

# Data-driven spectral decomposition and forecasting of ergodic dynamical systems

Dimitrios Giannakis\*

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

---

## Abstract

We develop a framework for dimension reduction, mode decomposition, and nonparametric forecasting of data generated by ergodic dynamical systems. This framework is based on a representation of the Koopman and Perron-Frobenius groups of unitary operators in a smooth orthonormal basis of the  $L^2$  space of the dynamical system, acquired from time-ordered data through the diffusion maps algorithm. Using this representation, we compute Koopman eigenfunctions through a regularized advection-diffusion operator, and employ these eigenfunctions in dimension reduction maps with projectible dynamics and high smoothness for the given observation modality. In systems with pure point spectra, we construct a decomposition of the generator of the Koopman group into mutually commuting vector fields that transform naturally under changes of observation modality, which we reconstruct in data space through a representation of the pushforward map in the Koopman eigenfunction basis. We also establish a correspondence between Koopman operators and Laplace-Beltrami operators constructed from data in Takens delay-coordinate space, and use this correspondence to provide a rigorous interpretation of diffusion-mapped delay coordinates for this class of systems. Moreover, we take advantage of a special property of the Koopman eigenfunction basis, namely that the basis elements evolve as simple harmonic oscillators, to build nonparametric forecast models for arbitrary probability densities and observables. In systems with more complex spectral behavior, including mixing systems, we develop a method inspired from time change in dynamical systems to transform the generator to a new operator with improved spectral properties, and use that operator for vector field decomposition and nonparametric forecasting.

*Keywords:* Koopman operators, Perron-Frobenius operators, dynamic mode decomposition, ergodic dynamical systems, time change, nonparametric forecasting, kernel methods, diffusion maps

---

## 1. Introduction

### 1.1. Background and motivation

In many branches of science and engineering, one is faced with the problems of dimension reduction and forecasting of high-dimensional time series. When these time series are generated by ergodic dynamical systems (or by ergodic components of non-ergodic systems), they possess an important special property, namely that long-time averages are equivalent to expectation values with respect to the invariant measure of the dynamics. This property enables sampling of the full phase space from long time series and sufficiently high-dimensional observables, and data analysis algorithms can be employed to perform dimension reduction and forecasting of these systems with no a priori knowledge of the equations of motion. This nonparametric approach is useful in a variety of contexts, such as when the equations of motion are unknown or partially known, when the system cannot be feasibly simulated, or when the solutions of the equations of motion are complicated and a decomposition into simpler components (modes) is desired.

---

\*Corresponding author

*Email address:* dimitris@cims.nyu.edu (Dimitrios Giannakis)

The vast array of methods developed to address these goals can be broadly categorized as state-space or operator-theoretic methods [1]. A popular state-space approach is to approximate a nonlinear dynamical system by a collection of local linear models on the tangent planes of the attractor [2–4]; other approaches construct global nonlinear models for the dynamical evolution map [5], or nonlinearly project the attractor to lower-dimensional Euclidean spaces and construct reduced models operating in those spaces [6–8]. A common element of these techniques is that the forward operators of the reduced models are defined in state space; i.e., they map the state at a given time to another state in the future. On the other hand, operator-theoretic techniques [1, 9–34, and others] shift attention away from the state-space perspective, and focus instead on the action of dynamical systems on spaces of observables or measures.

Remarkably, the action of a nonlinear dynamical system on appropriately constructed linear spaces of observables and measures can be characterized without approximation by groups (or semigroups) of *linear* operators, known as Koopman [35] or Perron-Frobenius operators, respectively. In particular, Koopman operators (also known as composition operators) act on observables by shifts along the dynamical flow, while Perron-Frobenius operators (also known as Ruelle [36] or transfer operators) act on measures by pullbacks. Typically, the spaces of observables and measures in question are infinite-dimensional, so one can think of a tradeoff between a finite-dimensional nonlinear system and a group of linear operators acting on an infinite-dimensional space. Nevertheless, the intrinsically linear structure of these spaces makes Koopman and Perron-Frobenius operators amenable to treatment through the full machinery of linear operator theory and the associated finite-dimensional approximation schemes (e.g., Galerkin methods). Data-driven operator-theoretic techniques exploit this structure to perform tasks such as spectral analysis of complex systems [9–11], identification of coherent sets (e.g., invariant, almost-invariant, periodic, and almost-periodic sets) [23–32], identification of dynamic modes [12–20], computation of ergodic and harmonic quotients [1, 22], modeling of metastable [33] and slow-fast systems [34], control [21], and other applications. In particular, [1, 22] developed a method for analyzing ergodic and harmonic quotients based on the diffusion maps algorithm [37]; in what follows, we will use diffusion maps to learn smooth orthonormal basis sets of functions from the data, and employ these bases in various Galerkin and spectral methods for dimension reduction, mode decomposition, and nonparametric forecasting based on Koopman and Perron-Frobenius operators.

For appropriately chosen spaces of observables and measures, the Koopman and Perron-Frobenius operators are dual pairs, and therefore provide theoretically equivalent information. For instance, in the setting of an ergodic dynamical system, natural spaces of observables are  $L^2$  spaces of complex-valued scalar functions associated with the system’s invariant probability measures, and natural spaces of measures are complex measures with  $L^2$  densities. Yet, at an applied level, the fact that Koopman and Perron-Frobenius operators act on fundamentally different entities appears to have led to the development of two fairly distinct families of approximation techniques (though the dichotomy is not rigid, and there are references in the literature utilizing results from both the Koopman and Perron-Frobenius perspectives; e.g, [19]).

Historically, data-driven techniques based on Perron-Frobenius operators [23–34] began with the work of Dellnitz and Junge [23] in 1999. A common approach in these techniques is to approximate the spectrum of the Perron-Frobenius operator through a scheme known as Ulam’s method [38]. Essentially, this involves partitioning the state space into a finite collection of disjoint subsets, and estimating the transition probabilities between these subsets by counting the corresponding transitions in a large ensemble of simulations or experiments. The resulting transition probability matrix can be interpreted as a Galerkin projection of a smoothed (compact) transfer operator for the system perturbed by a small amount of noise, and the eigenvectors of that matrix at eigenvalues  $\lambda$  on or near the unit circle are used to identify coherent sets. In subsequent work, Dellnitz, Froyland, et al. [24–28] studied certain classes of systems possessing quasi-compact Perron-Frobenius operators, where it was rigorously established that Ulam’s method is able to provide accurate approximations of isolated Perron-Frobenius eigenvalues and their corresponding eigenfunctions.

Data-driven techniques based on Koopman operators were first developed by Mezić and Banaszuk [9] and Mezić [10] in 2004 and 2005, respectively. In particular, [10] established the Koopman mode expansion for measure-preserving systems and  $L^2$  observables. In this setting, the Koopman and Perron-Frobenius operators are unitary and adjoint to one another, and have a well-defined spectral expansion consisting in general of both discrete (pure point) and continuous parts. In [9, 10], a method for estimating the

point spectrum of Koopman operators was developed using generalized Laplace analysis. This approach involves computing harmonic averages (Fourier transforms) of time series of observables for different initial conditions, and identifying the frequencies leading to non-vanishing averages. For those frequencies, the harmonic average corresponds to a projection of the observable onto an eigenspace of the Koopman operator, allowing one to identify Koopman eigenvalues and their corresponding periodic eigenfunctions, as well as spatial patterns called Koopman modes.

In later work, Rowley et al. [13] established that the Koopman mode expansion has close connections with the dynamic mode decomposition (DMD) technique of Schmid and Sesterhenn [12] and Schmid [15] for decomposing time-ordered spatiotemporal datasets into so-called dynamic modes. Similarly to the proper orthogonal decomposition (POD) [39], DMD extracts modes by solving an eigenvalue problem for a matrix  $A$  constructed from experimental snapshots, but instead of the covariance matrix used in POD, DMD employs a time-shifted cross-correlation matrix capturing the linear dependence of the snapshots at the next time step on the snapshots at the current time step. To address the computational cost associated with the high dimensionality of typical spatiotemporal data, the DMD algorithm in [13] employs an Arnoldi-type iterative method which does not require explicit formation of  $A$ . In particular, it was shown that the eigenvalues and eigenvectors obtained via this algorithm approximate the eigenvalues and modes in the Koopman mode expansion in [10]. Tu et al. [17] subsequently developed an alternative formulation of DMD based on matrix pseudoinverses (or truncated pseudoinverses) that does not require the data to be in the form of a single time series, and moreover returns estimates of the Koopman eigenfunctions (in addition to the eigenvalues and Koopman modes).

The connections between the Koopman mode expansion and DMD were further studied by Williams et al. [19], who developed an extended DMD (EDMD) framework. In EDMD, the eigenvalue problem for the Koopman operator is projected to a finite-dimensional matrix eigenvalue problem associated with the action of the Koopman operator on a general dictionary of observables. In particular, standard DMD can be viewed as a particular instance of EDMD for the dictionary formed by the components of the state vector, but the general approach advocated in EDMD is to employ richer dictionaries than those formed by the state vector components in order to improve the accuracy of the computed eigenvalues and Koopman modes. The dictionaries proposed in [19] include Hermite polynomials in ambient data space, radial basis functions, and discontinuous spectral elements analogous to the bases used in Ulam approximations of Perron-Frobenius operators. When these dictionaries are sufficiently rich, EDMD converges to a Galerkin method for the Koopman eigenvalue problem, although issues related to non-compactness and potentially continuous spectrum of the Koopman operator were not addressed in [19].

