of  $\tilde{K}_{Q,NS}$  at the same eigenvalue  $\lambda_{NS,j}$ .

As is well known, the space  $C(\mathcal{V})$  equipped with the uniform norm  $\|f\|_{C(\mathcal{V})} = \max_{\omega \in \mathcal{V}} |f(\omega)|$  becomes a Banach space, and it can further be shown that  $\tilde{K}_{Q,NS}$  and  $\tilde{K}_Q$  are compact operators on this space. In other words,  $C(\mathcal{V})$  can be used as a universal space to establish spectral convergence of  $\tilde{K}_{Q,NS}$  to  $\tilde{K}_Q$ , using approximation techniques for compact operators on Banach spaces (Chatelin 2011). Von Luxburg et al. (2008) use this approximation framework to establish convergence results for spectral clustering techniques, and their approach can naturally be adapted to show that, under natural assumptions,  $\tilde{K}_{Q,NS}$  indeed converges in spectrum to  $\tilde{K}_Q$ . In Appendix D, we prove the following result:

**Theorem 4** *Suppose that  $\mathcal{V} \subseteq \Omega$  is a compact set containing the supports of  $\rho$  and the family of measures  $\rho_{NS}$ , and assume that  $\rho_{NS}$  converges weakly to  $\rho$ , in the sense that*

$$\lim_{N,S \rightarrow \infty} \int_{\Omega} f \, d\rho_{NS} = \int_{\Omega} f \, d\rho, \quad \forall f \in C(\Omega). \quad (31)$$

*Then, for every nonzero eigenvalue  $\lambda_j$  of  $K_Q$ , including multiplicities, there exist positive integers  $N_0, S_0$  such that the eigenvalues  $\lambda_{NS,j}$  of  $K_{Q,NS}$  with  $N \geq N_0$  and  $S \geq S_0$  converge, as  $N, S \rightarrow \infty$ , to  $\lambda_j$ . Moreover, for every eigenfunction  $\phi_j \in H_{\Omega}$  of  $K$  corresponding to  $\lambda_j$ , there exist eigenfunctions  $\phi_{NS,j}$  of  $K_{Q,NS}$  corresponding to  $\lambda_{NS,j}$ , whose continuous representatives  $\tilde{\phi}_{NS,j}$  from (30) converge uniformly on  $\mathcal{V}$  to  $\tilde{\phi}_j$  from (29). Moreover, analogous results hold for the eigenvalues and eigenfunctions of the Markov operators  $P_{Q,NS}$  and  $P_Q$ .*

A natural setting where the conditions stated in Theorem 4 are satisfied are dynamical systems with compact absorbing sets and associated physical measures. Specifically, for such systems we shall assume that there exists a Lebesgue measurable subset  $\mathcal{U}$  of the state space manifold  $X$ , such that (i)  $\mathcal{U}$  is forward-invariant, i.e.,  $\Phi^t(\mathcal{U}) \subseteq \mathcal{U}$  for all  $t \geq 0$ ; (ii) the topological closure  $\bar{\mathcal{U}}$  is a compact set containing the support of  $\mu$ ; (iii)  $\mathcal{U}$  has positive Lebesgue measure in  $X$ ; and (iv) for any starting state  $x_0 \in \mathcal{U}$ , the corresponding sampling measures  $\mu_N$  converge weakly to  $\mu$ , i.e.,  $\lim_{N \rightarrow \infty} \int_X f \, d\mu_N = \int_X f \, d\mu$  for all  $f \in C(X)$ . Invariant measures exhibiting



Properties (iii) and (iv) are known as physical measures (Young 2002); in such cases, the set  $\mathcal{U}$  is called a basin of  $\mu$ . Clearly, Properties (i)–(iv) are satisfied if  $\Phi^t : X \rightarrow X$  is a flow on a compact manifold with an ergodic invariant measure supported on the whole of  $X$ , but are also satisfied in more general settings, such as certain dissipative flows on noncompact manifolds [e.g., the Lorenz 63 system on  $X = \mathbb{R}^3$  (Lorenz 1963)]. Assuming further that the measures  $\nu_S$  associated with the sampling points  $y_0, \dots, y_{S-1}$  and the corresponding quadrature weights  $\beta_{0,S}, \dots, \beta_{S-1,S}$  on the spatial domain  $Y$  converge weakly to  $\nu$ , i.e.,  $\lim_{S \rightarrow \infty} \int_Y g \, d\nu_S = \int_Y g \, d\nu$  for every  $g \in C(Y)$ , the conditions in Theorem 7 are met with  $\mathcal{V} = \overline{\mathcal{U}} \times Y$ , and the measures  $\rho_{NS}$  constructed as described in Sect. 5.1 for any starting state  $x_0 \in \mathcal{U}$ . Under these conditions, the data-driven spatiotemporal patterns  $\phi_{NS,j}$  recovered by VSA converge for an experimentally accessible set of initial states in  $X$ .

## 6 Application to the Kuramoto–Sivashinsky Model

### 6.1 Overview of the Kuramoto Sivashinsky Model

The KS model, originally introduced as a model for wave propagation in a dissipative medium (Kuramoto and Tsuzuki 1976), or laminar flame propagation (Sivashinsky 1977), is one of the most widely studied dissipative PDE models 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  $u(t, \cdot) : Y \rightarrow \mathbb{R}$ ,  $t \geq 0$  is given by

$$\dot{u} = -u\nabla u + \Delta u - \Delta^2 u, \quad (32)$$

where  $\nabla$  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 for  $L \gtrsim 4 \times 2\pi$  (Greene and Kim 1988; Arbruster et al. 1989; Kevrekidis et al. 1990; Cvitanović et al. 2009; Takeuchi et al. 2011).

A fundamental property of the KS system is that it possesses a global compact attractor, embedded within a finite-dimensional inertial manifold of class  $C^r$ ,  $r \geq 1$  (Foias et al. 1986, 1988; Constantin et al. 1989; Jolly et al. 1990; Chow et al. 1992; Robinson 1994). That is, there exists a  $C^r$  submanifold  $\mathcal{X}$  of the Hilbert space  $H_Y = L^2(Y, \nu)$  with  $\nu$  set to the Lebesgue measure, which is 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. Dimension estimates of inertial manifolds (Robinson 1994; Jolly et al. 2000) and attractors (Tajima and Greenside 2002) 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. 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,  $\mu$ . On  $X$ , the dynamics is described by a  $C^r$  flow map  $\Phi^t : X \rightarrow X$ ,  $t \in \mathbb{R}$ , as in Sect. 2.1. In particular, for every initial condition  $x_0 \in X$ , the orbit  $t \mapsto x(t) = \Phi^t(x_0)$  with  $t \geq 0$  is the unique solution  $u(t, \cdot) = x(t)$  to (32) with initial condition  $x_0$ . While in practice the initial data will likely not lie on  $X$ , the exponential tracking property of the dynamics 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 Sect. 5.2, for data-driven approximation purposes, we will formally assume that the measure  $\mu$  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 (Lu et al. 2013; Lian et al. 2016) on infinite-dimensional dynamical systems that include the class of dissipative systems in which the KS system belongs to indicate that analogs of the assumptions made in Sect. 5.2 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 (Greene and Kim 1988; Arbruster et al. 1989; Kevrekidis et al. 1990; Cvitanović et al. 2009). In particular, it is a direct consequence of the structure of the governing equation (32) 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 Sect. 4.2, this implies that the dynamics on the inertial manifold is equivariant under the actions induced by the orthogonal group  $O(2)$  and the reflection group on the circle. In particular, under the assumption that the  $O(2)$  action preserves  $\mu$ , the theoretical spatial patterns recovered by POD and comparable eigendecomposition techniques would be linear combinations of finitely many Fourier modes (Holmes et al. 1996), which are arguably non-representative of the complex spatiotemporal patterns generated by the KS system. We emphasize 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 (Aubry et al. 1993), the presence of symmetries still dominates the leading spatial patterns recovered by POD and comparable eigendecomposition techniques utilizing scalar-valued kernels.

## 6.2 Analysis Datasets

In what follows, we present applications of VSA to data generated by the KS model in the regimes  $L = 12, 18, 22,$  and  $94$ . The former two are periodic regimes, exhibiting a traveling wave associated with a stable limit cycle ( $L = 12$ ), and an oscillatory solution likely corresponding to a homoclinic trajectory associated with an unstable fixed point ( $L = 18$ ). See Figure 3.2 in Kevrekidis et al. (1990) for a bifurcation diagram that includes these regimes, where  $\alpha = L^2/\pi^2$  is used as the bifurcation parameter; the cases  $L = 12$  and  $18$  correspond to  $\alpha \approx 14.6$  and  $32.3$ , respectively. The regimes at  $L = 22$  and  $94$  exhibit spatiotemporal chaos, with heteroclinic orbits within  $O(2)$  families of fixed points playing an important dynamical role (Arbruster et al. 1989). These two chaotic regimes have been investigated extensively in the literature (e.g., Cvitanović et al. 2009; Takeuchi et al. 2011).

We have integrated the KS model using the publicly available MATLAB code accompanying Cvitanovic et al. (2016). 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  $S = 65$  gridpoints and uniform quadrature weights,  $w_{s,S} = 1/S$ ), 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  $\phi_{NS,j}$  of the data-driven Markov operator  $P_{Q,NS}$  as described in Sect. 5. This operator is constructed using the family of kernels  $k_{Q,NS}$  in (47), in conjunction with the diffusion maps normalization in (50) to obtain the Markov kernel  $p_{Q,NS}$ . Note that  $k_{Q,NS}$  and  $p_{Q,NS}$  are data-driven approximations of  $k_Q$  and  $p_Q$  from (19) and (44), respectively. Further information on numerical implementation, including explicit formulas for the kernels and pseudocode, can be found in Appendix E.

In the  $L = 12, 18,$  and  $22$  experiments, we also compare the VSA results with spatiotemporal patterns computed via POD/PCA and NLSA (see Sect. 2 and Appendix F). The POD and NLSA methods are applied to the same KS data as VSA, and in the case of NLSA we use the same number of delays. The POD patterns are computed via (1), whereas those from NLSA are obtained via a procedure originally introduced in the context of SSA (Ghil et al. 2002). This procedure involves first reconstructing in delay-coordinate space through (1) applied to the observation map  $\tilde{F}_Q$  from (14), and then projecting down to physical data space by averaging over consecutive delay windows; see Appendix F for additional details. Empirically, this reconstruction approach is known to be more adept at capturing propagating signals than direct reconstruction of the observation map  $F$  via (1), though in the KS experiments discussed below the results from the two-step NLSA/SSA reconstruction and direct reconstruction are very similar.

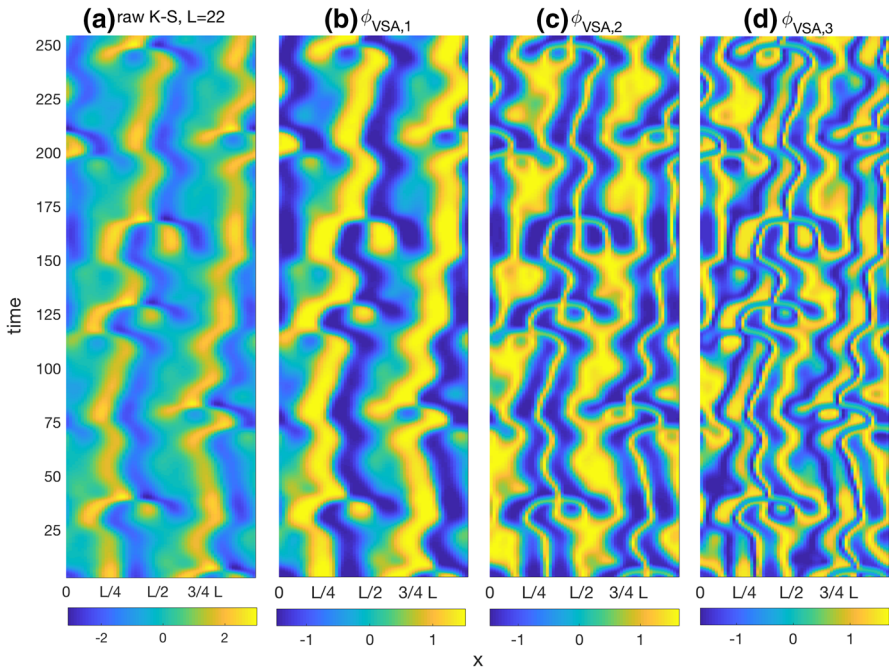
Before presenting our results, we recall that the patterns from all three methods are ordered in decreasing order of the corresponding eigenvalue of the kernel integral operator employed. We also note that since the vector-valued eigenfunctions from VSA are

directly interpretable as spatiotemporal patterns (see Sect. 3), and the VSA decomposition from (8) is given by linear combinations of eigenfunctions with scalar-valued coefficients  $c_j$ , comparing VSA eigenfunctions with POD and NLSA spatiotemporal patterns (which are formed by products of scalar-valued eigenfunctions of the corresponding kernel integral operators with spatial patterns) is meaningful. However, our depicted VSA eigenfunctions do not include multiplication by  $c_j$ , and therefore these comparisons are to be made only up to scale. To assess the efficacy of the VSA patterns in reconstructing the input signal, we compute their “fractional explained variances”,  $|\langle \vec{F}_{NS,j}, \vec{F} \rangle_{H_{NS}}|^2 / \|\vec{F}\|_{H_{NS}}^2$ , where  $\vec{F}_{NS,j}$  is the reconstructed pattern from (28), associated with eigenfunction  $\phi_{NS,j}$ . Similarly, in the case of NLSA and POD we compute  $|\langle \vec{F}_{NLSA,j}, \vec{F} \rangle_{H_{NS}}|^2 / \|\vec{F}\|_{H_{NS}}^2$  and  $|\langle \vec{F}_{POD,j}, \vec{F} \rangle_{H_{NS}}|^2 / \|\vec{F}\|_{H_{NS}}^2$ , where  $\vec{F}_{NLSA,j}$  and  $\vec{F}_{PCA,j}$  are the NLSA- and POD-reconstructed patterns [see (56) and (1)], respectively. Note that because the spatiotemporal patterns from VSA and NLSA are not necessarily orthogonal on  $H_{NS}$ , the explained variances from individual eigenfunctions are not necessarily additive, but they nevertheless provide a useful quantification of the degree of correlation between a given eigenfunction and the input signal. For simplicity, for the rest of this section we will drop the  $N$  and  $S$  subscripts from our notation for  $\phi_{NS,j}$ .

### 6.3 Results and Discussion

We begin by presenting VSA results obtained from a dataset of  $N = 1000$  samples at  $L = 22$ , using a small number of delays,  $Q = 15$ . According to Sect. 4.3.1, at this small  $Q$  value, VSA is expected to yield eigenfunctions  $\phi_j$ , which are approximately constant on the level sets of the input signal, and, with increasing  $j$ , capture smaller-scale variations in the directions transverse to the level sets. As is evident in Fig. 1, the leading three nonconstant eigenfunctions,  $\phi_1$ ,  $\phi_2$ , and  $\phi_3$ , indeed display this behavior, featuring wavenumbers 2, 3, and 4, respectively, in the directions transverse to the level sets. This behavior continues for eigenfunctions  $\phi_j$  with higher  $j$ . The corresponding fractional explained variances are 0.91,  $5.2 \times 10^{-4}$ , and 0.016, respectively, which demonstrates that even the one-term ( $l = 1$ ) reconstruction via (8) captures most of the signal variance.

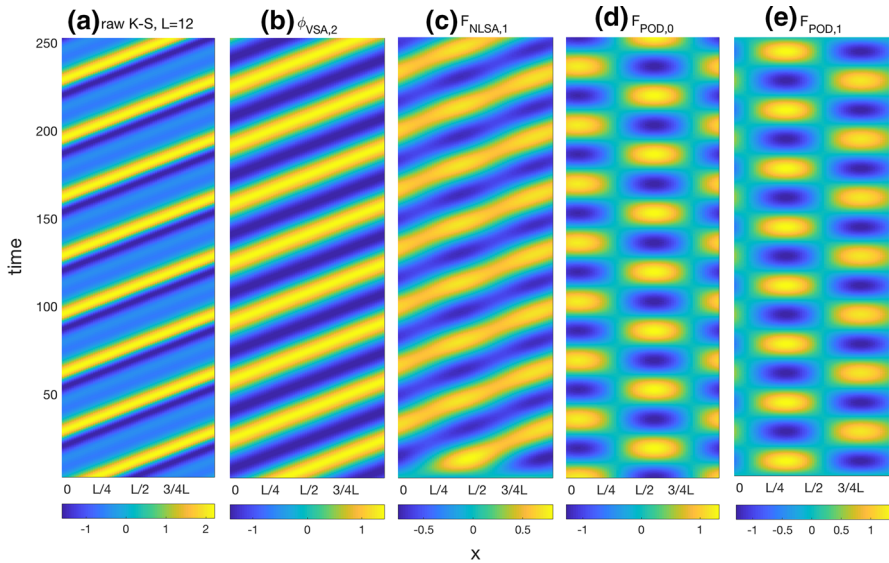
Next, we consider a dataset at the traveling-wave regime,  $L = 12$ , also with  $N = 1000$  samples, analyzed using  $Q = 150$  delays. The raw data and representative VSA eigenfunctions from this analysis, as well as NLSA and POD results, are displayed in Fig. 2. In this traveling-wave regime, the level sets of the signal (Fig. 2a) coincide with orbits on  $\Omega$  under the action  $\Gamma_{\Omega}^S$  associated with the  $O(2)$  symmetry group of the KS model (see Sect. 4.2). As a result, by Proposition 1, we expect the eigenfunctions  $\phi_j$  to be constant on the level sets of the input signal (i.e., the characteristics of the traveling wave). Indeed, as is evident from Fig. 2b, the traveling wave that the KS model develops at  $L = 12$  is well captured by VSA eigenfunction  $\phi_2$ . This eigenfunction forms a twofold degenerate pair with  $\phi_1$  (not shown), which exhibits a traveling wave similar to the one in  $\phi_2$ , but shifted by  $90^\circ$  in phase. The fractional explained variances due to  $\phi_1$  and  $\phi_2$  are 0.45 and 0.32, respectively. In contrast, due their tensor product structure, the POD reconstructions cannot represent a traveling



**Fig. 1** Raw data (a) and representative VSA eigenfunctions  $\phi_j$  (b–d) for the KS model with  $L = 22$ , using  $Q = 15$  delays. Notice that the patterns are approximately constant on the level sets of the raw data

mode through individual eigenfunctions. Instead, as shown in Fig. 2a, b, the two leading POD reconstructions are products of spatial Fourier modes of wavenumber  $2\pi/L$  with temporal sinusoids at the frequency of the KS traveling wave. The fractional explained variances due to these patterns are 0.33 and 0.32, respectively. Turning now to the NLSA results, the reconstruction based on the leading nonconstant eigenfunction from this method (Fig. 2c) is able to capture the traveling-wave characteristic of the signal due to the delay-averaging employed in the reconstruction procedure, despite the method utilizing a scalar-valued kernel as POD. As in the case of VSA, the second nonconstant NLSA pattern exhibits a traveling wave,  $90^\circ$  out-of-phase with the pattern in Fig. 2c. The fractional explained variances due to the first and second nonconstant NLSA eigenfunctions are both numerically equal to 0.40, at two significant figures.

Results from VSA, NLSA, and POD applied to our second periodic regime,  $L = 18$ , are shown in Fig. 3. As with the  $L = 12$  experiments, these results were obtained from a dataset of  $N = 1000$  samples, using  $Q = 150$  delays in the case of VSA and NLSA. As shown in Fig. 3a, the periodic solution at  $L = 18$  has the structure of a pair of standing oscillatory patterns, each pair comprising of a bimodal spatial profile taking both positive and negative values. Based on results in the bifurcation study in Kevrekidis et al. (1990), it is likely that this pattern is the outcome of a homoclinic orbit associated with an unstable steady state with a wavenumber- $4\pi/L$  spatial profile [see, e.g., the  $\alpha = 27.16$  pattern in Kevrekidis et al. (1990, Figure 2.1)]. The leading patterns from VSA and NLSA, shown in Fig. 3b, c, respectively, are consistent with this

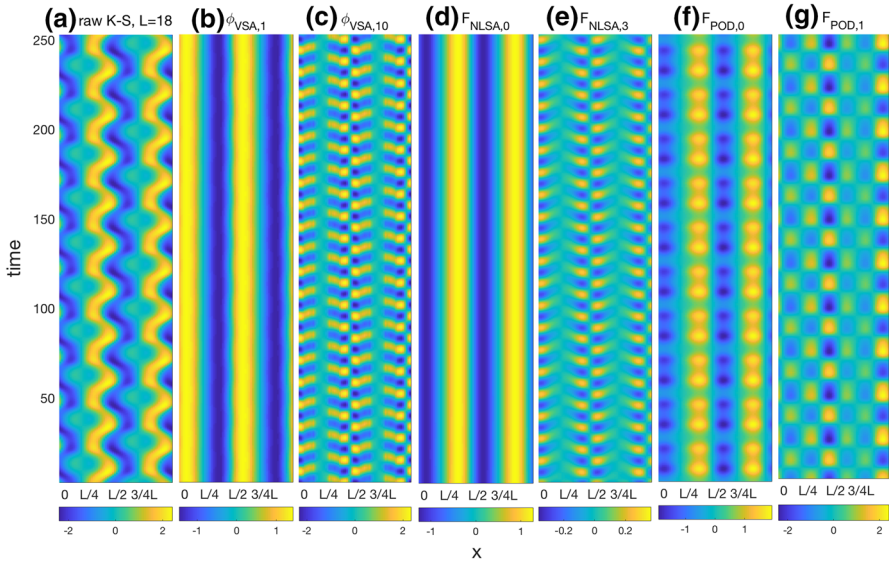


**Fig. 2** Raw data **(a)** and representative spatiotemporal patterns obtained through VSA **(b)**, NLSA **(c, d)**, and POD **(e, f)** from the KS model with  $L = 12$ , using  $Q = 150$  delays

behavior as they clearly capture a stationary pattern consistent with the unstable fixed point expected in this regime. In addition to the stationary patterns, VSA and NLSA both recover families of patterns featuring localized traveling waves, which exhibit an apparent amplitude modulation from the stationary patterns. The steady and traveling patterns from VSA (NLSA) explain 0.46 and 0.005 (0.26 and 0.002) of the signal variance. It is worthwhile noting that Kevrekidis et al. (1990) report unstable traveling-wave solutions in this dynamical regime, although they do not provide visualizations of these solutions that one could compare Fig. 3 with. The two leading patterns from POD (Fig. 3d, e) resemble superpositions of the stationary and traveling patterns from VSA/NLSA, with a fractional explained variance of 0.75 and 0.005, respectively.

Next, we consider longer datasets with  $N = 10,000$  samples (2500 natural time units), at  $L = 22$  and 94, analyzed using  $Q = 500$  delays. The raw data and representative VSA eigenfunctions from these analyses, as well as NLSA and POD results for  $L = 22$ , are displayed in Figs. 4 and 5, respectively. Figure 6 highlights a portion of the raw data and VSA eigenfunctions for  $L = 22$  over an interval spanning 1000 time units.

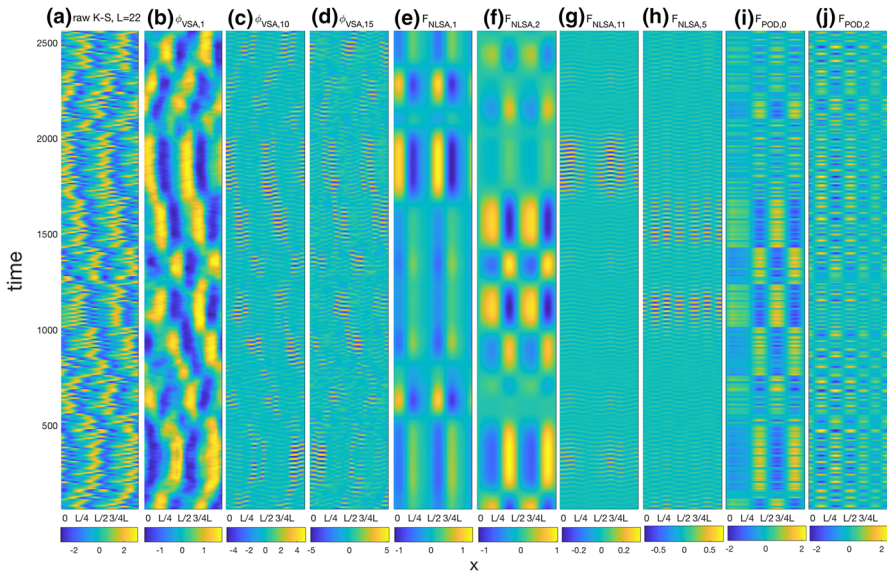
At large  $Q$ , we expect the eigenfunctions from VSA to lie approximately in finite-dimensional subspaces of the Hilbert space  $H$  of vector-valued observables associated with the point spectrum of the Koopman operator, thus acquiring timescale separation. This is clearly the case in the  $L = 22$  eigenfunctions in Figs. 4 and 6, where  $\phi_1$  is seen to capture the evolution of wavenumber- $4\pi/L$  structures, whereas  $\phi_{10}$  and  $\phi_{15}$  recover smaller-scale traveling waves embedded within the large-scale structures with a general direction of propagation either to the right ( $\phi_{10}$ ) or left ( $\phi_{15}$ ). The fractional explained variances associated with eigenfunctions  $\phi_1$ ,  $\phi_{10}$ , and  $\phi_{15}$  are 0.23, 0.047,



**Fig. 3** Raw data **(a)** and representative spatiotemporal patterns obtained through VSA **(b)**, NLSA **(c, d)**, and POD **(e, f)** from the KS model with  $L = 18$ , using  $Q = 150$  delays. Note that the NLSA eigenfunction associated with the pattern in **d** is a constant (as it is the leading eigenfunction of an ergodic Markov operator; see Appendix F), but nevertheless produces a pattern with nonconstant spatial structure since the signal in **a** has spatially nonconstant time mean

and 0.017, respectively. As expected, these values are smaller than the 0.91 value due to eigenfunction  $\phi_1$  for  $Q = 15$  in Fig. 1b, but are still fairly high despite the intermittent nature of the input signal. Ranked with respect to fractional explained variance,  $\phi_1$ ,  $\phi_{10}$ , and  $\phi_{15}$  are the first, fourth, and fifth among the  $Q = 200$  VSA eigenfunctions. Qualitatively, the spatiotemporal evolution of eigenfunction  $\phi_2$  in Figs. 4b and 6b is consistent with heteroclinic connections within an  $O(2)$  family of unstable equilibria at  $L = 22$  (Arbuster et al. 1989). Observe, in particular, that eigenfunction  $\phi_2$  at  $L = 22$  resembles a perturbed version of  $\phi_2$  at  $L = 18$  (Fig. 3b) associated with homoclinic dynamics. Similarly, the traveling-wave eigenfunctions  $\phi_{10}$  and  $\phi_{15}$  at  $L = 22$  (Figs. 4c, d and 6c, d) loosely resemble perturbed versions of  $\phi_{10}$  at  $L = 18$  (Fig. 3c).

In contrast, while the patterns from NLSA successfully separate the slow and fast timescales in the input signal [as expected theoretically at large  $Q$  (Giannakis 2017; Das and Giannakis 2019)], they are significantly less efficient in capturing its salient spatial features. Consider, for example, the leading two NLSA patterns shown in Fig. 4e, f. These patterns are clearly associated with the  $O(2)$  family of wavenumber- $4\pi/L$  structures in the raw data, but because they have a low rank, they are unable to represent the intermittent spatial translations of these patterns produced by chaotic dynamics in this regime. Their fractional explained variances are 0.13 and 0.15, respectively. Qualitatively, it appears that the NLSA patterns in Fig. 4e, f isolate periods during which the wavenumber- $4\pi/L$  structures are quasistationary and translated relative to each other by  $L/4$ . In other words, it appears that NLSA captures the unstable



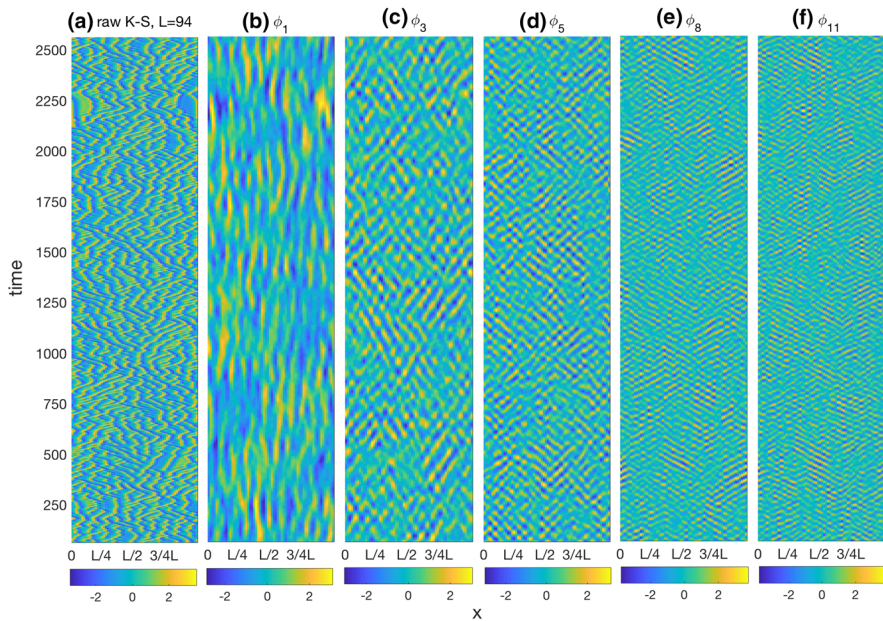
**Fig. 4** Raw data **(a)** and representative spatiotemporal patterns obtained through VSA **(b–d)**, NLSA **(e–h)**, and POD **(i, j)** from the KS model with  $L = 22$ , using  $Q = 500$  delays. The VSA pattern in **b** captures an  $O(2)$  family of unstable equilibria through an individual eigenfunction. Notice the manifestly non-separable character of this pattern with respect to the spatial and temporal coordinates. The 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 structures **(g, h)**, they are not representative of the intermittent character of the signal in space. The POD patterns in **i, j** exhibit low-rank structure as the NLSA patterns, but also mix timescales

equilibria that the system visits in the analysis time period through individual patterns, but does not adequately represent the transitory behavior associated with heteroclinic orbits connecting this family of equilibria. Moreover, due to the presence of the continuous  $O(2)$  symmetry, a complete description of the spatiotemporal signal associated with the wavenumber- $4\pi/L$  structures would require several modes. In contrast, VSA effectively captures this dynamics through a small set of leading eigenfunctions.