Irrespective of the method employed to compute them, the triplets of eigenvalues, eigenfunctions, and Koopman modes produce a decomposition a complex spatiotemporal signal into simpler “building blocks”, which are intrinsic to the dynamical system generating the data. In particular, the Koopman eigenvalues and eigenfunctions are independent of observation modality, meaning that data acquired from different sensors would yield the same results for the eigenvalues and eigenfunctions, so long as the data contain sufficiently rich information. In contrast, the results of POD-type techniques and many nonlinear dimension reduction techniques [40] depend on the observation modality. Of course, the efficacy of the Koopman and Perron-Frobenius methods in real-world applications depends strongly on a number of factors, including the properties of the underlying dynamical system, the discretization scheme, and the measurement and data acquisition apparatus. To motivate our work, we end this section by summarizing some of the challenges encountered in data-driven operator-theoretic techniques for dynamical systems.

First, at a fundamental level, for systems of high complexity the spectra of the Koopman and Perron-Frobenius operators are not amenable to eigendecomposition and can be highly exotic. For instance, a necessary and sufficient condition for a dynamical system to be weak-mixing (a weak form of chaos, which implies ergodicity) is that the Koopman operators have no  $L^2$  eigenfunctions other than the constant function [41, 42]. While such cases with continuous spectra can be theoretically handled through the Koopman mode expansion [10], in numerical implementations involving matrix algebra one always obtains eigenvectors, whose properties become difficult to analyze. These observations suggests that numerical methods for Koopman eigenfunctions must involve, either implicitly or explicitly, some form of regularization or annealing. For instance, in dictionary or Galerkin discretizations regularization is implicitly provided through

projection on a finite-dimensional solution space. In fact, regularization is warranted even in simple cases involving systems with pure point spectra and complete bases of Koopman eigenfunctions. For example, in irrational flows on tori (arguably, among the simplest ergodic dynamical systems) and at dimension greater than 1, the frequencies associated with the Koopman eigenvalues are dense on the real line, and this can lead to poorly conditioned numerical schemes unless care is taken to filter out highly oscillatory solutions.

A second important issue concerns the choice of dictionary or basis employed for operator approximation. This issue becomes especially pertinent in situations where the ambient space dimension is high, and/or the data lies on an a priori unknown subset of ambient data space of small or zero measure. As a concrete example, consider again an irrational flow on an  $m$ -torus embedded in  $\mathbb{R}^n$ . In this case, computing Koopman eigenfunctions in a smooth function basis for  $\mathbb{R}^n$  (e.g., Hermite polynomials) can be problematic due to computational cost (which increases exponentially with  $n$  independently of the intrinsic dimension  $m$ ), but more importantly due to the fact that the eigenfunctions are supported on a set of Lebesgue measure zero, and therefore cannot be well represented in a smooth basis for functions on  $\mathbb{R}^n$ . In such situations (which are quite prevalent in real-world applications) it is preferable to work in a basis which is adapted to the intrinsic geometry of the data.

A third issue, somewhat different in nature to the previous two, concerns the identification of spatial modes. Since the Koopman and Perron-Frobenius eigenfunctions are intrinsic to the dynamical system generating the data, one would like that the associated modes are also intrinsic; i.e., that they transform naturally as vector-valued functions under changes of observation modality. However, in the standard DMD and EDMD construction, the Koopman modes are given by global averages of the data weighted by the corresponding Koopman eigenfunctions, and such averages do not transform naturally under general nonlinear transformations of the data.

## 1.2. Contributions of this work

In this paper, we develop operator-theoretic techniques for dimension reduction and mode decomposition of data generated by ergodic dynamical systems. Building on these methods, we also develop nonparametric techniques for forecasting observables and probability measures. A key ingredient of our approach is a data-driven orthonormal basis for the  $L^2$  space of the dynamical system constructed using the diffusion maps algorithm [37] in conjunction with variable-bandwidth kernels [43]. The construction of this basis follows the approach of Berry et al. [44] for approximating Kolmogorov and Fokker-Planck operators (the stochastic analogs of the Koopman and Perron-Frobenius operators, respectively) of stochastic dynamical systems on manifolds. Here, we employ this basis in a family of spectral and Galerkin methods for the Koopman eigenvalue problem, reconstruction of vector fields (which can be thought of as a generalization of the notion of Koopman modes), and nonparametric forecasting. Ergodicity enables the implementation of these techniques from a single time series of measurements of the state vector without requiring (potentially costly or infeasible) ensembles of experiments with different initial conditions. The techniques are also applicable in the case of partial observations, so long as the observation map is such that one of the variants of Takens' delay embedding theorem applies [45–49]. We demonstrate our approach through analytical and numerical applications to ergodic dynamical systems on the 2- and 3-torus, including systems with mixing [50] and fixed points [51].

The data-driven basis from diffusion maps has several attractive properties that help address the challenges in operator-theoretic techniques outlined in Section 1.1:

1. The basis functions are naturally supported on the subset of ambient data space on which the system evolves (e.g., a low-dimensional attractor), and their computation requires no a priori knowledge of that subset. In particular, under suitable geometrical smoothness assumptions, the diffusion maps basis converges in the limit of large data to a complete orthonormal basis of the  $L^2$  space of the dynamical system. By orthogonality of the basis, passing from a Koopman to a Perron-Frobenius operator and vice versa can be accomplished by a standard matrix transpose.
2. The cost for computing the basis functions scales linearly with the ambient space dimension  $d$ , and after the basis has been computed, the cost of the Koopman and Perron-Frobenius approximation is independent of  $d$ . This property is particularly important in high-dimensional applications such as analysis of engineering and geophysical fluid flows (e.g., [52]).

3. Associated with the diffusion eigenfunctions is a Dirichlet energy (given by the corresponding eigenvalues), which measures the roughness of the eigenfunctions in a Riemannian geometry that depends on the kernel. Here, the kernel is constructed explicitly so that the Riemannian metric has compatible volume form with the invariant measure of the dynamics. As a result, our Galerkin approximation spaces for the Koopman eigenvalue problem are spanned by orthonormal functions with minimal expected roughness (with respect to the invariant measure).
4. Through an eigenvalue-dependent normalization, the diffusion eigenfunctions form a natural basis for the  $H_1$  Sobolev space associated with the invariant measure. In this basis, isotropic diffusion operators are represented by identity matrices, and therefore regularization can be carried out in a computationally efficient and well-conditioned manner.
5. Through the use of kernels in Takens delay-coordinate space [53, 54], the basis functions from diffusion maps can be made to converge to Koopman eigenfunctions in pure point spectrum systems, leading to efficient and noise-robust Galerkin schemes.
6. The variable-bandwidth family of kernels provides a flexible framework to perform a change of measure which is equivalent to a time change [55] in the dynamical system—we use such a change of measure as a regularization tool in certain classes of systems with continuous Koopman spectra.

In what follows, we first develop our approach in the case of systems with pure point spectra, where the Koopman eigenfunctions form a complete orthonormal basis of the  $L^2$  space associated with the invariant measure. We formulate a Galerkin method for the Koopman eigenvalue problem in the appropriate  $H_1$  space on the phase space manifold—this is important in systems with dense sets of eigenfrequencies (including pure point spectrum systems), as it eliminates highly oscillatory eigenfunctions from the spectrum through Tikhonov regularization. Using the Dirichlet energies available from diffusion maps, we select the least rough set of Koopman eigenfunctions with corresponding rationally independent eigenvalues, and take advantage of a group structure of the Koopman eigenfunctions to generate a complete basis of the  $L^2$  space on the manifold recursively from the group generators. This basis is employed to decompose the vector field of the dynamics into a set of linearly-independent, nowhere-vanishing, mutually commuting vector fields (i.e., a set of dynamically independent components), and to represent the pushforward map carrying along these vector fields in data space. The reconstructed vector fields are spatial patterns analogous to Koopman modes, which transform naturally (as tensors) under changes of observation modality. The generating eigenfunctions are also used to construct factor maps [9, 10] from the phase space manifold to the complex plane, under which the dynamical vector field is projectible, and the dynamics in the image spaces are simple harmonic oscillations. We show that through the use of delay-coordinate maps, the Laplace-Beltrami operator approximated by diffusion maps in the limit of infinitely many delays commutes with the generator of the Koopman group; as a result, the two operators have common eigenfunctions. This result bridges these two important families of dimension reduction and mode decomposition techniques, and naturally leads to efficient Galerkin schemes for the Koopman eigenvalue problem with high robustness to i.i.d. measurement noise.

Besides dimension reduction and mode decomposition, we utilize the Koopman eigenvalues and eigenfunctions in nonparametric forecasting schemes for probability measures and observables. This approach is closely related to the nonparametric forecasting method for stochastic systems developed in [44], with the difference that we advance probability densities using the simple harmonic oscillator structure of the Koopman eigenfunctions, as opposed to taking powers of a finite-dimensional Fokker-Planck operator. In particular, Koopman eigenfunctions of pure point spectrum systems evolve as uncoupled oscillators with frequency given by the corresponding eigenvalue, and by completeness, these eigenfunctions can be used to predict the time evolution of arbitrary observables in  $L^2$  and probability measures with  $L^2$  densities.

Next, we consider systems whose Koopman operators do not have pure point spectra, including a class of weak mixing systems with no nonconstant eigenfunctions. There, a regularization technique inspired from the theory of time change in dynamical systems [55] is developed which attempts to construct a new dynamical system having the same orbits as the original one, but with improved spectral properties for eigendecomposition and forecasting. In particular, we perform a time-change transformation using the norm of the dynamical vector field in the ambient data space as the time-change function. This technique can be

applied to arbitrary dynamical systems as it requires no a priori knowledge of the governing equations, and in special cases it recovers a pure point system with the same orbits as the mixing system generating the data. We generalize our vector field decomposition and nonparametric forecasting techniques to the time-changed framework, where the recovered vector field components become non-commuting, and the simple harmonic oscillators acquire couplings and time-dependent frequencies.

This paper is organized as follows. In Section 2, we lay out our notation and summarize basic results from ergodic theory. In Section 3, we formulate our dimension reduction, mode decomposition, and forecasting techniques for systems with pure point spectra. We discuss the numerical implementation of these techniques, including the Galerkin method with diffusion regularization for the Koopman eigenvalue problem, in Section 4, where we also present numerical applications to variable-speed flows on the 2-torus. In Section 5, we extend this scheme to delay-coordinate mapped data, and establish the correspondence between the Laplace-Beltrami and Koopman operators arising in the limit of infinitely many delays along with methods for removing i.i.d. observational noise. In Section 6, we formulate our time-change approach for improving the spectral properties of the Koopman group, and present applications to mixing dynamical systems on the 3-torus and fixed-point ergodic flows on the 2-torus. We conclude in Section 7. Appendices A–C contain auxiliary material and technical results, including high-level listings of the algorithms developed in the main text, and an asymptotic analysis of denoising using diffusion maps in conjunction with delay-coordinate maps. Videos showing the temporal evolution of probability densities in the applications of Section 4 and 6 are included as supporting online material (SOM). Also included as SOM is a comparison of our Koopman eigenfunction results to Koopman eigenfunctions computed via EDMD. An application of the techniques developed here to atmospheric dynamics can be found in [52].

## 2. Notation and basic results from ergodic theory

Let  $M$  be a closed, connected, compact, orientable, smooth (of class  $C^\infty$ ),  $m$ -dimensional manifold with a Borel  $\sigma$ -algebra  $\mathcal{B}$ . In what follows,  $M$  will be the phase space of a continuous-time dynamical system with a smooth, invertible evolution map  $\Phi_t : M \mapsto M$ ,  $t \in \mathbb{R}$ . The evolution map induces a map  $\Phi_{t*}$  on measures on  $\mathcal{B}$  such that for a measure  $\mu$  and  $A \in \mathcal{B}$ ,  $\Phi_{t*}\mu(A) = \mu(\Phi_t^{-1}A) = \mu(\Phi_{-t}A)$ . We will take  $\mu$  to be a  $\Phi_t$ -invariant probability measure (i.e.,  $\mu(M) = 1$  and  $\Phi_{t*}\mu = \mu$ ), characterized by a smooth, positive density relative to local coordinates. We consider that the dynamical system is observed at a fixed sampling interval  $T$  through a smooth, vector-valued observation function  $F : M \mapsto \mathbb{R}^d$ . That is, we have a dataset consisting of time-ordered samples  $\{x_0, x_1, \dots, x_{N-1}\}$  in the  $d$ -dimensional data space, with

$$x_i = F(a_i), \quad a_i = \Phi_{t_i}a_0, \quad t_i = T(i-1). \quad (1)$$

Without loss of generality, we assume that  $F$  is an embedding of  $M$  into  $\mathbb{R}^d$ , i.e., it is one-to-one and its derivative,  $F_*|_a : T_aM \mapsto T_{F(a)}\mathbb{R}^d$ , has full rank at every  $a \in M$ . If  $F$  is not an embedding, then with high probability an embedding can be constructed from  $F$  using delay-coordinate maps [45–49], and the methods described below can be applied in that space (e.g., [53, 54]). The embedding of  $M$  into data space induces a smooth Riemannian metric tensor  $g$  on  $M$  given by pulling back the canonical inner product of  $\mathbb{R}^d$ , i.e.,  $g(u_1, u_2) = F_*u_1 \cdot F_*u_2$  for any two tangent vectors  $u_1, u_2 \in T_aM$ . We denote the Riemannian volume form associated with  $g$  and the corresponding Riemannian measure by  $d\text{vol}_g$  and  $\text{vol}_g$ , respectively. From our point of view  $F$ ,  $g$ , and  $\text{vol}_g$  are extrinsic objects to the dynamical system and we would like the results of our dimension reduction and mode decomposition methods to not depend on them.

Next, we introduce the Koopman and Perron-Frobenius groups of operators and their generator which will play a central role in our study. We refer the reader to one of the many books and reviews on ergodic theory (e.g., [56, 57]) for more detailed expositions on these and other related topics. Let  $L^2(M, \mu)$  be the Hilbert space of square-integrable, complex valued functions on  $M$  with inner product  $\langle f_1, f_2 \rangle = \int_M f_1^* f_2 d\mu$  and the corresponding norm  $\|f\| = \langle f, f \rangle^{1/2}$ . In general,  $L^2(M, \mu)$  is different from the Hilbert space  $L^2(M, \text{vol}_g)$  associated with the Riemannian inner product  $\langle f_1, f_2 \rangle_g = \int_M f_1^* f_2 d\text{vol}_g$ , but because  $M$  is compact the two spaces are isomorphic. The Koopman operator  $U_t : L^2(M, \mu) \mapsto L^2(M, \mu)$  for time  $t \in \mathbb{R}$  is defined via  $U_t f = f \circ \Phi_t$ ; that is, the function  $f_t = U_t f$  evaluated at  $a \in M$  is equal to  $f$  evaluated at

the shifted point  $a_t = \Phi_t a$  along the dynamics,  $f_t(a) = f(a_t)$ . The set of Koopman operators  $\{U_t\}$  forms an Abelian group under composition of operators,  $U_{t_1}U_{t_2} = U_{t_1+t_2}$ , with  $U_0 = I$  acting as the identity element of the group. These operators are unitary on  $L^2(M, \mu)$  with  $U_t^* = U_t^{-1} = U_{-t}$ , and therefore they induce an isometry,  $\|U_t f\| = \|f\|$ . The generator of the group,  $v = dU_t/dt|_{t=0}$ , is a skew-symmetric operator on  $L^2(M, \mu)$  with  $v^* = -v$ , and one of its important properties is the Leibniz rule,

$$v(f_1 f_2) = v(f_1) f_2 + f_1 v(f_2). \quad (2)$$

Therefore,  $v$  is a smooth vector field in  $TM$  giving the directional derivative of functions along the dynamical flow,  $v(f)(a) = \lim_{t \rightarrow 0} (f(\Phi_t a) - f(a))/t$ . Because  $\Phi_t$  preserves  $\mu$ ,  $v$  has the Liouville property,  $\operatorname{div}_\mu v = 0$ , i.e., it generates an incompressible flow on  $M$  with respect to the invariant measure. Let now  $\mathcal{M}$  be the space of complex-valued measures on  $\mathcal{B}$  with densities in  $L^2(M, \mu)$ . The Perron-Frobenius operator,  $\tilde{U}_t : L^2(M, \mu) \mapsto L^2(M, \mu)$  characterizes the action of the flow on the densities of measures in  $\mathcal{M}$ ; that is, given  $\nu \in \mathcal{M}$  with  $d\nu/d\mu = \rho$ , then  $\rho_t = \tilde{U}_t \rho$  is equal to the density  $d(\Phi_{t*} \nu)/d\mu$ .

In the Hilbert space setting of interest here the Koopman and Perron-Frobenius operators form an adjoint pair,  $\tilde{U}_t = U_t^*$ , so we can characterize both through the generating vector field  $v$ . In particular,  $v$  maps to a vector field  $V = F_* v$  in the tangent data space,  $T\mathbb{R}^d$ , and because  $T\mathbb{R}^d$  is canonically isomorphic to  $\mathbb{R}^d$ , we can intuitively think of  $V$  as a collection of Euclidean vectors (“arrows”) tangent to the data manifold  $F(M)$  along the direction of the dynamical flow. Note that  $V$  can be approximated at  $O(T^p)$  accuracy by taking finite differences of the time-ordered snapshots  $\{x_i\}$  [58]; e.g.,  $(x_{i+1} - x_{i-1})/(2T)$  is an  $O(T^2)$  approximation of  $V$  at  $x_i$ . In Section 3.2, we will represent the pushforward map  $F_*$  in a smooth orthonormal basis of  $L^2(M, \mu)$ , allowing us to reconstruct arbitrary vector fields on  $M$ , and in particular, the vector field decomposition of the generator.