As can be seen in Fig. 4i, j, POD would also require several modes to capture the wavenumber- $4\pi/L$  unstable equilibria, but in this case the recovered patterns also exhibit an appreciable amount of mixing of the slow timescale characteristic of this family with faster timescales. Modulo this high-frequency mixing, the first POD pattern (Fig. 4i) appears to resemble the first NLSA pattern (Fig. 4f). The fractional explained variance of the leading two POD patterns, amounting to 0.23 and 0.22, respectively, is higher than the corresponding variances from NLSA, but this is not too surprising given their additional frequency content.

To summarize, the results at  $L = 22$  demonstrate that NLSA improves upon POD in that it achieves timescale separation through the use of delay-coordinate maps, and VSA further improves upon NLSA in that it quotients out the  $O(2)$  symmetry of the system, allowing efficient representation of intermittent space-time signals associated with heteroclinic dynamics in the presence of this symmetry. In separate

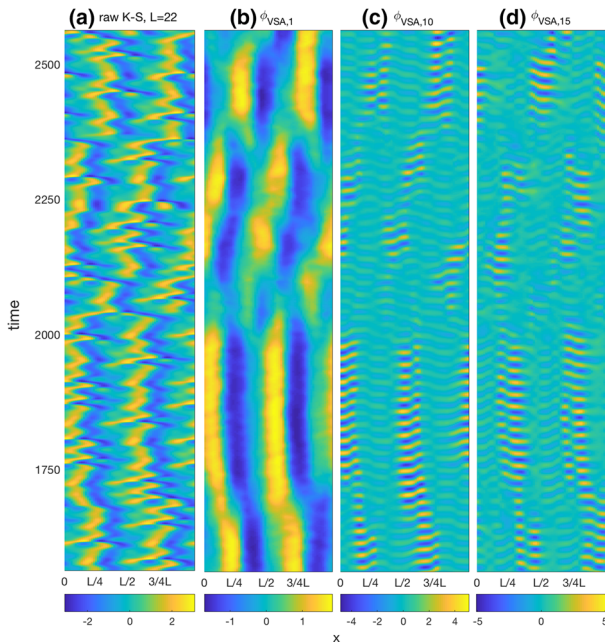




**Fig. 5** As in Fig. 4a–d, but for the KS system with  $L = 94$ . **b** Vector-valued eigenfunctions capturing an  $O(2)$  family of unstable equilibria. **c–f** Traveling-wave patterns

calculations, we have verified that the  $L = 22$  VSA patterns are robust under corruption of the data by i.i.d. Gaussian noise of variance up to 40% of the raw signal variance.

Turning now to the  $L = 94$  experiments, it is evident from Fig. 5a that the dynamical complexity in this regime is markedly higher than for  $L = 22$ , as multiple traveling and quasistationary patterns can now be accommodated in the domain, resulting in a spatiotemporal signal with high intermittency in both space and time. Despite this complexity, the recovered eigenfunctions (Fig. 5b–f) decompose the signal into a pattern  $\phi_1$  that captures the evolution of unstable fixed points and the heteroclinic connections between them, and other patterns,  $\phi_3$ ,  $\phi_5$ ,  $\phi_8$ , and  $\phi_{11}$ , dominated by traveling waves. The fractional explained variances associated with these patterns are  $6.0 \times 10^{-3}$  ( $\phi_1$ ), 0.020 ( $\phi_3$ ), 0.040 ( $\phi_5$ ), 0.067 ( $\phi_8$ ), and 0.066 ( $\phi_{11}$ ); 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  $\phi_1$  bears some qualitative similarities with the covariant Lyapunov vector (CLV) patterns identified at a nearby ( $L = 96$ ) KS regime in Takeuchi et al. (2011) (see Fig. 2 of that reference). Other VSA patterns also display qualitatively similar features to  $\phi_1$  and to CLVs. While such similarities are intriguing, they should be interpreted with caution as the existence of connections between VSA and CLV techniques is an open question.



**Fig. 6** As in Fig. 4a–d, but for the time interval [1625, 2625], highlighting the features of the small-scale waves in c, d

## 7 Conclusions

We have presented a method for extracting spatiotemporal patterns from complex dynamical systems, 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, called vector-valued spectral analysis (VSA), is that it operates directly on spaces of vector-valued observables appropriate for dynamical systems generating spatially extended patterns. This allows the 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, our approach enables efficient and physically meaningful decomposition of signals with intermittency in both space and time, while naturally factoring out dynamical symmetries present in the data. By incorporating delay-coordinate maps, the recovered patterns lie, in the 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 the ability to decompose multiscale signals into distinct coherent modes. We demonstrated with applications to the KS model in periodic and chaotic regimes that VSA recovers dynamically significant patterns, such as traveling waves and unstable steady states. The method also provides representations of intermittent patterns, such as heteroclinic 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.

As future research directions, it would be fruitful to explore applications of VSA to systems with non-Abelian group actions on the spatial domain; e.g., geophysical systems on the 2-sphere and molecular systems observed through scattering, both admitting  $SO(3)$  group actions. Moreover, the present formulation is based on delay-coordinate maps performed pointwise in the spatial domain, but one could also imagine variants of the approach utilizing patches, as frequently done in image processing. Another avenue of future research would be to use the eigenfunctions recovered by VSA as basis functions to represent the action of the Koopman operator on vector-valued observables, as done in Berry et al. (2015) and Zhao and Giannakis (2016) in the case of scalar-valued observables. Such data-driven representations of Koopman operators could be employed in spatiotemporal predictive models, with potentially improved skill due to the natural incorporation of dynamical symmetries. We anticipate the current VSA framework, as well possible future extensions, to be applicable across a broad range of disciplines dealing with complex spatiotemporal data, including climate dynamics (Wang et al. 2019) and neuroscience (Marrouch et al. 2018).

**Acknowledgements** 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.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A Koopman Operators on Scalar- and Vector-Valued Observables

### A.1 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  $H_X$  and its lift  $\tilde{U}^t$  acting on scalar-valued observables in  $H_\Omega$  (and, by the isomorphism  $H \simeq H_\Omega$ , 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., Sinai (2000), Budisić et al. (2012) and Eisner et al. (2015).

We begin by noting that for the class of  $C^1$  measure-preserving dynamical systems on manifolds studied here (see Sect. 2.1), the group  $U = \{U^t\}_{t \in \mathbb{R}}$  of Koopman operators is a strongly continuous unitary group. This means that for every  $f \in H_X$ , the map  $t \mapsto U^t f$  is continuous with respect to the  $H_X$  norm at every  $t \in \mathbb{R}$ . By Stone's theorem, strong continuity of  $U$  implies that there exists an unbounded, skew-adjoint operator  $V : D(V) \rightarrow H_X$  with dense domain  $D(V) \subset H_X$ , called the generator of  $U$ , such that  $U^t = e^{tV}$ . This operator completely characterizes Koopman group. Its

action on an observable  $f \in D(V)$  is given by

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

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

A distinguished class of observables in  $H_X$  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  $\alpha_j$  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  $\mu$ -almost every (a.e.)  $x \in X$ . 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  $\mu$ -a.e.  $x \in X$ . 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  $\mathcal{D}_X = \overline{\text{span}\{z_j\}}$  of  $H_X$  spanned by Koopman eigenfunctions has the closed-form expansion

$$U^t f = \sum_j e^{i\alpha_j t} c_j z_j, \quad c_j = \langle z_j, f \rangle_{H_X}. \quad (33)$$

This shows that the evolution of observables in  $\mathcal{D}_X$  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  $i\alpha_j$  and  $i\alpha_k$ , 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, these groups are finitely generated, there exists a collection of rationally-independent eigenfrequencies  $\tilde{\alpha}_1, \dots, \tilde{\alpha}_l$ , such that every eigenfrequency has the form  $\alpha_j = \sum_{k=1}^l j_k \tilde{\alpha}_k$ , where  $j = (j_1, \dots, j_l)$  is a vector of integers. Moreover, the Koopman eigenfunction corresponding to eigenfrequency  $\alpha_j$  is given by  $z_j = \tilde{z}_1^{j_1} \cdots \tilde{z}_l^{j_l}$ , 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 (33) that the evolution of every observable in  $\mathcal{D}_X$  can then

be 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 (33), that is,  $\mathcal{D}_X$  will generally be a strict subspace of  $H_X$ . In such cases, we have the orthogonal decomposition

$$H_X = \mathcal{D}_X \oplus \mathcal{D}_X^\perp, \tag{34}$$

which is invariant under the action of  $U^t$  for all  $t \in \mathbb{R}$ . For observables in the orthogonal complement  $\mathcal{D}_X^\perp$  of  $\mathcal{D}_X$  dynamical evolution is not determined by (33), but rather by a spectral expansion involving a continuous spectrum (intuitively, an uncountable set of frequencies). This evolution exhibits the characteristic behaviors associated with chaotic dynamics, such as decay of temporal correlations. In particular, it can be shown that for any  $f \in \mathcal{D}_X^\perp$  and  $g \in H_X$ , the quantity  $\int_0^t |\langle g, U^s f \rangle_{H_X}| ds/t$  vanishes as  $|t| \rightarrow \infty$ .

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  $H_\Omega$ . As stated in Sect. 4.2.2, these operators are obtained by a trivial lift  $\tilde{U}^t = U^t \otimes I_{H_Y}$  of the Koopman operators on  $H_X$ ; equivalently, we have  $\tilde{U}^t f = f \circ \tilde{\Phi}^t$ , where  $\tilde{\Phi}^t = \Phi^t \otimes I_Y$ , and  $I_Y$  is the identity map on  $Y$ . The group  $\tilde{U}$  is generated by the densely-defined, skew-adjoint operator  $\tilde{V} : D(\tilde{V}) \rightarrow H_\Omega$ ,

$$\tilde{V} f = \lim_{t \rightarrow 0} \frac{f \circ \tilde{\Phi}^t - f}{t}, \tag{35}$$

which is an extension of  $V \otimes I_{H_Y}$ . Moreover, analogously to the decomposition in (34), there exists an orthogonal decomposition

$$H_\Omega = \mathcal{D}_\Omega \oplus \mathcal{D}_\Omega^\perp, \quad \mathcal{D}_\Omega = \mathcal{D}_X \otimes H_Y, \quad \mathcal{D}_\Omega^\perp = \mathcal{D}_X^\perp \otimes H_Y,$$

which is invariant under  $\tilde{U}^t$  for all  $t \in \mathbb{R}$ . It is straightforward to verify that the eigenvalues of  $\tilde{U}^t$  are identical to those of  $U^t$ , i.e.,  $\Lambda^t = e^{\alpha t}$  for some eigenfrequency  $\alpha \in \mathbb{R}$ , and every eigenfunction  $\tilde{z}$  at eigenvalue  $\Lambda^t$  has the form

$$\tilde{z} = z \otimes \psi, \tag{36}$$

where  $z \in H_X$  is an eigenfunction of  $U^t$  at the same eigenvalue (unique up to normalization by ergodicity), and  $\psi$  an arbitrary spatial pattern in  $H_Y$ .

### A.2 Common Eigenfunctions with Kernel Integral Operators

We now examine the properties of common eigenfunctions between the Koopman operators  $\tilde{U}^t$  on  $H_\Omega$  and the kernel integral operators  $K_\infty$  from Theorem 3 with which they commute. In particular, let  $\tilde{z} \in H_\Omega$  be an eigenfunction of  $\tilde{U}^t$  at eigenvalue  $\Lambda^t$ . Then,

$$\tilde{U}^t K_\infty \tilde{z} = K_\infty \tilde{U}^t \tilde{z} = \Lambda^t K_\infty \tilde{z}, \tag{37}$$

which implies that  $K_\infty z$  is also an eigenfunction of  $\tilde{U}^t$  at the same eigenvalue. As stated in Appendix A.1, 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  $K_\infty \tilde{z} = \lambda \tilde{z}$  for some number  $\lambda$ , i.e., it is not necessarily the case that  $\tilde{z}$  is also an eigenfunction of  $K_\infty$  (despite the fact that (37) implies that every eigenspace of  $\tilde{U}^t$  is invariant under  $K_\infty$ ). 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  $K_\infty$  is a compact operator with finite-dimensional eigenspaces corresponding to nonzero eigenvalues. For each such eigenspace, there exists an orthonormal basis consisting of simultaneous eigenfunctions of  $K_\infty$  and  $\tilde{U}^t$ . To verify this explicitly, let  $W_l \subset H_\Omega$  be the eigenspace of  $K_\infty$  corresponding to eigenvalue  $\lambda_l \neq 0$ , and  $f$  an arbitrary element of  $W_l$ . Since

$$K_\infty \tilde{U}^t f = \tilde{U}^t K_\infty f = \lambda_l \tilde{U}^t f,$$

we can conclude that  $\tilde{U}^t f \in W_l$ , i.e., that  $W_l$  is a finite-dimensional invariant subspace of  $H_\Omega$  under  $\tilde{U}^t$ . Choosing an orthonormal basis  $\{\phi_{1l}, \dots, \phi_{m_l l}\}$  for this space, where  $m_l = \dim W_l$ , we can expand  $f = \sum_{j=1}^{m_l} c_j \phi_{jl}$  with  $c_j = \langle \phi_{jl}, f \rangle_{H_\Omega}$ , and compute

$$\tilde{U}^t f = \sum_{i,j=1}^{m_l} \phi_{il} \tilde{U}_{ij} c_j, \quad \tilde{U}_{ij} = \langle \phi_{il}, \tilde{U}^t \phi_{jl} \rangle_{H_\Omega}. \tag{38}$$

By unitarity of  $\tilde{U}^t$ , the  $m_l \times m_l$  matrix  $U$  with elements  $\tilde{U}_{ij}$  is unitary, and therefore unitarily diagonalizable. Let then  $\{v_j\}_{j=1}^{m_l}$  with  $v_j = (v_{1j}, \dots, v_{m_l j})^T$  be an orthonormal basis of  $\mathbb{C}^{m_l}$  consisting of eigenvectors of  $U$ , and  $\Lambda_{1l}^t, \dots, \Lambda_{m_l l}^t$  be the corresponding eigenvalues. It is a direct consequence of (38) that the set  $\{\tilde{z}_{1l}, \dots, \tilde{z}_{m_l l}\}$  with  $\tilde{z}_{jl} = \sum_{k=1}^{m_l} v_{kj} \phi_{kl}$  is an orthonormal basis of  $W_l$  consisting of Koopman eigenfunctions corresponding to the eigenvalues  $\Lambda_{jl}^t$ , which must thus be given by  $\Lambda_{jl} = e^{i\alpha_{jl}t}$  for some Koopman eigenfrequency  $\alpha_{jl} \in \mathbb{R}$ . Since every element of  $W_l$  is an eigenfunction of  $K_\infty$ , we conclude that the  $\tilde{z}_{jl}$  are simultaneous eigenfunctions of  $\tilde{U}^t$  and  $K_\infty$ .