**Remark 1 (Continuous vs. discrete time).** In practical applications, the data is acquired at the finite sampling interval  $T$ , and we can consider the discrete-time system  $(M, \mathcal{B}, \mu, \hat{\Phi}_n)$ , where  $\hat{\Phi}_n = \Phi_{nT}$  and  $n \in \mathbb{Z}$ . In the discrete case, the Koopman group  $\{\hat{U}_n\}$  is generated by  $U_T$  so that  $\hat{U}_n = U_T^n$ . While some of the constructions discussed below can also be made using  $U_T$  instead of the vector field  $v$  (replacing the eigenvalue  $\lambda$  in (3) ahead by  $e^{\lambda T}$ ), it is generally more natural to carry out our analysis in the continuous-time setting, and use finite differences for numerical implementation. In particular, the discrete-time generator does not have the Leibniz property of the vector field, and cannot be reconstructed in data space via the pushforward map. Nevertheless, discrete analogs of these properties can be constructed using finite-difference approximations of  $v(f)$ , as described in Section 4.3. The diffusion regularization and time-change techniques of Sections 4 and 6 are also more naturally formulated in the continuous-time setting and implemented with finite-difference approximations.

Consider now the eigenvalue equation for the generator,

$$v(z) = \lambda z, \quad \lambda \in \mathbb{C}, \quad z \in L^2(M, \mu). \quad (3)$$

Because  $v$  is skew-symmetric on  $L^2(M, \mu)$ , for any solution to (3) we have  $\lambda = i\omega$ , where  $\omega$  is a real number that corresponds to an intrinsic (i.e., independent of observation modality) frequency of the dynamical system. Thus, the set of eigenvalues lies on the imaginary line, and if  $\lambda$  is the eigenvalue corresponding to the eigenfunction  $z$  then  $\lambda^* = -\lambda$  is also an eigenvalue corresponding to the eigenfunction  $z^*$ . It also follows from the skew-symmetry of  $v$  that eigenfunctions corresponding to distinct eigenvalues are orthogonal on  $L^2(M, \mu)$ . Moreover, by the Leibniz property in (2) (or by unitarity of the  $U_t$ ), it follows that the eigenvalues and eigenfunctions of  $v$  have a group structure under addition and multiplication, respectively. Specifically, if  $z_1$  and  $z_2$  are eigenfunctions corresponding to the eigenvalues  $\lambda_1$  and  $\lambda_2$ , respectively, then  $z_1 z_2$  is also an eigenfunction corresponding to the eigenvalue  $\lambda_1 + \lambda_2$ . Because  $v$  is a directional derivative operator annihilating constant functions, (3) always has the solution  $\lambda = 0$ ,  $z = 1$ , where 1 denotes here a constant function equal to one  $\mu$ -a.e. When other solutions to (3) are present (which, as stated in the Introduction, is not necessarily the case), this group structure is highly beneficial for mode decomposition and forecasting using  $\lambda$  and  $z$ . Note that the spectra of the diffusion operators used traditionally for geometrical analysis of

data do not have this group structure, but (on compact manifolds) these operators are compact and hence guaranteed to have pure point spectra with the associated complete orthonormal eigenfunctions.

Hereafter, we will assume that the continuous-time dynamical system under study is ergodic, and that the discrete-time system for the given sampling interval is also ergodic. It then follows by Birkhoff's pointwise ergodic theorem that for  $\mu$ -a.e.  $a \in M$  and for every  $f \in L^2(M, \mu)$  the temporal means,  $\bar{f}_t(a) = \int_0^t f(\Phi_\tau a) d\tau/t$  and  $\bar{f}_N(a) = \sum_{i=0}^{N-1} f(\hat{\Phi}_i a)/N$  in the continuous- and discrete-time cases, respectively, both converge in  $L^2(M, \mu)$  to the ergodic average  $\int_M f d\mu$ . Under these assumptions, inner products on  $L^2(M, \mu)$  and other related Sobolev spaces can be approximated by Monte Carlo sums of sampled time series, i.e., given the time series  $\{f_{1i}\}_{i=0}^{N-1}$  and  $\{f_{2i}\}_{i=0}^{n-1}$  with  $f_{ji} = f_j(\hat{\Phi}_i a)$  and  $f_j \in L^2(M, \mu)$ , then for  $\mu$ -a.e.  $a \in M$  we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} f_{1i}^* f_{2i} = \int_M f_1^* f_2 d\mu = \langle f_1, f_2 \rangle. \quad (4)$$

This property makes the data-driven techniques in Sections 4 and 6 ahead and in [44] possible. Note that if the dynamical system is not ergodic but preserves  $\mu$ , then the Monte Carlo sum in (4) will converge to the  $L^2$  inner product over the ergodic component associated with the point  $a$ , and the techniques developed below can be applied for that component provided that the assumptions laid out in the beginning of Section 2 are satisfied. An important spectral implication of ergodicity is that the eigenfunctions of  $v$  have  $\mu$ -a.e. constant modulus, i.e., they can be normalized to lie on the unit circle  $|z| = 1$  on the complex plane. This is because  $|z|^2 = z^* z$  is an eigenfunction with eigenvalue 0 (due to the group structure of the eigenvalues and eigenfunctions) and therefore a  $\mu$ -a.e. constant function.

It follows from the spectral theorem for unitary operators on Hilbert spaces [55] that in mixing systems  $v$  has continuous spectrum (but no residual spectrum). This observation, as well as the existence of many known ergodic dynamical systems with exotic spectral behavior (e.g., [41, 55]) raises concerns about the suitability of data-driven eigendecomposition techniques involving the generator in systems of high complexity. In Section 6, we will present a regularization scheme based on the theory of time change in dynamical systems that attempts to transform  $v$  to a generator of a dynamical system which is more amenable to eigendecomposition via numerical methods. For now, however, we restrict attention to systems with pure point spectra where the spectral properties of  $v$  are "optimal" for both dimension reduction and nonparametric forecasting.

**Definition 2 (Pure point spectrum).** A dynamical system  $(M, \mathcal{B}, \mu, \Phi_t)$  is said to have pure point spectrum if there exists an orthonormal basis of  $L^2(M, \mu)$  consisting of eigenfunctions of the generator. We say that the spectrum is generated by  $l$  basic frequencies if there exist  $l$  rationally independent real numbers  $\{\Omega_i\}_{i=1}^l$  such that the eigenvalues in (3) can be expressed as

$$\lambda_k = i \sum_{i=1}^l k_i \Omega_i, \quad k = (k_1, k_2, \dots, k_l) \in \mathbb{Z}^l.$$

Denoting the eigenfunction corresponding to the eigenvalue  $i\Omega_i$  by  $\zeta_i$ , where we take  $\|\zeta_i\| = 1$  by convention, the eigenfunction corresponding to eigenvalue  $\lambda_k$  is given by  $z_k = \prod_{j=1}^l \zeta_j^{k_j}$ .

In systems with pure point spectra, every observable  $f \in L^2(M, \mu)$  has the Fourier expansion

$$f = \sum_{k_1, k_2, \dots, k_l = -\infty}^{\infty} \hat{f}_{k_1 k_2 \dots k_l} \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_l^{k_l}, \quad \text{with } \hat{f}_{k_1 k_2 \dots k_l} = \langle \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_l^{k_l}, f \rangle.$$

Whenever convenient, we will use the shorthand notation  $f = \sum_k \hat{f}_k z_k$  with  $k = (k_1, \dots, k_l)$ . Note that by virtue of the group structure of the spectrum, one needs to store only the  $l$  generating eigenfunctions and the corresponding basic frequencies; the rest of the basis can be generated recursively whenever needed. The group structure of the eigenfunctions also leads to a useful convolution relationship giving the expansion

coefficients of products of functions (which we do not quote here in the interest of brevity). Moreover, applying the decomposition componentwise to the vector-valued observation map  $F$  we have

$$F = \sum_k \hat{F}_k z_k, \quad \hat{F}_k = \langle z_k, F \rangle, \quad (5)$$

where the Fourier coefficients are now spatial patterns in  $\mathbb{R}^d$ . In particular, these patterns are the Koopman modes introduced in [9, 10]. In Section 3.2, we will see that Koopman modes also arise naturally in the spectral representation of the pushforward map  $F_*$  for vector fields.

A classical result in ergodic theory [42] states that ergodic dynamical systems with pure point spectra are measure-theoretically isomorphic to translations on compact Abelian groups equipped with the Haar measure. In particular, for pure point systems realized through diffeomorphisms of smooth manifolds (which is the case studied here), and possessing  $m$  basic frequencies and the associated smooth eigenfunctions, the phase space manifold must necessarily be the  $m$ -torus,  $\mathbb{T}^m$ . Pure-point systems can be constructed from diffeomorphisms of more general smooth manifolds, but these systems have discontinuous eigenfunctions [59]. Throughout this paper, the case  $M = \mathbb{T}^m$  will be the canonical setting where we develop our dimension reduction and forecasting strategies for pure point spectrum systems, as well as for systems with more general spectral properties.