## B Symmetry Group Actions

### B.1 Basic Definitions

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

1.  $\Gamma_Y$  is continuous, and  $\Gamma_Y(g, \cdot) : Y \rightarrow Y$  is a homeomorphism for all  $g \in G$ .

2.  $\Gamma_Y$  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  $G$ .

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

Assume now that  $\Gamma_Y^g$  preserves null sets with respect to the measure  $\nu$ . Then, the action of  $G$  on  $Y$  induces a continuous left action on the Hilbert space  $H_Y$  such that the action map  $\Gamma_{H_Y}^g : H_Y \rightarrow H_Y$  sends  $u \in H_Y$  to  $u \circ \Gamma_Y^{g^{-1}}$ . Assuming, further, that the state space manifold  $X$  is a subset of  $H_Y$  (as done in Sect. 4.2),  $G$  is considered to be a dynamical symmetry group if the following hold for all  $g \in G$  (e.g., Holmes et al. 1996):

1.  $X \subset H_Y$  is invariant under  $\Gamma_{H_Y}^g$ . Thus, we obtain a left group action  $\Gamma_X^g$  on  $X$  by restriction of  $\Gamma_{H_Y}$ .
2.  $\Gamma_X^g$  is differentiable for all  $g \in G$ , and the vector field  $v$  generating the dynamics  $\Phi^t : X \rightarrow X$  is invariant under the pushforward map  $\Gamma_{X^*}^g$  associated with  $\Gamma_X^g$ . 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  $\Gamma_X^g$  is a diffeomorphism (which is in turn a consequence of the fact that  $\Gamma_X^g$  is a differentiable group action).

### B.2 Common Eigenfunctions with Koopman Operators

In this section, we examine the structure of common eigenfunctions between the Koopman operator  $\tilde{U}^t = U^t \otimes I_{H_Y}$  on  $H_\Omega$  and the unitary representatives  $R_\Omega^g = R_X^g \otimes R_Y^g$  of the symmetry group  $G$  from Sect. 4.2.2, under the assumption that  $U^t$  commutes with  $R_X^g$  (which is equivalent to  $\tilde{U}^t$  commuting with  $R_\Omega^g$ ). As noted in Appendix A.1, every eigenfunction  $\tilde{z}$  of  $\tilde{U}^t$  has the form  $\tilde{z} = z \otimes \psi$ , where  $z$  is an eigenfunction of  $U^t$  corresponding to a simple eigenvalue  $\Lambda^t$ , and  $\psi$  a spatial pattern in  $H_Y$ . As result, because  $U^t$  and  $R_X^g$  commute, we have

$$U^t R_X^g z = R_X^g U^t z = \Lambda^t R_X^g z, \tag{39}$$

which shows that  $R_X^g z$  lies in the same Koopman eigenspace as  $z$ . Thus, since all eigenspaces of  $U^t$  are one dimensional, and  $R_X^g$  is unitary, there exists a complex

number  $\gamma_X^g$  with  $|\gamma_X^g| = 1$  such that

$$R_X^g z = \gamma_X^g z; \tag{40}$$

in other words,  $z$  is an eigenfunction of  $R_X^g$  at eigenvalue  $\gamma_X^g$ . Note that the  $\gamma_X^g$  are not necessarily simple eigenvalues.

Next, the commutativity between  $\tilde{U}^t$  and  $R_\Omega^g$ , in conjunction with (40), leads to

$$R_\Omega^g \tilde{z} = (R_X^g \otimes R_Y^g)(z \otimes \psi) = (\gamma_X^g z) \otimes (R_Y^g \psi),$$

which implies that  $\tilde{z}$  is an eigenfunction of  $R_\Omega^g$  if and only if  $\psi$  is an eigenfunction of  $R_Y^g$ . The  $R_\Omega^g$  eigenvalue corresponding to  $\tilde{z}$  is then given by

$$\gamma_\Omega^g = \gamma_Y^g / \gamma_X^g,$$

where  $\gamma_Y^g$  is the  $R_Y^g$  eigenvalue corresponding to  $\psi$ . Further, because  $R_X^g$  is unitary, we have

$$\gamma_\Omega^g = \gamma_Y^g \gamma_X^{g*} = \gamma_Y^g \gamma_X^{g^{-1}}.$$

We have thus obtained a characterization of the common eigenspaces of  $\tilde{U}^t$  and  $R_\Omega^g$ .

### C Construction and Properties of VSA Kernels

In this appendix, we describe in detail some aspects of the VSA kernel construction, namely the choice of distance scaling function for the kernels  $k_Q$  in (19) (Appendix C.1), and the normalization procedure to obtain the Markov kernels  $p_Q$  (Appendix C.2). We also establish some results on the behavior of these kernels in the limit of infinite delays,  $Q \rightarrow \infty$  (Appendix C.3). Throughout,  $M = A \times Y \subseteq \Omega$  will denote the compact support of the measure  $\rho$ .

#### C.1 Choice of Scaling Function

The distance scaling function  $a_Q$  is based on the corresponding function introduced in Giannakis (2017), which has dependencies on both the local sampling density and time tendency of the data, as follows.

##### C.1.1 Local Density Function

Let  $\bar{k}_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  denote the unscaled Gaussian kernel from (18), and  $\bar{K}_Q : H_\Omega \rightarrow H_\Omega$  the corresponding integral operator. Following Berry and Harlim (2016), we define the function

$$\sigma_Q = \bar{K}_Q 1_\Omega = \int_\Omega \bar{k}_Q(\cdot, \omega) d\rho(\omega), \tag{41}$$



which is continuous, strictly positive, and bounded away from zero on compact sets. 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  $\tilde{F}_Q : \Omega \rightarrow \mathbb{R}^Q$  from (14) is an embedding of  $M$  into  $\mathbb{R}^Q$ , 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  $\rho$  with respect to the Riemannian measure  $\text{vol}$  on  $M$  induced by the embedding. Moreover,  $\bar{\sigma}_Q$  has physical dimension (units) of  $\text{length}^{-m}$ , 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 distance in delay-coordinate space  $\mathbb{R}^Q$ , we can still interpret  $\sigma_Q$  from (41) as a local density-like function.

### C.1.2 Phase Space Velocity

Throughout this section, we will assume that the pointwise observation map  $F_y : X \rightarrow \mathbb{R}$  is of class  $C^2$  for every  $y \in Y$ , which is equivalent to assuming that the vector-valued observation map  $\vec{F}$  lies in the space  $C^2(X; C(Y))$ . 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  $\int_{\Omega} |\tilde{V}(\vec{F})|^2 d\rho$  is finite. Under this condition, the phase space speed function  $\xi_Q : \Omega \rightarrow \mathbb{C}$ , defined as

$$\xi_Q^2(x, y) = \frac{1}{Q} \sum_{q=0}^{Q-1} \zeta^2(\Phi^{-q\tau}(x), y), \quad x \in X, \quad y \in Y, \tag{42}$$

where

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

is continuously differentiable with respect to  $x$ . Note that  $\xi_Q(x, y)$  may vanish, e.g., if  $y$  lies in the boundary of  $Y$  and  $\vec{F}$  obeys time-independent boundary conditions. In the special case  $Q = 1$ ,  $\xi_Q$  will also vanish at local maxima/minima of the signal with respect to time. Phase space speed functions analogous to  $\xi_Q$  were previously employed in NLSA (Giannakis and Majda 2012) and related kernel (Giannakis 2015) and Koopman operator techniques (Giannakis 2017). For reasons that will be made clear below, we will adopt the approach introduced in Giannakis (2017, Section 6), which utilizes  $\xi_Q$  in such a way so that if  $\xi_Q(\omega)$  is zero, then  $a_Q(\omega)$  vanishes too.

### C.1.3 Scaling Function

With the density and phase space speed functions from Appendices C.1.1 and C.1.2, respectively, we define the continuous scaling function

$$a_Q = (\sigma_Q \xi_Q)^\gamma, \quad (43)$$

where  $\gamma$  is a positive parameter. This definition is motivated by Giannakis (2017), 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  $\gamma$  and in a suitable limit of vanishing kernel bandwidth parameter  $\epsilon$ , 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 Sect. 3.3 was shown to approximate the heat semigroup generated by a Laplace–Beltrami operator associated with this conformally transformed metric. In Giannakis (2017), 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 (Katok and Thouvenot 2006)] 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 state 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  $a_Q$ ) such that it vanishes at the fixed point is tantamount to increasing the bandwidth  $1/a_Q$  of the kernel by arbitrarily large amounts, 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),  $a_Q$  in (43) should behave similarly in regions of the product space  $\Omega$  where the rate-of-change of the observed data [measured by  $\xi_Q$  in (42)] is small. As stated in Sect. 3.1,  $\xi_Q$  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  $a_Q$  is to set the exponent parameter  $\gamma$ . According to Giannakis (2017), if  $M$  is an  $m$ -dimensional manifold embedded in  $\mathbb{R}^Q$ , the Riemannian metric associated with  $P_Q$  for the choice  $\gamma = 1/m$  has compatible volume form with the invariant measure of the dynamics, in the sense that the corresponding density  $\frac{d\rho}{d\text{vol}}$  is a constant. Moreover, the induced metric is also invariant under a class of conformal changes of observation map  $F_Q$ . 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  $\kappa = \int_{\Omega \times \Omega} \bar{k}_Q d\rho \times d\rho$  (or the corresponding data-driven quantity  $\kappa_{NS} = \int_{\Omega \times \Omega} \bar{k}_Q d\rho_{NS} \times d\rho_{NS}$ , where the measures  $\rho_{NS}$  are defined in Sect. 5) as a function of the bandwidth parameter  $\epsilon$ . As shown in Coifman et al. (2008) and Berry and Harlim (2016),  $d \log \kappa / d \log \epsilon$  can be interpreted as an effective dimension at a scale associated by the bandwidth parameter  $\epsilon$ . This motivates an automatic bandwidth tuning procedure where  $\epsilon$  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. The results presented in Sect. 6 are not too sensitive with respect to changes of  $\gamma$  around that value. In fact, for the systems studied here, the results remain qualitatively robust even if the velocity-dependent terms are not included in  $a_Q$  and  $a_{Q,N}$ . That is, qualitatively similar results can also be obtained using the scaling function  $a_Q = \sigma_Q^\gamma$ , which is continuous even if  $F$  is not continuously differentiable.

### C.2 Markov Normalization

Following the approach taken in NLSA algorithms and in Berry et al. (2013), Giannakis et al. (2015), Giannakis (2017) and Das and Giannakis (2019), we will construct a Markov kernel  $p_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  from a strictly positive, symmetric kernel  $k_Q : \Omega \times \Omega \rightarrow \mathbb{R}$ , meeting the conditions in Sect. 3.2, by applying the normalization procedure introduced in the diffusion maps algorithm (Coifman and Lafon 2006) and further developed in the context of general exponentially decaying kernels in Berry and Sauer (2016). For that, we first compute the functions

$$v_Q = K_Q 1_\Omega, \quad u_Q = K_Q(1_\Omega/v_Q),$$

where  $1_\Omega$  is the function on  $\Omega$  equal to 1 at every point. By the properties of  $k_Q$  and compactness of the support of  $\rho$ , both  $v_Q$  and  $u_Q$  are continuous, positive functions on  $\Omega$ , bounded away from zero on compact sets. We then define the kernel  $p_Q$  by

$$p_Q(\omega, \omega') = \frac{k_Q(\omega, \omega')}{u_Q(\omega)v_Q(\omega')}, \tag{44}$$

and the Markov property follows by construction. In Berry and Sauer (2016), the division of  $k_Q(\omega, \omega')$  by  $u_Q(\omega)$  and  $v_Q(\omega')$  to form  $p_Q(\omega, \omega')$  is referred to as left and right normalization, respectively. Because  $v_Q$  and  $u_Q$  are positive and bounded away from zero on compact sets,  $p_Q$  is continuous.

In general, a kernel of the class in (44), is not symmetric, and as a result the corresponding integral operator  $P_Q : H_\Omega \rightarrow H_\Omega$  is not self-adjoint. Nevertheless, by symmetry of  $k_Q$ ,  $P_Q$  is related to a self-adjoint compact operator  $\hat{P}_Q : H_\Omega \rightarrow H_\Omega$  by a similarity transformation. In particular, let  $f$  be a bounded function in  $L^\infty(\Omega, \rho)$ , and  $T_f : H_\Omega \rightarrow H_\Omega$  the corresponding multiplication operator by  $f$ . That is, for  $g \in H_\Omega$ ,  $T_f g$  is the function equal to  $f(\omega)g(\omega)$  for  $\rho$ -a.e.  $\omega \in \Omega$ . Defining  $\hat{P}_Q : H_\Omega \rightarrow H_\Omega$  as the self-adjoint kernel integral operator associated with the symmetric kernel

$$\hat{p}_Q(\omega, \omega') = \frac{k_Q(\omega, \omega')}{\hat{w}_Q(\omega)\hat{w}_Q(\omega')}, \quad \hat{w}_Q = \sqrt{u_Q v_Q}, \tag{45}$$

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

$$\hat{P}_Q = T_{w_Q} \circ P_Q \circ T_{w_Q}^{-1}, \quad w_Q = \sqrt{u_Q/v_Q}. \tag{46}$$

Due to (46),  $P_Q$  and  $\hat{P}_Q$  have the same eigenvalues  $\lambda_j$ , which are real by self-adjointness of  $\hat{P}_Q$ , and thus admit the ordering  $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots$  since  $P_Q$  is ergodic and Markov. Moreover, by compactness of  $P_Q$  and  $\hat{P}_Q$ , the eigenvalues have a single accumulation point at zero, and the eigenspaces corresponding to the nonzero eigenvalues are finite-dimensional.

Since  $\hat{P}_Q$  is self-adjoint and real, there exists an orthonormal basis  $\{\hat{\phi}_j\}_{j=0}^\infty$  of  $H_\Omega$  consisting of real eigenfunctions  $\hat{\phi}_j$  of  $\hat{P}_Q$  corresponding to  $\lambda_j$ . Moreover, the eigenfunctions corresponding to nonzero eigenvalues are continuous by the assumed continuity of kernels and compactness of  $M$ . In addition, due to (46), for every element  $\hat{\phi}_j$  of this basis, the continuous functions  $\phi_j = T_{w_Q}\hat{\phi}_j = w_Q\hat{\phi}_j$  and  $\phi'_j = T_{w_Q}^{-1}\hat{\phi}_j = \hat{\phi}_j/w_Q$  are eigenfunctions of  $P_Q$  and  $P_Q^*$ , respectively, corresponding to the same eigenvalue  $\lambda_j$ . The sets  $\{\phi_j\}_{j=0}^\infty$  and  $\{\phi'_j\}_{j=0}^\infty$  are then (non-orthogonal) Riesz bases of  $H_\Omega$  satisfying the bi-orthogonality relation  $\langle \phi'_i, \phi_j \rangle_{H_\Omega} = \delta_{ij}$ . In particular, every  $f \in H_\Omega$  can be uniquely expanded as  $f = \sum_{j=0}^\infty c_j \phi_j$  with  $c_j = \langle \phi'_j, f \rangle_{H_\Omega}$ , and we have  $P_Q f = \sum_{j=0}^\infty \lambda_j c_j \phi_j$ .

### C.3 Behavior in the Infinite-Delay Limit

In this section, we establish that the covariance and Gaussian kernels  $k_Q$  in (17)–(19), as well as the Markov kernels in Sect. 3.3, converge to well-defined, shift-invariant limits in the infinite-delay ( $Q \rightarrow \infty$ ) limit, in accordance with the conditions for VSA kernels listed in Sect. 3.2. Since all of these kernels are based on distances between datapoints in delay-coordinate space under the maps  $\tilde{F}_Q$  from (14), we begin by considering the family of distance-like functions  $d_Q : \Omega \times \Omega \rightarrow \mathbb{R}$ , with  $Q \in \mathbb{N}$  and

$$d_Q(\omega, \omega') = \frac{1}{Q^{1/2}} \|\tilde{F}_Q(\omega) - \tilde{F}_Q(\omega')\|_{\mathbb{R}^Q}.$$

This family of functions has the following important property:

**Proposition 5** *Suppose that the sampling interval  $\tau$  is such that there exists no eigenfrequency  $\alpha_j$  of the generator  $V$  such that  $e^{i\alpha_j\tau} = 1$ . Then, the sequence  $d_1, d_2, \dots$  converges in  $H_\Omega \times H_\Omega$  norm to a  $\tau$ -independent limit  $d_\infty \in H_\Omega \otimes H_\Omega$ , satisfying  $\tilde{U}^t \otimes \tilde{U}^t d_\infty = d_\infty$  for all  $t \in \mathbb{R}$ .*

**Proof** Let  $\omega = (x, y)$  and  $\omega' = (x', y')$  with  $x, x' \in X$  and  $y, y' \in Y$  be arbitrary points in  $\Omega$ . The  $H_\Omega \otimes H_\Omega$  convergence of  $d_Q$  to  $d_\infty$  follows from the Von Neumann mean ergodic theorem and the fact that

$$\begin{aligned} d_Q^2(\omega, \omega') &= \frac{1}{Q} \sum_{q=0}^{Q-1} |F_y(\Phi^{-q\tau}(x)) - F_{y'}(\Phi^{-q\tau}(x'))|^2 \\ &= \frac{1}{Q} \sum_{q=0}^{Q-1} d_1^2(\tilde{\Phi}^{-q\tau}(\omega), \tilde{\Phi}^{-q\tau}(\omega')) \end{aligned}$$

is a Birkhoff average under the product dynamical system  $\tilde{\Phi}^{q\tau} \otimes \tilde{\Phi}^{q\tau}$  on  $\Omega \times \Omega$  of the function  $d_1^2 : \Omega \times \Omega \rightarrow \mathbb{R}$ , which is bounded on the compact support of the corresponding invariant measure  $\rho \times \rho$ .

Next, let  $\{\tilde{U}^t \otimes \tilde{U}^t\}_{t \in \mathbb{R}}$  be the strongly continuous unitary group induced by  $\tilde{\Phi}^t \times \tilde{\Phi}^t$ . Denote the generator of this group by  $\hat{V}$ . To establish  $\tau$ -independence and invariance of  $d_\infty$  under  $\tilde{U}^t \otimes \tilde{U}^t$ , it suffices to show that  $d_\infty$  lies in the nullspace of  $\hat{V}$ . Indeed, by invariance of infinite Birkhoff averages, we have  $\tilde{U}^\tau \otimes \tilde{U}^\tau d_\infty = d_\infty$ , i.e.,  $d_\infty$  is an eigenfunction of  $\tilde{U}^\tau \otimes \tilde{U}^\tau$  at eigenvalue 1, and by the condition on  $\tau$  stated in the proposition and the fact that  $U^t$ ,  $\tilde{U}^t$ , and  $\tilde{U}^t \otimes \tilde{U}^t$  have the same eigenvalues (see Appendix A), this implies that  $d_\infty$  is an eigenfunction of  $\hat{V}$  at eigenvalue zero.  $\square$

Note that the condition on  $\tau$  in Proposition 5 holds for Lebesgue almost every  $\tau \in \mathbb{R}$ , as the set of Koopman eigenfrequencies  $\alpha_j$  is countable.

An immediate consequence of Proposition 5 is that given any continuous shape function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , the kernel  $k_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  with  $k_Q(\omega, \omega') = h(d_Q(\omega, \omega'))$  satisfies the conditions listed in Sect. 3.2. In particular, setting  $h$  to a Gaussian shape function,  $h(s) = e^{-s^2/\epsilon}$ , with  $\epsilon > 0$ , shows that the Gaussian kernel in (18) has the desired properties. That the covariance kernel in (17) also has these properties follows from an analogous result to Proposition 5 applied directly to the kernel  $k_Q$ , which in this case is equal to a Birkhoff average of  $k_1$ .

Next, we turn to the family of kernels in (19) utilizing scaled distances. These kernels have the general form

$$k_Q(\omega, \omega') = h(a_Q(\omega)a_Q(\omega')d_Q(\omega, \omega')),$$

where  $h : \mathbb{R} \rightarrow \mathbb{R}$  is continuous, so by Proposition 5, the required properties will follow if it can be shown that the sequence of scaling functions  $a_1, a_2, \dots$  converges in  $H_\Omega$  norm to a bounded function  $a_\infty \in L^\infty(\Omega, \rho)$ . That this is indeed the case for the choice of scaling functions described in Appendix C.1 follows from the facts that (i) the local density function  $\sigma_Q$  in (41) is itself derived from an unscaled Gaussian kernel  $\bar{k}_Q$ , which was previously shown to meet the required conditions, and (ii) the phase space velocity function  $\xi_Q$  from (42) is equal to a Birkhoff average of a continuous function.

Finally, the class of Markov kernels from Sect. 3.3 meets the necessary conditions because the diffusion maps normalization function  $v_Q$  is a continuous function determined by action of  $K_Q$  on a constant function, thus converging as  $Q \rightarrow \infty$  to a  $\tilde{U}^t$ -invariant function by the previous results on  $k_Q$ , and similarly  $u_Q$  is determined by action of  $K_Q$  on  $1/v_Q$  (see Appendix C.2).

## D Data-Driven Approximation

In this appendix, we present a proof of Theorem 4 for the most general class of integral operators on  $H_\Omega$  studied in this paper, namely the Markov operators  $P_Q$  associated with the kernels  $k_Q : \Omega \times \Omega \rightarrow \mathbb{R}$  from (19) utilizing the distance scaling functions  $a_Q$  in Appendix C.1, followed by Markov normalization as described in Sect. 3.3

and Appendix C.2. The convergence results for the operators not employing distance scaling and/or Markov normalization follow by straightforward modification of the arguments below.

### D.1 Data-Driven Markov Kernels

Because the distance scaling functions  $a_Q$  involve integrals with respect to  $\rho$  and time derivatives, in a data-driven setting we must work with kernels  $k_{Q,NS} : \Omega \times \Omega \rightarrow \mathbb{R}_+$  approximating  $k_Q$ , where

$$k_{Q,NS}(\omega, \omega') = \exp \left( -\frac{a_{Q,NS}(\omega)a_{Q,NS}(\omega')}{\epsilon_Q} \sum_{q=0}^{Q-1} |F_y(\Phi^{q\tau}(x)) - F_{y'}(\Phi^{q\tau}(x'))|^2 \right), \tag{47}$$

$\omega = (x, y)$ ,  $\omega' = (x', y')$ , and  $a_{Q,NS} \in C(\Omega)$  are scaling functions approximating  $a_Q$ .

Our construction of  $a_{Q,NS}$  follows closely that of  $a_Q$  in (43), that is, we set

$$a_{Q,NS} = (\sigma_{Q,NS} \xi_{Q,N})^\gamma,$$

where  $\gamma$  is the same exponent parameter as in (43), and  $\sigma_{Q,NS}$  and  $\xi_{Q,N}$  are continuous functions approximating  $\sigma_Q$  and  $\xi_Q$ , respectively. To construct  $\sigma_{Q,NS}$ , we introduce the integral operator  $\bar{K}_{Q,NS} : H_{\Omega,NS} \rightarrow H_{\Omega,NS}$ ,

$$\bar{K}_{Q,NS}f(\omega) = \int_{\Omega} \bar{k}_Q(\omega, \omega')f(\omega') \, d\rho_{NS}(\omega'),$$

where  $\bar{k}_Q$  denotes the unscaled Gaussian kernel as in (41), and define

$$\sigma_{Q,NS} = \bar{K}_{Q,NS}1_{\Omega} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{s=0}^{S-1} \bar{k}_Q(\cdot, \omega_{ns})w_{s,S}. \tag{48}$$

Moreover, following Giannakis and Majda (2012), Giannakis (2015), and Giannakis (2017), in the data-driven setting we approximate the function  $\zeta$  used in the definition of  $\xi_Q$  in (42) by a continuous function  $\zeta_{\tau} : \Omega \rightarrow \mathbb{R}$  that provides a finite-difference approximation of  $\zeta$  with respect to the sampling interval  $\tau$ . As a concrete example, we consider a first-order, backward difference scheme,

$$\zeta_{\tau}(x, y) = \frac{|F_y(x) - F_y(\Phi^{-\tau}(x))|}{\tau}, \quad x \in X, \quad y \in Y, \tag{49}$$

which, under the assumed differentiability properties of  $F_y$  (see Appendix C.1.2), converges as  $\tau \rightarrow 0$  to  $\zeta$ , uniformly on compact sets (in particular,  $\mathcal{V}$ ). 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). We define  $\xi_{Q,N}$  as the function resulting by substituting  $\zeta$  by  $\zeta_{\tau(N)}$  in (42). As we will establish in Appendix D.5, with these definitions, and under weak convergence of the measures  $\rho_{NS}$  in (31),  $a_{Q,NS}$  converges to  $a_Q$  uniformly on the compact set  $\mathcal{V}$ .

Next, we construct the Markov kernels  $p_{Q,NS} : \Omega \times \Omega \rightarrow \mathbb{R}$  of the operators  $P_{Q,NS} : H_{\Omega,NS} \rightarrow H_{\Omega,NS}$  in the statement of the theorem by applying diffusion maps normalization to  $k_{Q,NS}$  as in Appendix C.2, viz.

$$\begin{aligned}
 p_{Q,NS}(\omega, \omega') &= \frac{k_{Q,NS}(\omega, \omega')}{u_{Q,NS}(\omega)v_{Q,NS}(\omega')}, \\
 v_{Q,NS} &= K_{Q,NS}1_{\Omega}, \quad u_{Q,NS} = K_{Q,NS} \left( \frac{1_{\Omega}}{v_{Q,NS}} \right).
 \end{aligned}
 \tag{50}$$

where  $K_{Q,NS}$  is the kernel integral operator on  $H_{\Omega,NS}$  associated with  $k_{Q,NS}$ . Note that  $u_{Q,NS}$  and  $v_{Q,NS}$  are positive, continuous functions on  $\Omega$ , bounded away from zero on compact sets.

### D.2 Proof of Theorem 4

We will need the important notion of *compact convergence* of operators on Banach spaces (von Luxburg et al. 2008; Chatelin 2011).

**Definition 6** A sequence of bounded operators  $T_n : E \rightarrow E$  on a Banach space  $E$  is said to converge compactly to a bounded operator  $T : E \rightarrow E$  if  $T_n$  converges to  $T$  pointwise (i.e.,  $T_n f \rightarrow T f$  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).