**Remark 3 (Density of the spectrum).** The set of frequencies  $\{\sum_{i=1}^m k_i \Omega_i \mid k_1, \dots, k_m \in \mathbb{Z}\}$  is countable and therefore has zero Lebesgue measure on  $\mathbb{R}$ . However, for  $m = \dim M \geq 2$ , and by rational independence of the  $\Omega_i$ , the set of frequencies is dense in  $\mathbb{R}$ . Moreover, the basic frequencies  $\{\Omega_i\}$  are non-unique as linear combinations of the  $\Omega_i$  with nonzero integer coefficients are also rationally independent. This fact raises the questions of how to select appropriate generators of the spectrum, and how to ensure the good conditioning of approximate generators obtained via finite-dimensional numerical algorithms. We will address these issues in Section 3.1 and 4.3.

### 3. Dimension reduction and forecasting in systems with pure point spectra

#### 3.1. Intrinsic dimension reduction coordinates

Our approach for systems with pure point spectra is to construct a family of nonlinear projection maps  $\{\pi_i\}_{i=1}^m$  of the phase space manifold  $M$  (or, equivalently, its diffeomorphic copy  $F(M)$  in data space) to the complex plane, using the generating eigenfunctions  $\{\zeta_i\}_{i=1}^m$  from Definition 2 as dimension-reduction coordinates. Specifically, we set

$$\pi_i : M \mapsto \mathbb{C} \quad \text{with} \quad \pi_i(a) = \zeta_i(a), \quad (6)$$

where, by ergodicity, the image  $\pi_i(M)$  lies on the unit circle  $\mathbb{T} \subset \mathbb{C}$ . Below, we will see that  $\pi_i$  is actually a factor map that projects the dynamical system on  $M$  to a rotation on  $\mathbb{T}$  with frequency  $\Omega_i$ . The use of such factor maps for dimension reduction was originally proposed in [9, 10].

The projection maps in (6) follow the widely adopted paradigm of applied harmonic analysis and machine learning, which is to perform dimension reduction of data on nonlinear manifolds using eigenfunctions of linear operators on function spaces on these manifolds [37, 60–65]. While these methods are typically based on eigenfunctions of diffusion operators or heat kernels, in this case we use eigenfunctions of a skew-symmetric operator intrinsic to the dynamical system generating the data, namely the generator of the Koopman group, with a small amount of diffusion added for annealing. We will formulate algorithms for approximating these eigenfunctions from data in Section 4.3. For now, we discuss the main features of dimension reduction with eigenfunctions of the generator of the Koopman group, which follow from basic results of ergodic theory and maps of manifolds. In what follows, we use the notation  $\{\theta^1, \dots, \theta^m\}$  for general local coordinates defined in a neighborhood of  $a \in M$ . Moreover, we denote the corresponding coordinate basis vectors of  $T_a M$  and their duals by  $\{\frac{\partial}{\partial \theta^1}, \dots, \frac{\partial}{\partial \theta^m}\}$  and  $\{d\theta^1, \dots, d\theta^m\}$ , respectively, where  $d\theta^\mu(\frac{\partial}{\partial \theta^\nu}) = \delta^\mu_\nu$ . In this coordinate system, we have the expansion  $v = \sum_{\mu=1}^m v^\mu \frac{\partial}{\partial \theta^\mu}$  for the generator, and  $v(f) = \sum_{\mu=1}^m v^\mu \frac{\partial f}{\partial \theta^\mu}$ .

First, the eigenfunctions of  $v$  are intrinsic to the dynamical system  $(M, \mathcal{B}, \mu, \Phi_t)$ , and in particular do not depend on the observation map  $F$  and its associated Riemannian metric and volume form. As stated in Remark 3, the generating eigenfunctions and the corresponding basic frequencies are non-unique, but given any two observation maps it is possible to find generating eigenfunctions such that the corresponding projection maps from (6) are consistent. This means that the  $\pi_i$  provide a unified space to parameterize data acquired from different sensors (different choices of  $F$ ), but generated by the same dynamical system. This property has several data analysis applications (which we do not study here), such as fusion and inference of data acquired from different sensors and from potentially distinct states in  $M$ .

In the setting of a single observation modality, we choose the generating eigenfunctions with the least oscillatory behavior in the Riemannian metric  $g$  associated with the observation map. A natural measure to quantify this behavior for a function  $f$  is the Dirichlet energy,  $E_g(f) = \|\text{grad}_g f\|^2$ , where  $\text{grad}_g$  is the gradient operator associated with  $g$  and  $\|\cdot\|$  the  $L^2$  norm for vector fields. We will provide more precise definitions for the gradient operator and related Hilbert space notions in Section 4.1, but for now it suffices to note that the set of Dirichlet energies  $\{E_g(z_k)\}$  of the eigenfunctions is discrete in  $\mathbb{R}$ , and therefore a finite collection of eigenfunctions can be stably ordered in order of increasing  $E_g(z_k)$ . (Actually, in Section 4 we will work with a conformally transformed metric,  $h$ , with stronger invariance properties under changes of observation modality than  $g$ , but for now we work with  $g$  to illustrate ideas.) Intuitively, we expect that functions with small Dirichlet energy can be more accurately approximated from finite datasets than highly oscillatory functions, so our selection criterion for the generating eigenfunctions is to set  $\{\zeta_i\}_{i=1}^m$  to the first  $m$  eigenfunctions in that ordering corresponding to rationally independent frequencies. We adopt this selection criterion throughout the paper, and discuss its numerical implementation in Section 4.3. In general, the generating eigenfunctions selected in this way will not always agree among different observation modalities, but our selection criterion exhibits some rigidity in the sense that there exist equivalence classes of observation modalities for which the selected generating eigenfunctions and generating frequencies are the same.

Further useful properties of the Koopman eigenfunctions pertain to the dynamics in the image space from (6).

**Proposition 4.** *The vector field  $v$  is projectible under  $\pi_i$  in the sense that for any  $a, b \in M$  such that  $\pi_i(a) = \pi_i(b)$ , the equality  $\pi_{i*}v|_{T_{\pi_i(a)}\mathbb{C}} = \pi_{i*}v|_{T_{\pi_i(b)}\mathbb{C}}$  holds, where  $\pi_{i*} : TM \mapsto T\mathbb{C}$  is the derivative map of  $\pi_i$ , mapping tangent vectors on  $M$  to tangent vectors on  $\mathbb{C}$ .*

*Proof.* Let  $\{\hat{e}_1, \hat{e}_2\} = \{1, i\}$  be the canonical basis of  $\mathbb{C}$ . This basis can be canonically identified with a basis of  $T_{\pi_i(a)}\mathbb{C}$  so that  $v_* = \pi_{i*}v|_{\pi_i(a)} = \sum_{\alpha=1}^2 v_*^\alpha \hat{e}_\alpha$ , where

$$v_*^\nu = \sum_{\mu=1}^m v^\mu \frac{\partial \chi^\nu}{\partial \theta^\mu} = v(\chi^\nu), \quad \chi^1 = \text{Re } \zeta_i, \quad \chi^2 = \text{Im } \zeta_i.$$

Because  $v(\zeta_i) = i\Omega_i\zeta_i$ , we have  $v_*^\nu|_{T_{\pi_i(a)}\mathbb{C}} = v_*^\nu|_{T_{\pi_i(b)}\mathbb{C}}$  whenever  $\pi_i(a) = \pi_i(b)$ . In particular, note that  $v(\chi^1) = -\Omega_i\chi^2$  and  $v(\chi^2) = \Omega_i\chi^1$ .  $\square$

That  $v$  is projectible under  $\pi_i$  is a non-trivial property which is not satisfied by general manifold submersions, and ensures that the time evolution of the data in the image spaces  $\pi_i(M)$  can be described by means of autonomous deterministic dynamical systems, avoiding closure issues. In this particular case, it is straightforward to check that the projected dynamics are simple harmonic oscillations (rotations) with frequencies  $\Omega_i$ . Specifically, introducing the time series  $\tilde{\zeta}_{i,a}(t) = \zeta_i(\Phi_t a)$  we have the ordinary differential equation  $d\tilde{\zeta}_i/dt = i\Omega_i\tilde{\zeta}_i$  with the solution

$$\tilde{\zeta}_{i,a}(t) = \zeta_i(\Phi_t a) = \zeta_i(a)e^{i\Omega_i t}. \quad (7)$$

It follows from (7) that the image  $\pi_i(M)$  under each of the projection maps  $\pi_i$  is the full unit circle. Moreover, we have  $\pi_i \circ \Phi_t = \Xi_{i,t} \circ \pi_i$ , where  $\Xi_{i,t} : \mathbb{T} \mapsto \mathbb{T}$  is the circle rotation with frequency  $\Omega_i$  for time  $t \in \mathbb{R}$ ; i.e.,  $\Xi_{i,t}(\vartheta) = e^{i(\vartheta + \Omega_i t)}$ . Thus,  $\pi_i$  is a factor map [9, 10], sending the original flow (which can have

variable speed; see Section 4.6 ahead) to a constant-frequency rotation on the circle, and this map is chosen such that it is optimally smooth with respect to the Dirichlet energy  $E_g$ . Further, recalling that the  $\Omega_i$  are rationally independent, we see that for our canonical case,  $M = \mathbb{T}^m$ , the composite map  $\pi : M \mapsto \mathbb{C}^m$  with  $\pi(a) = (\pi_1(a), \dots, \pi_m(a))$  induces a diffeomorphism of the torus, mapping the original flow to the irrational flow  $\Xi_t = (\Xi_{1,t}, \dots, \Xi_{m,t})$  on  $\pi(M) = \mathbb{T}^m$  such that  $\pi \circ \Phi_t = \Xi_t \circ \pi$ .

### 3.2. Vector field decomposition

So far, we have constructed a family of projection maps taking the dynamical vector field  $v$  on  $M$  to the vector fields  $v_{i*}$  on the circle generating simple harmonic oscillations. However, given the  $v_{i*}$  it is generally not possible to perform the reverse operation and pull them back on  $M$ . In this Section, we construct a decomposition of  $v$  into a sum of vector fields which are mutually commuting, and describe “simpler” (but non-ergodic) dynamics than  $v$  in that they have non-trivial nullspaces.

**Theorem 5.** *Let  $v_i : L^2(M, \mu) \mapsto L^2(M, \mu)$  be the linear operator defined through its action on the eigenfunction basis from Definition 2:*

$$v_i(\zeta_1^{k_1} \cdots \zeta_i^{k_i} \cdots \zeta_m^{k_m}) = ik_i \Omega_i \zeta_1^{k_1} \cdots \zeta_i^{k_i} \cdots \zeta_m^{k_m}.$$

(i)  $v_i$  is a vector field on  $M$  with nullspace  $\ker v_i = \prod_{j \neq i} \mathcal{Z}_j$ , where  $\{\mathcal{Z}_j\}_{j=1}^m$  are the orthogonal subrings of  $L^2(M, \mu)$  generated by  $\zeta_j$ .

(ii) At every  $a \in M$ , the  $v_i$  are non-vanishing, linearly independent, and have vanishing commutator,  $[v_i, v_j] = v_i v_j - v_j v_i = 0$ .

(iii) The transformations  $\Phi_{i,t} : M \mapsto M$  generated by  $v_i$  preserve the invariant measure of the full dynamics.

(iv) The  $\Phi_{j,t}$  act on the  $v_i$  by translations, in the sense that the tangent vector  $u \in T_b M$  with  $u = \Phi_{j,t*}(v_i|_a)$  and  $b = \Phi_{j,t}(a)$  acts on  $f \in L^2(M, \mu)$  according to  $u(f) = v_i|_b(f)$  for every  $a \in M$ , where  $\Phi_{j,t*} : TM \mapsto TM$  is the derivative map associated with  $\Phi_{j,t}$ .

(v) The decomposition  $v = \sum_{i=1}^m v_i$  holds.

*Proof.* To prove that  $v_i$  is a vector field, it suffices to show that it obeys the Leibniz rule for all  $f_1, f_2 \in L^2(M, \mu)$ , and we will verify that this is the case by direct computation. Introducing the multi-indices  $k = (k_1, \dots, k_m)$  and  $l = (l_1, \dots, l_m)$  with  $k, l \in \mathbb{Z}^m$ , we have  $f_1 = \sum_k c_k \zeta_1^{k_1} \cdots \zeta_m^{k_m}$ ,  $f_2 = \sum_l d_l \zeta_1^{l_1} \cdots \zeta_m^{l_m}$ , and  $f_1 f_2 = \sum_{k,l} c_k d_l \zeta_1^{k_1+l_1} \cdots \zeta_m^{k_m+l_m}$ , where  $c_k = \langle \zeta_1^{k_1} \cdots \zeta_m^{k_m}, f_1 \rangle$  and  $d_l = \langle \zeta_1^{l_1} \cdots \zeta_m^{l_m}, f_2 \rangle$ . Therefore,  $v_i(f_1) = i\Omega_i \sum_k k_i c_k \zeta_1^{k_1} \cdots \zeta_i^{k_i} \cdots \zeta_m^{k_m}$ ,  $v_i(f_2) = i\Omega_i \sum_l l_i d_l \zeta_1^{l_1} \cdots \zeta_i^{l_i} \cdots \zeta_m^{l_m}$ , and

$$v_i(f_1 f_2) = i\Omega_i \sum_{k,l} (k_i + l_i) c_k d_l \zeta_1^{k_1+l_1} \cdots \zeta_i^{k_i+l_i} \cdots \zeta_m^{k_m+l_m}.$$

The Leibniz rule follows directly from the last three equations. Moreover, that  $\ker v_i = \prod_{j \neq i} \mathcal{Z}_j$  follows by noting that  $f$  in this subspace has the expansion coefficients with  $c_{k_1 \dots k_i \dots k_m} = 0$  for all  $k_i \neq 0$ , and therefore  $v_i(f) = 0$ . This proves part (i) of the Lemma.

The linear independence of  $v_i$  in part (ii) can be checked by noticing that  $v_j(\zeta_i) = 0$  for all  $j \neq i$ , but  $v_i(\zeta_i) = i\Omega_i \zeta_i$  is nowhere-vanishing on  $M$ . Thus,  $v_i(\zeta_i)$  cannot be expressed as a linear combination of  $\{v_j(\zeta_i)\}_{j \neq i}$ . Repeating this argument for all  $i \in \{1, \dots, m\}$  leads to the desired result. To establish commutativity of the  $v_i$ , we use again the decomposition for  $f_1$  and compute  $v_i v_j f_1 = -\Omega_i \Omega_j \sum_k k_i k_j c_k \zeta_1^{k_1} \cdots \zeta_m^{k_m}$ . This expression is invariant under permutations of  $i$  and  $j$ , leading to  $[v_i, v_j] f_1 = 0$  and completing the proof of part (ii).

Turning to part (iii), we first note that starting from any  $a \in M$  we can follow the integral curves of the  $v_i$  for an arbitrary time  $t \in \mathbb{R}$ , and therefore we can construct the flows  $\Phi_{i,t}$  in the statement of the Lemma. Then, a necessary and sufficient condition that these flows are measure preserving is that the functions  $\operatorname{div}_\mu v_i$  vanish everywhere on  $M$ , where  $\operatorname{div}_\mu$  is the divergence operator on vector fields associated with the

invariant measure  $\mu$  (see (15) for an expression in local coordinates). The adjoint of  $\text{div}_\mu$  is the negative gradient with respect to the ambient space metric  $g$ , and for any eigenfunction  $z_k$  we have

$$\langle z_k, \text{div}_\mu v_i \rangle = -\langle \text{grad}_g z_k, v_i \rangle = -\int_M v_i(z_k^*) d\mu = i\Omega_i k_i \int_M z_k d\mu = 0,$$

where the last equality holds because  $k_i = 0$  if  $z_k$  is constant, and  $\int_M z_k d\mu = \langle 1, z_k \rangle$  vanishes if  $z_k$  is nonconstant by ergodicity. Therefore,  $\text{div}_\mu v_i$  vanishes everywhere since it is orthogonal to every element in a complete basis of  $L^2(M, \mu)$ .

Next, since  $\{z_k\}$  is a basis of  $L^2(M, \mu)$ , to verify part (iv) it is sufficient to show that  $u(z_k) = v_i|_b(z_k)$ . Indeed, according to (7),

$$u(z_k) = (\Phi_{j,t*}(v_i|_a))(z_k) := v_i|_a(z_k \circ \Phi_{j,t}) = e^{i\Omega_j t k_j} v_i|_a(z_k) = e^{i\Omega_j t k_j} i\Omega_i k_i z_k(a) = i\Omega_i k_i z_k(b) = v_i|_b(z_k).$$

Finally, we verify the decomposition in part (iv) by operating with  $v$  on the function  $f_1$  and using the Leibniz rule, viz.

$$v(f_1) = \sum_{i=1}^m i\Omega_i \sum_k k_i c_k \zeta_1^{k_1} \cdots \zeta_m^{k_m} = \sum_{i=1}^m v_i(f_1). \quad \square$$

The vanishing commutator of the  $v_i$  from Theorem 5 is an intrinsic (observation map independent) dynamical independence property. This property can also be seen by considering local coordinates  $(\tilde{\theta}^1, \dots, \tilde{\theta}^m)$  constructed through the canonical angle coordinates of the circles  $\pi_i(M)$  from the projection maps in (6). In these coordinates, we have  $v_i = \Omega_i \frac{\partial}{\partial \tilde{\theta}^i}$ , and the full vector field,  $v = \sum_{i=1}^m \Omega_i \frac{\partial}{\partial \tilde{\theta}^i}$ , is decomposed into manifestly independent components. This decomposition has connections with the nonlinear independent component analysis technique of Singer and Coifman [66], which recovers independent components of stochastic differential equations on manifolds using kernel methods. Globally on  $M$ , the  $v_i$  generate measure-preserving, but non-ergodic, transformations  $\Phi_{i,t}$  (together with associated Koopman operators), giving the full evolution map through the composition  $\Phi_t = \Phi_{1,t} \circ \cdots \circ \Phi_{m,t}$ , where the order of the components does not matter. Moreover, as is the case with  $\Phi_t$ , the projection maps  $\pi_i$  in (6) are factor maps mapping the dynamics on  $M$  associated with  $\Phi_{i,t}$  to the corresponding circle rotation  $\Xi_{i,t}$  introduced in Section 3.1; i.e.,  $\pi_i \circ \Phi_{i,t} = \Xi_{i,t} \circ \pi_i$ .