Compact convergence is stronger than pointwise convergence, but weaker than convergence in operator norm. For our purposes, it is useful as it is sufficient to imply convergence of isolated eigenvalues of bounded operators (Chatelin 2011), and hence convergence of nonzero eigenvalues of compact operators and their corresponding eigenspaces. In particular, Theorem 4 is a corollary of the following theorem, proved in Appendix D.3:

**Theorem 7** *Under the assumptions of Theorem 4, the following hold:*

- (a)  $\tilde{P}_{Q,NS}$  and  $\tilde{P}_Q$  are both compact operators on  $C(\mathcal{V})$ . As a result, their nonzero eigenvalues have finite multiplicities, and accumulate only at zero.
- (b) As  $N, S \rightarrow \infty$ ,  $\tilde{P}_{Q,NS}$  converges compactly to  $\tilde{P}_Q$ .
- (c)  $\lambda_j$  is a nonzero eigenvalue of  $\tilde{P}_Q$  if and only if it is a nonzero eigenvalue of  $P_Q$ . Moreover, if  $\phi_j$  is an eigenfunction of  $P_Q$  corresponding to that eigenvalue, then  $\tilde{\phi}_j \in C(\mathcal{V})$  with

$$\tilde{\phi}_j(\omega) = \frac{1}{\lambda_j} \int_{\Omega} p_Q(\omega, \omega') \phi_j(\omega') \, d\rho(\omega') \quad (51)$$

is an eigenfunction of  $\tilde{P}_Q$  corresponding to the same eigenvalue. Analogous results hold for every nonzero eigenvalue of  $\lambda_{NS,j}$ , of  $\tilde{P}_{Q,NS}$ , and corresponding eigenfunctions  $\phi_{NS,j}$  and  $\tilde{\phi}_{NS,j}$  of  $P_{Q,NS}$  and  $\tilde{P}_{Q,NS}$ , respectively, where

$$\tilde{\phi}_{NS,j}(\omega) = \frac{1}{\lambda_{NS,j}} \int_{\Omega} p_{Q,NS}(\omega, \omega') \phi_{NS,j}(\omega') \, d\rho_{NS}(\omega'). \quad (52)$$

To verify that Theorem 7 indeed implies Theorem 4 (with  $K_Q$  replaced by  $P_Q$  and  $K_{Q,NS}$  by  $P_{Q,NS}$ ), note that since the eigenvalue  $\lambda_j$  in the statement of Theorem 4 is nonzero, it follows from Theorem 7(c) that it is an eigenvalue of  $\tilde{P}_Q$ , and that  $\tilde{\phi}_j$  from (51) is a corresponding eigenfunction. Moreover, since, by Theorem 7(b),  $\tilde{P}_{Q,NS}$  converges to  $\tilde{P}_Q$  compactly (and thus in spectrum for nonzero eigenvalues Chatelin 2011), there exist  $N_0, S_0 \in \mathbb{N}$  such that the  $j$ -th eigenvalues  $\lambda_{NS,j}$  of  $\tilde{P}_{Q,NS}$  are all nonzero for  $N \geq N_0$  and  $S \geq S_0$ , and thus, by Theorem 7(c), they are eigenvalues of  $P_{Q,NS}$  converging to  $\lambda_j$ , as claimed in Theorem 4. The existence of eigenfunctions  $\phi_{NS,j}$  of  $P_{Q,NS}$  corresponding to  $\lambda_{NS,j}$ , such that  $\tilde{\phi}_{NS,j}$  from (52) converges uniformly to  $\tilde{\phi}_j$  is shown in Das and Giannakis (2019). This completes our proof of Theorem 4.

### D.3 Proof of Theorem 7

Our proof of Theorem 7 draws heavily on the spectral convergence results on data-driven kernel integral operators established in von Luxburg et al. (2008) and Das and Giannakis (2019), though it requires certain modifications appropriate for the class of kernels in (19) utilizing scaled distances, 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  $\tilde{\phi}_{NS,j}$  in (52). Throughout this section, 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  $\tilde{P}_{Q,NS}$ . We will also drop  $S$  subscripts representing the number of sampled points in the spatial domain  $Y$ , with the understanding that  $N \rightarrow \infty$  limits correspond to  $S \rightarrow \infty$  followed by  $N \rightarrow \infty$  limits.

### D.4 Proof of Claim (a)

Let  $\bar{d} : \Omega \times \Omega \rightarrow \mathbb{R}$  be any metric on  $\Omega$ . We begin by establishing the following result on the kernel  $p_Q$ :

**Lemma 8** *The map  $\omega \mapsto p_Q(\omega, \cdot)$  is a continuous map from  $\mathcal{V}$  to  $C(\mathcal{V})$ , that is, for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for all  $\omega, \omega' \in \mathcal{V}$  satisfying  $\bar{d}(\omega, \omega') < \delta$ ,*



$$\|p_Q(\omega, \cdot) - p_Q(\omega', \cdot)\|_{C(\mathcal{V})} < \epsilon.$$

**Proof** Suppose that the claim is not true. Then, there exists  $\epsilon > 0$  and sequences  $\omega_n, \omega'_n \in \mathcal{V}$ , such that, as  $n \rightarrow \infty$ ,  $\bar{d}(\omega_n, \omega'_n) \rightarrow 0$  and  $\|p_Q(\omega_n, \cdot) - p_Q(\omega'_n, \cdot)\|_{C(\mathcal{V})} > \epsilon$ . As a result, there exists  $\omega''_n \in \mathcal{V}$  such that  $|p_Q(\omega_n, \omega''_n) - p_Q(\omega'_n, \omega''_n)| > \epsilon$ . However, this contradicts the fact that  $p_Q$  is continuous since  $(\omega_n, \omega''_n) \in \mathcal{V} \times \mathcal{V}$  converges to  $(\omega'_n, \omega''_n)$ .  $\square$

We now return to the 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

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

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

$$|g_n(\omega) - g_n(\omega')| \leq \|p_Q(\omega, \cdot) - p_Q(\omega', \cdot)\|_{C(\mathcal{V})} \|f_n\|_{C(\mathcal{V})},$$

and the equicontinuity of  $\{g_n\}$  follows from Lemma 8. It therefore follows from the Arzelà–Ascoli theorem that  $g_n$  has a limit point, and thus that  $P_Q$  is compact, as claimed.

### D.5 Proof of Claim (b)

According to Definition 6, 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  $\hat{p}_{Q,N} : \Omega \times \Omega \rightarrow \mathbb{R}_+$ ,

$$\hat{p}_{Q,N}(\omega, \omega') = \frac{k_Q(\omega, \omega')}{u_{Q,N}(\omega)v_{Q,N}(\omega')}, \tag{53}$$

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

$$\tilde{P}_{Q,N} f = \int_{\Omega} p_Q(\cdot, \omega) f(\omega) d\rho_N(\omega), \quad \hat{P}_{Q,N} f = \int_{\Omega} \hat{p}_{Q,N}(\cdot, \omega) f(\omega) d\rho_N(\omega),$$

we have

$$\begin{aligned} \eta_N \leq & \|P_Q f - \tilde{P}_{Q,N} f\|_{C(\mathcal{V})} + \|\tilde{P}_{Q,N} f - \hat{P}_{Q,N} f\|_{C(\mathcal{V})} \\ & + \|\hat{P}_{Q,N} f - P_Q f\|_{C(\mathcal{V})}. \end{aligned} \quad (54)$$

That is, we can bound  $\eta_N$  by a sum of contributions due to (i) errors in approximating integrals with respect to the invariant measure  $\rho$  by the sampling measure  $\rho_N$  (the first term in the right-hand side); (ii) errors in approximating the left and right normalization functions  $u_Q$  and  $v_Q$  by their data-driven counterparts,  $u_{Q,N}$  and  $v_{Q,N}$ , respectively (the second term in the right-hand side); and (iii) 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,

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

By the weak convergence of the measures  $\rho_N$  to  $\rho$  (see (31)) in conjunction with the continuity of  $p_Q$ , it follows that  $\tilde{P}_{Q,N} f(\omega)$  converges to  $P_Q f(\omega)$ , 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 9** Let  $\mathbb{E} : C(\mathcal{V}) \rightarrow \mathbb{C}$  and  $\mathbb{E}_N : C(\mathcal{V}) \rightarrow \mathbb{C}$  be the expectation operators with respect to the measures  $\rho_N$  and  $\rho$ , respectively, i.e.,

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

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

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

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  $\|\tilde{P}_{Q,N} f - P_Q f\|_{C(\mathcal{V})}$  vanishes as  $N \rightarrow \infty$ . That this is indeed the case follows from von Luxburg et al. (2008, Proposition 11).

Next, we turn to the second and third terms in (54). To bound these terms, we first establish convergence of the data-driven distance scaling functions  $a_{Q,N}$  to  $a_Q$ .

**Lemma 10** *Restricted to  $\mathcal{V}$ , the scaling functions  $a_{Q,N}$  from Appendix C.1 converge uniformly as  $N \rightarrow \infty$  to  $a_Q$ .*

**Proof** It follows from the definition of  $a_Q$  and  $a_{Q,N}$  in (43) that for all  $\omega \in \mathcal{V}$ ,

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

Thus, since  $\xi_{Q,N}$  converges uniformly to  $\xi_Q$  by continuous differentiability of the observation map  $F$  on the compact set  $\mathcal{V}$  (see Appendix C.1.2),  $a_{Q,N}$  will converge uniformly to  $a_Q$  if  $\sigma_{Q,N}$  converges uniformly to  $\sigma_Q$ . 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. □

Lemma 10, in conjunction with the continuity of the kernel shape function used throughout this work (see Sect. 3.1), implies the following: □

**Corollary 11** *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 (54),  $\|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})}$ . It follows from the definition of the kernels  $p_{Q,N}$  and  $\hat{p}_{Q,N}$  via (50) and (53), respectively, that

$$\begin{aligned} &\|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} \\ &\leq \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \left\| \frac{1}{u_{Q,N} \otimes v_{Q,N}} - \frac{1}{u_Q \otimes v_Q} \right\|_{C(\mathcal{V} \times \mathcal{V})}. \end{aligned}$$

By our assumptions on kernels stated in Sect. 3.2, the functions  $u_Q, v_Q, u_{Q,N}$ , and  $v_{Q,N}$  are bounded away from zero on  $\mathcal{V}$ , uniformly with respect to  $N$ . Therefore, there exists a constant  $c > 0$ , independent of  $N$ , such that

$$\begin{aligned} \|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} &\leq c \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \|u_Q \otimes v_Q - u_{Q,N} \otimes v_{Q,N}\|_{C(\mathcal{V} \times \mathcal{V})} \\ &= c \|k_Q\|_{C(\mathcal{V} \times \mathcal{V})} \|f\|_{C(\mathcal{V})} \|u_Q - u_{Q,N}\|_{C(\mathcal{V})} \|v_Q - v_{Q,N}\|_{C(\mathcal{V})}. \end{aligned}$$

Observe now that

$$\|v_Q - v_{Q,N}\|_{C(\mathcal{V})} = \max_{\omega \in \mathcal{V}} |v_Q(\omega) - v_{Q,N}(\omega)|$$

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

Since  $\|k_{Q,N} - k_Q\|_{C(\mathcal{V} \times \mathcal{V})}$  converges to zero by Corollary 11, it follows that

$$\lim_{N \rightarrow \infty} \|v_Q - v_{Q,N}\|_{C(\mathcal{V})} = 0 \tag{55}$$

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 (55), in conjunction with the fact that  $\|u_Q - u_{Q,N}\|_{C(\mathcal{V})}$  is bounded, is sufficient to deduce that  $\lim_{N \rightarrow \infty} \|\tilde{P}_{Q,N}f - \hat{P}_{Q,N}f\|_{C(\mathcal{V})} = 0$ .

We now turn to the third term in (54),  $\|\hat{P}_{Q,N}f - P_{Q,N}f\|_{C(\mathcal{V})}$ . 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  $u_{Q,N}$  and  $v_{Q,N}$  are both uniformly bounded away from zero, that there exists a 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, the convergence of  $\|\hat{P}_{Q,N}f - P_{Q,N}f\|_{\mathcal{V}}$  to zero follows from Corollary 11.

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 6, 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 von Luxburg et al. (2008, Proposition 13) for more details. This completes our proof of Claim (b).

### E Numerical Implementation and Pseudocode

In this appendix, we describe the numerical procedure used to obtain the VSA results in Sect. 6, and provide pseudocode for the main steps. This implementation uses the variable-bandwidth Gaussian kernels and diffusion maps normalization described

in Appendix C. Implementations using other kernels and/or normalizations can be performed similarly.

The starting point of the procedure is a spatiotemporal signal  $F_{y_s}(x_n)$ , with  $n \in \{-Q + 1, \dots, N\}$ ,  $s \in \{0, \dots, S - 1\}$ , as in Sect. 5. Here, we have taken the time-indexing of the data to start at  $n = -Q$  in anticipation of the fact that we will be performing delay-coordinate maps with  $Q$  delays via (14) and first-order finite-differences via (49). We have split the entire procedure to compute the VSA eigenvalues and eigenfunctions  $(\lambda_{NS,j}, \phi_{NS,j})$ , and perform the decomposition of the signal in (28), into five algorithms. Algorithms 1–3 are auxiliary algorithms for delay embedding (Algorithm 1) and computation of the phase speeds  $\xi_{Q,N}$  (Algorithm 2) and densities  $\sigma_{Q,NS}$  (Algorithm 3) from Appendix D.1. Then, in Algorithm 4, we employ the output of Algorithms 1–3 to evaluate the VSA kernel  $k_{Q,NS}$  in (47), and solve the eigenvalue problem associated with the Markov operator  $P_{Q,NS}$  from Appendix D.1 to obtain the eigenpairs  $(\lambda_{NS,j}, \phi_{NS,j})$  and the dual eigenvectors  $\phi'_{NS,j}$ . In Algorithm 5, the eigenfunctions from Algorithm 4 are employed to carry out the signal decomposition in (28), in terms of the patterns  $\vec{F}_j \in H_{NS}$ .

Note that the operator  $P_{Q,NS}$  is related to a self-adjoint operator  $\hat{P}_{Q,NS} : H_{\Omega,NS} \rightarrow H_{\Omega,NS}$  by a similarity transformation analogous to (46). In Algorithm 4, we take advantage of this structure by first computing the eigenvalues and eigenvectors of  $\hat{P}_{Q,NS}$  (which can be done with higher efficiency and stability through the use of specialized solvers for symmetric problems), and then acting on these eigenvectors by a diagonal operator to obtain eigenvectors of  $P_{Q,NS}$ . In accordance with Sect. 5.1,  $\hat{P}_{Q,NS}$  is represented by an  $(NS) \times (NS)$  matrix  $\hat{P}$  with elements

$$\hat{P}_{mr,ns} = \hat{p}_{Q,NS}(\omega_{mr}, \omega_{ns}) / (NS),$$

$$m, n \in \{0, \dots, N - 1\}, \quad r, s \in \{0, \dots, S - 1\}, \quad \omega_{ns} = (x_n, y_s),$$

where  $\hat{p}_{Q,NS}$  is a symmetric kernel analogous to  $\hat{p}_Q$  in (45). Moreover, the eigenvectors  $\phi_{NS,j}$  and  $\hat{\phi}_{NS,j}$  of  $P_{Q,NS}$  and  $\hat{P}_{Q,NS}$  are represented by  $NS$ -dimensional column vectors  $\underline{\phi}_j$  and  $\hat{\underline{\phi}}_j$ , respectively, with elements  $\phi_{j,ns} = \phi_{NS,j}(\omega_{ns})$  and  $\hat{\phi}_{j,ns} = \hat{\phi}_{NS,j}(\omega_{ns})$ . The dual eigenvectors  $\phi'_{NS,j}$  and vector-valued patterns  $\vec{F}_{NS,j}$  are similarly represented by column vectors  $\underline{\phi}'_j$  and  $\underline{F}_j$  in  $\mathbb{R}^{NS}$ , respectively. Note that our usage of double indices to label matrix and vector elements (e.g.,  $\hat{P}_{mr,ns}$ ) implies that an ordering has been chosen for the time-space index pairs  $(n, s)$ .