The vector fields  $v_i$  are intrinsically defined as differential operators on  $M$ , and to reconstruct them in data space  $\mathbb{R}^d$  we apply the pushforward map for tangent vectors introduced in Section 2.

**Proposition 6.** *Let  $u$  be a smooth vector field on  $M$ , and let  $F$  be a smooth embedding of  $M$  into  $\mathbb{R}^d$ . Then, the image  $u_* = F_*(u)$  under the pushforward map  $F_* : TM \mapsto T\mathbb{R}^d$  is given by  $u_* = u(F)$ , where  $u$  acts on  $F$  componentwise in a basis of  $\mathbb{R}^d$ .*

*Proof.* Let  $\{e_1, \dots, e_d\}$  be a basis of  $\mathbb{R}^d$  (not necessarily orthonormal). This basis induces a basis of the tangent space  $T_x \mathbb{R}^d$  at every  $x \in \mathbb{R}^d$  through the canonical isomorphism  $T_x \mathbb{R}^d \simeq \mathbb{R}^d$ , and we have  $u_* = \sum_{\nu=1}^d u_*^\nu e_\nu$ . Denoting the components of  $u$  in local coordinates  $\{\theta^1, \dots, \theta^m\}$  on  $M$  by  $\{u^1, \dots, u^m\}$ , for any smooth function  $f$  we have  $u(f) = \sum_{\mu=1}^m u^\mu \frac{\partial f}{\partial \theta^\mu}$ . In particular, expanding the observation map in the  $\{e_i\}$  basis,  $F = \sum_{\nu=1}^d F^\nu e_\nu$ , and using the transformation law for tangent vectors on manifolds, we obtain  $u_*^\nu = \sum_{\mu=1}^m \frac{\partial F^\nu}{\partial \theta^\mu} u^\mu = u(F^\nu)$  and therefore  $u_* = u(F)$ .  $\square$

Using Proposition 6 and the Fourier expansion of the observation map in (5), we compute the reconstructed vector fields in  $\mathbb{R}^d$  given by

$$V_i = F_* v_i = v_i(F) = \sum_k i k_i \Omega_i \hat{F}_k z_k, \quad V = F_* v = \sum_{i=1}^m V_i, \quad (8)$$

where  $\hat{F}_k = \langle z_k, F \rangle \in \mathbb{R}^d$  are the Koopman modes from (5), and the last equality follows by linearity of the pushforward map. Note that  $V$  can also be approximated from finite differences of time-ordered data as described in Section 2 and [58], but the components  $V_i$  in general cannot. In applications, the

$V_i$  can be visualized as spatiotemporal patterns (movies), or if  $d$  is sufficiently small, as arrow plots on the data manifold  $F(M)$ . In particular, because  $F_*$  is the pushforward map for vector fields, the vector  $V_i|_a = \sum_k ik_i \Omega_i \hat{F}_k z_k(a)$  is guaranteed (modulo numerical errors) to be tangent to  $F(M)$  at the point  $F(a)$ ; see Figs. 1, 5, 14, and 18 for examples. The following is a direct consequence of Proposition 6 and the definition of the  $V_i$  through (8).

**Corollary 7.** The reconstructed vector fields  $V_i$  transform naturally as type  $(1,0)$  tensors. That is, given a map  $G : \mathbb{R}^d \mapsto \mathbb{R}^{\bar{d}}$  such that  $\tilde{F} = G \circ F$  is an embedding of  $M$  into  $\mathbb{R}^{\bar{d}}$ , the reconstructed vector fields  $\tilde{V}_i = \tilde{F}_* v_i$  in  $\mathbb{R}^{\bar{d}}$  are related to the  $V_i$  via  $\tilde{V}_i = G_* V_i$ .

Note that the individual Koopman modes  $\hat{F}_k$  do not obey an analogous transformation law. That is, if  $\hat{\tilde{F}}_k = \langle z_k, \tilde{F} \rangle = \langle z_k, G \circ F \rangle$ , then in general  $\hat{\tilde{F}}_k$  is not equal to  $G(\hat{F}_k)$ .

### 3.3. Nonparametric forecasting

The availability of the closed form solution in (7) for the temporal evolution of the Koopman eigenfunctions in systems with pure point spectra, together with the fact that they form a complete basis of  $L^2(M, \mu)$ , makes these eigenfunctions well suited for data-driven nonparametric forecasting. In this section, we formulate a technique for forecasting probability densities and expectation values of observables with initial data specified as a probability measure. Our approach follows closely the nonparametric framework developed in [44], with the difference that here we use the additional structure in the temporal evolution of the eigenfunctions.

Consider the initial data given as a probability measure  $\mu_0$  on  $(M, \mathcal{B})$  with a smooth probability density  $\rho_0$  relative to the invariant measure  $\mu$ . The measure  $\mu_0$  evolves at time  $t$  according to  $\mu_t = \Phi_{t*} \mu_0$ , and the corresponding density is given by  $\rho_t = U_t^* \rho_0$ , where  $U_t^* = U_{-t}$  is the Perron-Frobenius operator, i.e., the adjoint of the Koopman operator on  $L^2(M, \mu)$ . Thus, for an initial density with Fourier expansion  $\rho_0 = \sum_k \hat{\rho}_k(0) z_k$ , where  $\hat{\rho}_k(0) = \langle z_k, \rho_0 \rangle$ , we have

$$\rho_t = \sum_k \hat{\rho}_k(t) z_k, \quad \text{with} \quad \hat{\rho}_k(t) = e^{-i\omega_k t} \hat{\rho}_k(0).$$

Note that the expansion coefficients form the  $\ell^2$  sequence  $\hat{\rho}(t) = (\dots, \hat{\rho}_{k-1}(0), \hat{\rho}_{k_0}(0), \hat{\rho}_{k_1}(0), \dots)$ , and we can write the formal solution

$$\hat{\rho}(t) = e^{i\omega t} \hat{\rho}(0), \tag{9}$$

where  $\omega$  is the diagonal operator on  $\ell^2$  such that

$$\omega \hat{\rho} = (\dots, \omega_{k-1} \hat{\rho}_{k-1}, 0, \omega_{k_1} \hat{\rho}_{k_1}, \dots). \tag{10}$$

Next, to compute the expectation value  $\bar{f}_t = \mathbb{E}_{\mu_t} f$  of an observable  $f \in L^2(M, \mu)$  with respect to  $\mu_t$ , we make use of the fact that

$$\mathbb{E}_{\mu_t} f = \int_M f \rho_t d\mu = \langle \rho_t^*, f \rangle.$$

This leads to the expression

$$\bar{f}_t = \sum_k e^{i\omega_k t} \hat{\rho}_k^*(0) \hat{f}_k, \tag{11}$$

where  $\hat{f}_k = \langle z_k, f \rangle$  are the expansion coefficients of  $f$  in the eigenfunction basis. Evaluating the mean square forecast  $\mathbb{E}_{\mu_t} f^2$  using a similar approach (or the convolution identity mentioned in Section 2), we obtain the variance forecast  $\sigma_t^2 = \mathbb{E}_{\mu_t} f^2 - \bar{f}_t^2$  which is useful for uncertainty quantification.

**Remark 8 (Deterministic forecast).** Forecasting with the Koopman eigenfunctions can also be performed with initial data given as a single observation  $y$  in the ambient data space  $\mathbb{R}^d$ . In this case, we first compute values  $\{\hat{\zeta}_i\}_{i=1}^m$  for the  $m$  generating eigenfunctions at the point  $a \in M$  with  $F(a) = y$  using out-of-sample extension techniques for functions (e.g., [67, 68]), and then evolve the initial values  $\{\hat{\zeta}_i\}$  via (7). We

then determine the value of other eigenfunctions using their group structure, and reconstruct the observable through its expansion coefficients and the eigenfunction values at the desired lead time. Note that if  $F(M)$  is the data manifold from an imperfect model with model error, then  $y$  may not lie on  $F(M)$ , but extended function values can also be computed in this case. This approach is closely related to a kernel analog forecasting framework developed in [69]. In numerical experiments not reported here, we have observed comparable skill with this method and the statistical forecast results of Section 4.6.

### 3.4. Irrational flow on the 2-torus

We demonstrate the techniques presented in Sections 3.1–3.3 in an analytically solvable example involving an irrational flow on the 2-torus. Denoting the azimuthal and polar angles on the 2-torus by  $(\theta^1, \theta^2)$ , respectively, we consider the dynamical vector field on  $M = \mathbb{T}^2$  given by

$$v = \sum_{\mu=1}^2 v^\mu \frac{\partial}{\partial \theta^\mu}, \quad \text{with } v^1 = 1, \quad v^2 = \alpha, \quad (12)$$

where  $\alpha$  is a positive angular frequency parameter which is set to an irrational number to produce an ergodic flow. This dynamical system is observed through the observation map  $F : M \mapsto \mathbb{R}^3$  corresponding to the standard embedding of the 2-torus into three-dimensional Euclidean space, i.e., for the point  $a \in M$  with coordinates  $(\theta^1, \theta^2)$ , we have  $F(a) = (F^1(a), F^2(a), F^3(a)) = (x^1, x^2, x^3)$ , where

$$x^1 = (1 + R \cos \theta^2) \cos \theta^1, \quad x^2 = (1 + R \cos \theta^2) \sin \theta^1, \quad x^3 = \sin \theta^2, \quad R \in (0, 1). \quad (13)$$

The equilibrium measure of this system has uniform density relative to the Haar (planar) measure  $\mu$ , where  $d\mu = d\theta^1 \wedge d\theta^2 / (2\pi)^2$ . Moreover, the eigenvalue problem for  $v$  in (3) has solutions  $\lambda_k = i(k_1 + k_2\alpha)$  and  $z_k(\theta^1, \theta^2) = e^{i(k_1\theta^1 + k_2\theta^2)}$ , with  $k = (k_1, k_2)$ ,  $k_i \in \mathbb{Z}$ .