As is common practice, we take advantage of the exponential decay of Gaussian kernels to approximate  $\hat{P}$  by a sparse matrix, whose sparsity is controlled by an integer neighborhood parameter  $k_{nn} \leq NS$  (approximately equal to the number of nonzero elements in each row of  $\hat{P}$ ). In the experiments in Sect. 6,  $k_{nn}/(NS)$  was typically 0.01. Algorithms 3 and 4 both employ an automatic procedure to tune the Gaussian kernel bandwidth parameter  $\epsilon$ , introduced in Coifman et al. (2008) and further refined in Berry and Harlim (2016). See Appendix A in Berry et al. (2015) for further details. In what follows, we will use the shorthand notations  $F_{ns} = F_{y_s}(x_n)$  and  $\vec{F}_{ns} = \vec{F}_Q(\omega_{ns})$  for the values of the raw and delay-embedded signal [the latter, given by (14)]. We will also abbreviate  $\xi_{ns} = \xi_{Q,N}(\omega_{ns})$  and  $\sigma_{ns} = \sigma_{Q,NS}(\omega_{ns})$ .

**Algorithm 1** (Delay embedding)

- Inputs
  - Number of delays  $Q$
  - Spatiotemporal signal  $F_{ns} \in \mathbb{R}, n \in \{-Q+1, \dots, N-1\}, s \in \{0, \dots, S-1\}$
- Outputs
  - Delay-embedded data vectors  $\tilde{F}_{ns} \in \mathbb{R}^Q, n \in \{-1, \dots, N-1\}, s \in \{0, \dots, S-1\}$
- Steps
  1. Set  $\tilde{F}_{ns} = (F_{ns}, F_{n-1,s}, \dots, F_{n-Q+1,s})^\top$ .

**Algorithm 2** (Phase space speed)

- Inputs
  - Delay-embedded vectors  $\tilde{F}_{ns}$  from Algorithm 1
  - Sampling interval  $\tau > 0$
- Outputs
  - Phase space speeds  $\xi_{ns} \geq 0, n \in \{0, \dots, N-1\}, s \in \{0, \dots, S-1\}$
- Steps
  1. Set  $\xi_{ns} = \|\tilde{F}_{ns} - \tilde{F}_{n-1,s}\|_2 / (\tau Q^{1/2})$ .

**Algorithm 3** (Density estimation)

- Inputs
  - Delay-embedded vectors  $\tilde{F}_{ns}$  from Algorithm 1
  - Neighborhood parameter  $k_{nn} \leq NS$
  - Candidate bandwidth parameter values  $\{\epsilon_{-b}, \dots, \epsilon_b\}$ , with  $b \in \mathbb{N}$  and  $\epsilon_i = 2^i$
- Outputs
  - Densities  $\sigma_{ns} > 0, n \in \{0, \dots, N-1\}, s \in \{0, \dots, S-1\}$
  - Estimated dimension  $\hat{m}$
- Steps
  1. For every  $m, n \in \{0, \dots, N-1\}, r, s \in \{0, \dots, S-1\}$ , compute the pairwise distances

$$d_{mr,ns} = \|\tilde{F}_{mr} - \tilde{F}_{ns}\|_2 / Q^{1/2}.$$

2. If  $d_{mr,ns}$  is not among the  $k_{nn}$  smallest values of neither  $\{d_{mr,ij}\}_{i,j}$  nor  $\{d_{ns,ij}\}_{i,j}$ , set  $d_{mr,ns} = \infty$ .

- For each  $i \in \{-b, \dots, b\}$ , compute the kernel sum  $\kappa_i = \sum_{m,n=0}^{N-1} \sum_{r,s=0}^{S-1} \bar{K}_{mr,ns}^{(i)} / (NS)^2$ , where

$$\bar{K}_{mr,ns}^{(i)} = \exp(-d_{mr,ns}^2 / \epsilon_i).$$

- Choose  $i \in \{-b + 1, \dots, b - 1\}$  that maximizes

$$\kappa'_i = (\log \kappa_{i+1} - \kappa_{i-1}) / (\log \epsilon_{i+1} - \epsilon_{i-1}).$$

- With  $i$  determined from Step 4, set  $\hat{m} = 2\kappa'_i$  and  $\sigma_{ns} = \sum_{m=0}^{N-1} \sum_{r=0}^{S-1} \bar{K}_{ns,mr}^{(i)}$ .

**Algorithm 4** (VSA eigenfunctions)

- Inputs

- Delay-embedded vectors  $\tilde{F}_{ns}$  from Algorithm 1
- Phase space speeds  $\tilde{\xi}_{ns}$  from Algorithm 2
- Densities  $\sigma_{ns}$  and dimension  $\hat{m}$  from Algorithm 3
- Neighborhood parameter  $k_{nn} \leq NS$
- Candidate bandwidth parameter values  $\{\epsilon_0, \dots, \epsilon_{b-1}\}$  with  $\epsilon_i = 2^i$
- Number  $l \leq N$  of eigenfunctions to compute

- Outputs

- Eigenvalues  $\lambda_0, \lambda_1, \dots, \lambda_{l-1} \leq 1$  and corresponding eigenvectors  $\underline{\phi}_0, \dots, \underline{\phi}_{l-1} \in \mathbb{R}^{NS}$
- Dual eigenvectors  $\underline{\phi}'_0, \dots, \underline{\phi}'_{l-1} \in \mathbb{R}^{NS}$

- Steps

- For every  $n \in \{0, \dots, N - 1\}$ ,  $s \in \{0, \dots, N - 1\}$ , compute the scaling factors  $a_{ns} = (\sigma_{ns} \tilde{\xi}_{ns})^{-1/\hat{m}}$ .
- For every  $m, n \in \{0, \dots, N - 1\}$ ,  $r, s \in \{0, \dots, S - 1\}$ , compute the scaled pairwise distances

$$\tilde{d}_{mr,ns} = a_{mr} a_{ns} \| \tilde{F}_{mr} - \tilde{F}_{ns} \|_2 / Q^{1/2}.$$

- If  $\tilde{d}_{mr,ns}$  is not among the  $k_{nn}$  smallest values of neither  $\{\tilde{d}_{mr,ij}\}_{i,j}$  nor  $\{\tilde{d}_{ns,ij}\}_{i,j}$ , set  $\tilde{d}_{mr,ns} = \infty$ .
- Using the same method as Steps 3 and 4 of Algorithm 3, select the bandwidth parameter  $\epsilon_i \in \{\epsilon_{-b+1}, \dots, \epsilon_{b-1}\}$  for the kernel values  $k_{mr,ns}^{(i)} = \exp(-\tilde{d}_{mn,rs}^2 / \epsilon_i)$ .
- Using  $i$  from Step 4, and for every  $n \in \{0, \dots, N - 1\}$ ,  $s \in \{0, \dots, S - 1\}$ , compute the normalization coefficients

$$v_{ns} = \sum_{m=0}^{N-1} \sum_{r=0}^{S-1} k_{ns,mr}^{(i)}, \quad u_{ns} = \sum_{m=0}^{N-1} \sum_{r=0}^{S-1} k_{ns,mr}^{(i)} / v_{mr}, \quad \hat{w}_{ns} = \sqrt{u_{ns} v_{ns}},$$

$$w_{ns} = \sqrt{u_{ns}/v_{ns}}.$$

6. Using  $i$  from Step 4, form the  $(NS) \times (NS)$  symmetric kernel matrix  $\hat{\mathbf{P}}$  with elements

$$\hat{P}_{mr,ns} = \frac{k_{mr,ns}^{(i)}}{NS \hat{w}_{mr} \hat{w}_{ns}}, \quad m, n \in \{0, \dots, N-1\}, \quad r, s \in \{0, \dots, S-1\}.$$

7. Set  $\lambda_0, \dots, \lambda_{l-1}$  to the  $l$  largest eigenvalues of  $\hat{\mathbf{P}}$ . Compute corresponding eigenvectors  $\hat{\phi}_0, \dots, \hat{\phi}_{l-1}$ , normalized such that  $\|\hat{\phi}_j\|_2 = NS$ .
8. Form the diagonal matrix  $\mathbf{W}$  with  $W_{ns,ns} = w_{ns}$ . For every  $j \in \{0, \dots, l-1\}$ , set  $\underline{\phi}_j = \mathbf{W} \hat{\phi}_j$  and  $\underline{\phi}'_j = \mathbf{W}^{-1} \hat{\phi}_j$ .

### Algorithm 5 (Signal decomposition)

- Inputs

- Spatiotemporal signal  $\tilde{F}_{ns} \in \mathbb{R}$ ,  $n \in \{0, \dots, N-1\}$ ,  $s \in \{0, \dots, S-1\}$
- Eigenvectors  $\underline{\phi}_0, \dots, \underline{\phi}_{l-1}$  and dual eigenvectors  $\underline{\phi}'_0, \dots, \underline{\phi}'_{l-1}$  from Algorithm 4

- Outputs

- Spatiotemporal patterns  $\underline{F}_0, \dots, \underline{F}_{l-1} \in \mathbb{R}^{NS}$

- Steps

1. Assemble the spatiotemporal signal into a column vector  $\underline{F} = [F_{ns}] \in \mathbb{R}^{NS}$ .
2. For each  $j \in \{0, \dots, l-1\}$ , compute the expansion coefficient  $c_j = \underline{\phi}'_j{}^\top \underline{F} / (NS)$ , and set  $\underline{F}_j = c_j \underline{\phi}_j$ .

All numerical experiments in this work were carried out using a MATLAB code, running on a medium-scale Linux cluster. Pairwise distances were computed in brute force, though the moderate dimensionality of the data (equal to the number of delays  $Q$ ) as treated by VSA could potentially enable the efficient use of approximate nearest neighbor algorithms (Arya et al. 1998; Jones et al. 2011), leading to a significant cost reduction in that step. The eigenvalue problem for  $\hat{\mathbf{P}}$  was solved using MATLAB's `eigs` iterative solver, which is based on implicitly restarted Arnoldi methods in the ARPACK library (Lehoucq et al. 1998).

## F Overview of NLSA

In this appendix we summarize the kernel construction and spatiotemporal reconstruction procedure utilized in NLSA. Additional details and pseudocode for this method can be found in Giannakis and Majda (2012) and Giannakis (2017).

First, the NLSA kernel construction parallels closely the VSA construction described in Appendix C, with the difference that in NLSA all kernels and their associated eigenfunctions are defined on state space  $X$ , as opposed to the product space



$\Omega = X \times Y$ . More specifically, NLSA is based on kernels  $k_Q^{(X)} : X \times X \rightarrow \mathbb{R}$  of the form [cf. (3)]

$$k_Q^{(X)}(x, x') = \exp\left(-\frac{a_Q^{(X)}(x)a_Q^{(X)}(x')d_Q^2(x, x')}{\epsilon}\right),$$

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

where  $Q \in \mathbb{N}$  is the number of delays,  $\epsilon$  a positive bandwidth parameter, and  $a_Q : X \rightarrow \mathbb{R}$  a continuous nonnegative scaling function. Among the different choices for that function studied in Giannakis and Majda (2012) and Giannakis (2017), here we work with

$$a_Q^{(X)}(x) = (\sigma_Q^{(X)}(x)\xi_Q^{(X)}(x))^\gamma, \quad \sigma_Q^{(X)}(x) = \int_X e^{-d_Q^2(x, x')/\epsilon} d\mu(x'),$$

$$\xi_Q^{(X)}(x) = \frac{1}{Q} \sum_{q=0}^{Q-1} \|\tilde{V}\vec{F}(\Phi^{-q\tau}(x))\|_{H_Y}^2,$$

where  $\gamma$  is a real parameter. This function has a similar structure as  $a_Q$  in (43), and  $\gamma$  is chosen as described in Appendix C.1.3. Equipped with this kernel, NLSA proceeds by applying diffusion maps normalization to obtain an ergodic Markov kernel  $p_Q^{(X)} : X \times X \rightarrow \mathbb{R}$ , with

$$p_Q^{(X)}(x, x') = \frac{k_Q^{(X)}(x, x')}{u_Q^{(X)}(x)v_Q^{(X)}(x')}, \quad v_Q^{(X)}(x) = \int_X k_Q^{(X)}(x, x') d\mu(x'),$$

$$u_Q^{(X)}(x) = \int_X \frac{k_Q^{(X)}(x, x')}{v_Q^{(X)}(x')} d\mu(x'),$$

and computing the eigenvalues and eigenfunctions  $(\lambda_j^{(X)}, \varphi_j)$  of the corresponding Markov operator  $P_Q^{(X)} : H_X \rightarrow H_X$ . Dual eigenfunctions  $\varphi'_i \in H_X$  satisfying  $\langle \varphi'_i, \varphi_j \rangle_{H_X} = \delta_{ij}$  are also computed analogously to the  $\phi'_i$  in VSA (see Sect. 3.3).

The NLSA eigenfunctions  $\varphi_j$  define temporal patterns of the system (see Sect. 2.2). To construct corresponding spatiotemporal patterns,  $\vec{F}_{\text{NLSA},j} \in H$ , the method employs the reconstruction procedure introduced in SSA (Ghil et al. 2002), suitably modified to account for non-orthogonality of the eigenfunctions. This involves first computing the family of spatial patterns  $\psi_j^{(q)} \in H_Y$  with  $\psi_j^{(q)}(y) = \langle \varphi'_j, U^{q\tau} F_y \rangle_{H_X}$ ,  $q \in \mathbb{Z}$ , and then reconstructing according to the formula

$$\vec{F} \approx \sum_{j=0}^{l-1} \vec{F}_{\text{NLSA},j}, \quad \vec{F}_{\text{NLSA},j} = \frac{1}{Q} \sum_{q=0}^{Q-1} U^{q\tau} \varphi_j \otimes \psi_j^{(-q)}. \tag{56}$$

As  $l \rightarrow \infty$ ,  $\sum_{j=0}^l \vec{F}_{\text{NLSA},j}$  converges to the observation map  $\vec{F}$  in  $H$  norm. Unlike the standard reconstruction approach in (1), the reconstructed patterns  $\vec{F}_{\text{NLSA},j}$  are not necessarily of pure tensor product form, and can span instead up to a  $Q$ -dimensional subspace of  $H$ . Moreover, the  $\vec{F}_{\text{NLSA},j}$  are not necessarily  $H$ -orthogonal. Nevertheless, the fact that the  $\vec{F}_{\text{NLSA},j}$  are derived from the eigenfunctions  $\varphi_j$  associated with a scalar-valued kernel, subjects them to similar shortcomings in the presence of symmetries as those described in Sect. 4.2. In practical applications, the patterns  $\vec{F}_{\text{NLSA},j} \in H$  are replaced by their counterparts in the data-driven Hilbert space  $H_{NS}$  with obvious modifications.

## References

- Ahlers, G., Grossmann, S., Lohse, D.: Heat transfer and large scale dynamics in turbulent Rayleigh–Bénard convection. *Rev. Mod. Phys.* **81**(2), 503–537 (2009). <https://doi.org/10.1103/revmodphys.81.503>
- Arbruster, D., Guckenheimer, J., Holmes, P.: Kuramoto–Sivashinsky dynamics on the center-unstable manifold. *SIAM J. Appl. Math.* **49**(3), 676–691 (1989). <https://doi.org/10.1137/0149039>
- Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching. *J. ACM* **45**, 891–923 (1998). <https://doi.org/10.1145/293347.293348>
- Aubry, N., Guyonnet, R., Lima, R.: Spatiotemporal analysis of complex signals: theory and applications. *J. Stat. Phys.* **64**, 683–739 (1991). <https://doi.org/10.1007/bf01048312>
- Aubry, N., Lian, W.-Y., Titi, E.S.: Preserving symmetries in the proper orthogonal decomposition. *SIAM J. Sci. Comput.* **14**, 483–505 (1993). <https://doi.org/10.1137/0914030>
- Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**, 1373–1396 (2003). <https://doi.org/10.1162/089976603321780317>
- Berry, T., Harlim, J.: Variable bandwidth diffusion kernels. *Appl. Comput. Harmon. Anal.* **40**(1), 68–96 (2016). <https://doi.org/10.1016/j.acha.2015.01.001>
- Berry, T., Sauer, T.: Local kernels and the geometric structure of data. *Appl. Comput. Harmon. Anal.* **40**(3), 439–469 (2016). <https://doi.org/10.1016/j.acha.2015.03.002>
- Berry, T., Cressman, R., Gregurić-Ferenček, Z., Sauer, T.: Time-scale separation from diffusion-mapped delay coordinates. *SIAM J. Appl. Dyn. Syst.* **12**, 618–649 (2013). <https://doi.org/10.1137/12088183x>
- Berry, T., Giannakis, D., Harlim, J.: Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E* **91**, 032915 (2015). <https://doi.org/10.1103/PhysRevE.91.032915>
- Broomhead, D.S., King, G.P.: Extracting qualitative dynamics from experimental data. *Phys. D* **20**(2–3), 217–236 (1986). [https://doi.org/10.1016/0167-2789\(86\)90031-x](https://doi.org/10.1016/0167-2789(86)90031-x)
- Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N.: Chaos as an intermittently forced linear system. *Nat. Commun.* **8**, 19 (2017). <https://doi.org/10.1038/s41467-017-00030-8>
- Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**(2), 490–530 (2005). <https://doi.org/10.1137/040616024>
- Budisić, M., Mohr, R., Mezić, I.: Applied Koopmanism. *Chaos* **22**, 047510 (2012). <https://doi.org/10.1063/1.4772195>
- Caponnetto, A., Micchelli, C.A., Pontil, M., Ying, Y.: Universal multi-task kernels. *J. Mach. Learn. Res.* **9**, 1615–1646 (2008)
- Carmeli, C., De Vito, E., Toigo, A., Umanità, V.: Vector valued reproducing kernel Hilbert spaces and universality. *Anal. Appl.* **08**(1), 19–61 (2010). <https://doi.org/10.1142/s0219530510001503>
- Chatelin, F.: Spectral Approximation of Linear Operators. *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia (2011)
- Chow, S.-N., Lu, K., Sell, G.R.: Smoothness of inertial manifolds. *J. Math. Anal. Appl.* **169**, 283–312 (1992). [https://doi.org/10.1016/0022-247X\(92\)90115-T](https://doi.org/10.1016/0022-247X(92)90115-T)
- Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006). <https://doi.org/10.1016/j.acha.2006.04.006>
- Coifman, R.R., et al.: Geometric diffusions as a tool for harmonic analysis and structure definition on data. *Proc. Natl. Acad. Sci.* **102**(21), 7426–7431 (2005). <https://doi.org/10.1073/pnas.0500334102>

- Coifman, R.R., Shkolnisky, Y., Sigworth, F.J., Singer, A.: Graph Laplacian tomography from unknown random projections. *IEEE Trans. Image Process.* **17**(10), 1891–1899 (2008). <https://doi.org/10.1109/tip.2008.2002305>
- Constantin, P., Foias, C., Nicolaenko, B., Témam, R.: *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations*. Springer, New York (1989). <https://doi.org/10.1007/978-1-4612-3506-4>
- Cross, M.P., Hohenberg, P.C.: Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**(3), 851–1123 (1993). <https://doi.org/10.1103/RevModPhys.65.851>
- Cvitanović, P., Davidchack, R.L., Siminos, E.: On the state space geometry of the Kuramoto–Sivashinsky flow in a periodic domain. *SIAM J. Appl. Dyn. Syst.* **9**(1), 1–33 (2009). <https://doi.org/10.1137/070705623>
- Cvitanovic, P., Artuso, R., Mainieri, R., Tanner, G.: *Chaos: Classical and Quantum*. Niels Bohr Institute, Copenhagen (2016)
- Das, S., Giannakis, D.: Delay-coordinate maps and the spectra of Koopman operators. *J. Stat. Phys.* (2019). <https://doi.org/10.1007/s10955-019-02272-w>
- Dellnitz, M., Junge, O.: On the approximation of complicated dynamical behavior. *SIAM J. Numer. Anal.* **36**, 491 (1999). <https://doi.org/10.1137/S0036142996313002>
- Deyle, E.R., Sugihara, G.: Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* **6**(3), e18295 (2011). <https://doi.org/10.1371/journal.pone.0018295>
- Eisner, T., Farkas, B., Haase, M., Nagel, R.: *Operator Theoretic Aspects of Ergodic Theory*, Volume 272 of Graduate Texts in Mathematics. Springer, Berlin (2015)
- Foias, C., Nicolaenko, B., Sell, G.R., Témam, R.: Inertial manifolds for the Kuramoto–Sivashinsky equation. In: *IMA Preprints Series, Number 279*. University of Minnesota Digital Conservancy (1986). <http://hdl.handle.net/11299/4494>. Accessed 8 May 2019
- Foias, C., Jolly, M.S., Kevrekidis, I.G., Sell, G.R., Titi, E.S.: On the computation of inertial manifolds. *Phys. Lett. A* **131**(7,8), 433–436 (1988). [https://doi.org/10.1016/0375-9601\(88\)90295-2](https://doi.org/10.1016/0375-9601(88)90295-2)
- Froyland, G., Dellnitz, M.: On the isolated spectrum of the Perron–Frobenius operator. *Nonlinearity* **13**(4), 1171–1188 (2000). <https://doi.org/10.1088/0951-7715/13/4/310>
- Fung, R., Hanna, A.M., Vendrell, O., Ramakrishna, S., Seideman, T., Santra, R., Ourmazd, A.: Dynamics from noisy data with extreme timing uncertainty. *Nature* **532**, 471–475 (2016). <https://doi.org/10.1038/nature17627>
- Ghil, M., et al.: Advanced spectral methods for climatic time series. *Rev. Geophys.* **40**, 1003 (2002). <https://doi.org/10.1029/2000rg000092>
- Giannakis, D.: Dynamics-adapted cone kernels. *SIAM J. Appl. Dyn. Syst.* **14**(2), 556–608 (2015). <https://doi.org/10.1137/140954544>
- Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* (2017). <https://doi.org/10.1016/j.acha.2017.09.001>. In press
- Giannakis, D., Majda, A.J.: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.* **109**(7), 2222–2227 (2012). <https://doi.org/10.1073/pnas.1118984109>
- Giannakis, D., Slawinska, J., Zhao, Z.: Spatiotemporal feature extraction with data-driven Koopman operators. *J. Mach. Learn. Res. Proc.* **44**, 103–115 (2015)
- Greene, J.M., Kim, J.S.: The steady states of the Kuramoto–Sivashinsky equation. *Phys. D* **33**, 99–120 (1988). [https://doi.org/10.1016/S0167-2789\(98\)90013-6](https://doi.org/10.1016/S0167-2789(98)90013-6)
- Holmes, P., Lumley, J.L., Berkooz, G.: *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge (1996)
- Jolly, M.S., Kevrekidis, I.G., Titi, E.S.: Approximate inertial manifolds for the Kuramoto–Sivashinsky equation: analysis and computations. *Phys. D* **44**(1–2), 38–60 (1990). [https://doi.org/10.1016/0167-2789\(90\)90046-R](https://doi.org/10.1016/0167-2789(90)90046-R)
- Jolly, M.S., Rosa, R., Temam, R.: Evaluating the dimension of an inertial manifold for the Kuramoto–Sivashinsky equation. *Adv. Differ. Equ.* **5**(1–3), 33–66 (2000)
- Jones, P.W., Osipov, A., Rokhlin, V.: Randomized approximate nearest neighbors algorithm. *Proc. Natl. Acad. Sci.* **108**(38), 15679–15686 (2011). <https://doi.org/10.1073/pnas.1107769108>
- Katok, A., Thouvenot, J.-P.: Spectral properties and combinatorial constructions in ergodic theory (chapter 11). In: Hasselblatt, B., Katok, A. (eds.) *Handbook of Dynamical Systems*, vol. 1B, pp. 649–743. North-Holland, Amsterdam (2006)

- Kevrekidis, I.G., Nicolaenko, B., Scovel, J.C.: Back in the saddle again: a computer-assisted study of the Kuramoto–Sivashinsky equation. *SIAM J. Appl. Math.* **50**(3), 760–790 (1990). <https://doi.org/10.1137/0150045>
- Kuramoto, Y., Tsuzuki, T.: Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. *Progr. Theor. Phys.* **55**(2), 356–369 (1976). <https://doi.org/10.1143/PTP.55.356>
- Lehoucq, R.B., Sorensen, D.C., Yang, C.: *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems With Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia (1998)
- Lian, Z., Liu, P., Lu, K.: SRB measures for a class of partially hyperbolic attractors in Hilbert spaces. *J. Differ. Equ.* **261**, 1532–1603 (2016). <https://doi.org/10.1016/j.jde.2016.04.006>
- Lorenz, E.N.: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963). [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Lu, K., Wang, Q., Young, L.-S.: Strange attractors for periodically forced parabolic equations. *Mem. Am. Math. Soc.* **224**(1054), 1–85 (2013). <https://doi.org/10.1090/S0065-9266-2012-00669-1>
- Marrouch, N., Read, H.L., Slawinska, J., Giannakis, D.: Data-driven spectral decomposition of ECoG signal from an auditory oddball experiment in a marmoset monkey: Implications for EEG data in humans. In: 2018 International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro, Brazil. IEEE (2018). <https://doi.org/10.1109/IJCNN.2018.8489475>
- Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**, 309–325 (2005). <https://doi.org/10.1007/s11071-005-2824-x>
- Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior. *Phys. D.* **197**, 101–133 (2004). <https://doi.org/10.1016/j.physd.2004.06.015>
- Micchelli, C.A., Pontil, M.: On learning vector-valued functions. *Neural Comput.* **17**(1), 177–204 (2005). <https://doi.org/10.1162/0899766052530802>
- Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–716 (1980). <https://doi.org/10.1103/physrevlett.45.712>
- Penland, C.: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Weather Rev.* **117**(10), 2165–2185 (1989)
- Robinson, J.C.: Inertial manifolds for the Kuramoto–Sivashinsky equation. *Phys. Lett. A* **184**(2), 190–193 (1994). [https://doi.org/10.1016/0375-9601\(94\)90775-7](https://doi.org/10.1016/0375-9601(94)90775-7)
- Robinson, J.C.: A topological delay embedding theorem for infinite-dimensional dynamical systems. *Nonlinearity* **18**(5), 2135–2143 (2005). <https://doi.org/10.1088/0951-7715/18/5/013>
- Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S.: Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127 (2009). <https://doi.org/10.1017/s0022112009992059>
- Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. *J. Stat. Phys.* **65**(3–4), 579–616 (1991). <https://doi.org/10.1007/bf01053745>
- Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28 (2010). <https://doi.org/10.1017/S0022112010001217>
- Schölkopf, B., Smola, A., Müller, K.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299–1319 (1998). <https://doi.org/10.1162/089976698300017467>
- Sinai, Y.G. (ed.): *Dynamical Systems, Ergodic Theory and Applications*, Volume 100 of *Encyclopedia of Mathematical Sciences*, 2nd edn. Springer, Berlin (2000)
- Singer, A.: From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmon. Anal.* **21**, 128–134 (2006). <https://doi.org/10.1016/j.acha.2006.03.004>
- Sivashinsky, G.I.: Nonlinear analysis of hydrodynamical instability in laminar flames. Part I. Derivation of basic equations. *Acta Astronaut.* **4**(11), 1177–1206 (1977). [https://doi.org/10.1016/0094-5765\(77\)90096-0](https://doi.org/10.1016/0094-5765(77)90096-0)
- Tajima, S., Greenside, H.S.: Microextensive chaos of a spatially extended system. *Phys. Rev. E* **66**, 017205 (2002). <https://doi.org/10.1103/PhysRevE.66.017205>
- Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.S. (eds.) *Dynamical Systems and Turbulence*, Volume 898 of *Lecture Notes in Mathematics*, pp. 366–381. Springer, Berlin (1981). <https://doi.org/10.1007/bfb0091924>
- Takeuchi, K.A., Yang, H.-L., Ginelli, F., Radons, G., Chaté, H.: Hyperbolic decoupling of tangent space and effective dimension of dissipative systems. *Phys. Rev. E* **84**, 046214 (2011). <https://doi.org/10.1103/PhysRevE.84.046214>
- Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Phys. D* **35**, 395–424 (1989). [https://doi.org/10.1016/0167-2789\(89\)90077-8](https://doi.org/10.1016/0167-2789(89)90077-8)

- von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. *Ann. Stat.* **26**(2), 555–586 (2008). <https://doi.org/10.1214/009053607000000640>
- Wang, X., Giannakis, D., Slawinska, J.: Antarctic circumpolar waves and their seasonality: intrinsic traveling modes and ENSO teleconnections. *Int. J. Climatol.* **39**(2), 1026–1040 (2019). <https://doi.org/10.1002/joc.5860>
- Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* **25**(6), 1307–1346 (2015). <https://doi.org/10.1007/s00332-015-9258-5>
- Yair, O., Talmon, R., Coifman, R.R., Kevrekidis, I.G.: No equations, no parameters, no variables: Data, and the reconstruction of normal forms by learning informed observation geometries. *Proc. Natl. Acad. Sci.* **114**(38), E7865–E7874 (2017). <https://doi.org/10.1073/pnas.1620045114>
- Young, L.-S.: What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.* **108**, 733–754 (2002). <https://doi.org/10.1023/A:1019762724717>
- Zhao, Z., Giannakis, D.: Analog forecasting with dynamics-adapted kernels. *Nonlinearity* **29**, 2888–2939 (2016). <https://doi.org/10.1088/0951-7715/29/9/2888>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Dimitrios Giannakis<sup>1</sup>  · Abbas Ourmazd<sup>2</sup>  · Joanna Slawinska<sup>2</sup> · Zhizhen Zhao<sup>3</sup>

<sup>1</sup> Courant Institute of Mathematical Sciences, New York University, New York, USA

<sup>2</sup> Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, USA

<sup>3</sup> Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, USA