To select generating eigenfunctions  $\{\zeta_1, \zeta_2\}$  with low roughness on the torus and their corresponding basic frequencies  $\{\Omega_1, \Omega_2\}$ , we compute the Dirichlet energy of the eigenfunctions in the induced Riemannian geometry from the observation map as described in Section 3.1. For the embedding in (13), the induced Riemannian metric has components  $g_{11} = (1 + R \cos \theta^2)^2$ ,  $g_{22} = R^2$ , and  $g_{12} = g_{21} = 0$  in the  $\{\theta^\mu\}$  coordinates, leading to the Dirichlet energy values  $E_g(z_k) = C_1 k_1^2 + C_2 k_2^2$ , where  $C_1 = 1/(1 - R^2)^{3/2}$  and  $C_2 = \alpha^2/R^2$ . Thus, in this geometry, the least-rough eigenfunctions corresponding to rationally independent frequencies are those with  $k = (0, 1)$  and  $k = (0, 2)$ , i.e., we have  $\Omega_1 = 1$ ,  $\Omega_2 = \alpha$ ,  $\zeta_1(\theta^1, \theta^2) = e^{i\theta^1}$ , and  $\zeta_2(\theta^1, \theta^2) = e^{i\theta^2}$ . Clearly, the image of the torus under each of the projection maps from (6) is the unit circle,  $\pi_i(a) = e^{i\Omega_i \theta^i}$ , and the system evolves in these coordinates as a simple harmonic oscillator in accordance with (7).

**Remark 9 (Highly oscillatory eigenfunctions and slow observables).** In this torus rotation, for every eigenfunction  $z_k$  corresponding to the eigenvalue  $\lambda_k$  there exist eigenfunctions with eigenvalues arbitrarily close to  $\lambda_k$  and with arbitrarily large Dirichlet energy. This is a consequence of the density of the spectrum of  $v$  on the real line (see Remark 3). In particular, the eigenvalue and Dirichlet energy of the constant eigenfunction are both zero, but for any  $\epsilon, \bar{E} > 0$  we can find integers  $i$  and  $j$  such that for  $k = (i, j)$ ,  $|\lambda_k| < \epsilon$  and  $E_g(z_k) > \bar{E}$ . In other words, there exist observables  $z_k$  with arbitrarily small frequency  $|\lambda_k|$  but arbitrarily large roughness  $E_g(z_k)$ . This behavior is generic in systems possessing two or more rationally independent Koopman eigenvalues, and can adversely affect the conditioning of numerical schemes for Koopman eigenvalues and eigenfunctions. More generally, the identification of slow observables is an important task in reduced dynamical modeling (e.g., [6, 34]), and the simple example discussed here suggests that care may be needed to ensure that the identified slow observables are also smooth. In Section 4.3, we will suppress the pathological Koopman eigenfunctions with large Dirichlet energy by adding a small amount of diffusion to  $v$ .

Next, we reconstruct the vector fields  $v_i$  from Theorem 5 associated with the generating eigenfunctions  $\{\zeta_1, \zeta_2\}$  using the spectral expansion of  $F$  in the  $\{z_k\}$  basis in accordance with (8). Setting  $k = (i, j) \in \mathbb{Z}^2$ ,

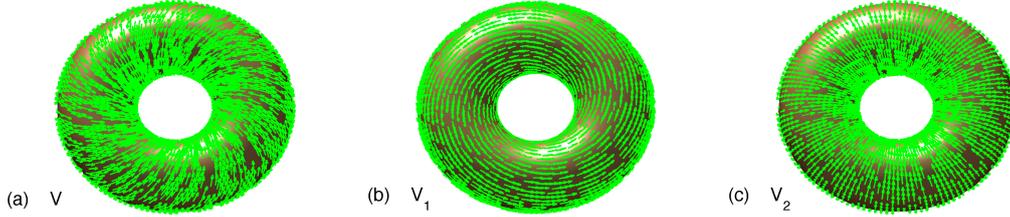


Figure 1: Decomposition of the vector field of an irrational flow on the torus into mutually commuting components. (a) Full vector field; (b,c) the components  $v_i = \Omega_i \frac{\partial}{\partial \theta^i}$  from Theorem 5, reconstructed in  $\mathbb{R}^3$  using the pushforward map from (8).

the expansion coefficients are

$$\begin{aligned}\hat{F}_{ij}^1 &= \langle z_{ij}, F^1 \rangle = \frac{1}{2}(\delta_{i1}\delta_{j0} - \delta_{i,-1}\delta_{j0}) + \frac{R}{4}(\delta_{i1}\delta_{j1} + \delta_{i,-1}\delta_{j1} + \delta_{i1}\delta_{j,-1} + \delta_{i,-1}\delta_{j,-1}), \\ \hat{F}_{ij}^2 &= \langle z_{ij}, F^2 \rangle = \frac{1}{2i}(\delta_{i0}\delta_{j1} - \delta_{i0}\delta_{j,-1}) + \frac{R}{4i}(\delta_{i1}\delta_{j1} - \delta_{i,-1}\delta_{j1} + \delta_{i1}\delta_{j,-1} - \delta_{i,-1}\delta_{j,-1}), \\ \hat{F}_{ij}^3 &= \langle z_{ij}, F^3 \rangle = \frac{R}{2i}(\delta_{i0}\delta_{j1} - \delta_{i0}\delta_{j,-1}),\end{aligned}$$

giving  $V_i = (V_i^1, V_i^2, V_i^3)$  with

$$V_1^1 = -\sin \theta^1 - \frac{R}{2}(\sin(\theta^1 + \theta^2) + \sin(\theta^1 - \theta^2)), \quad V_2^1 = \cos \theta^1 + \frac{R}{2}(\cos(\theta^1 + \theta^2) + \cos(\theta^1 - \theta^2)), \quad V_3^1 = 0,$$

and

$$V_2^1 = -\frac{R}{2}(\sin(\theta^1 + \theta^2) + \sin(\theta^1 - \theta^2)), \quad V_2^2 = \frac{R}{2}(\cos(\theta^1 + \theta^2) - \cos(\theta^1 - \theta^2)), \quad V_2^3 = R \cos \theta^2.$$

This decomposition is depicted in Fig. 1.

Consider now statistical forecasting of the irrational-flow system using the nonparametric approach of Section 3.3. In this example, we set the initial probability measure  $\mu_0$  to a von Mises (circular Gaussian) distribution on the torus with the density function

$$\rho_0(\theta^1, \theta^2) = e^{\kappa(\cos(\theta^1 - \bar{\theta}^1) + \cos(\theta^2 - \bar{\theta}^2))} / (I_0(\kappa))^2 \quad (14)$$

relative to the equilibrium measure. In (14),  $I_n$  is the modified Bessel function of order  $n$ , and we use the values  $(\bar{\theta}^1, \bar{\theta}^2) = (0, 0)$  and  $\kappa = 30$  for the location and concentration parameters, respectively. We take the component  $F^1$  of the observation map as the forecast observable  $f$ , and compute the time-dependent expectation value and standard deviation of  $f$  using (11). The latter equation can be evaluated analytically using properties of Bessel functions and the expansion of  $F^1$  in the eigenfunction basis. In particular, using the result  $\int_0^{2\pi} e^{in\theta + \kappa \cos \theta} d\theta = 2\pi I_n(\kappa) / I_0(\kappa)$ , we find

$$\begin{aligned}\bar{f}_t &= \frac{I_1(\kappa)}{I_0(\kappa)} \cos t + \frac{RI_1^2(\kappa)}{I_0^2(\kappa)} (\cos((1+\alpha)t) + \cos((1-\alpha)t)), \\ \bar{f}_t^2 &= \left(1 + \frac{I_2(\kappa)}{I_0(\kappa)} \cos(2t)\right) \left[\frac{1}{2} + \frac{RI_1(\kappa)}{I_0(\kappa)} \cos(\alpha t) + \frac{R^2}{4} \left(1 + \frac{I_2(\kappa)}{I_0(\kappa)} \cos(2\alpha t)\right)\right].\end{aligned}$$

The time evolution of  $\bar{f}_t$  and the standard deviation  $\sigma_t = (\bar{f}_t^2 - \bar{f}_t)^{1/2}$  are shown in Fig 2.

#### 4. Galerkin approximation in a data-driven orthonormal basis

In this section, we present a Galerkin method with regularization for the eigenvalue problem of the generator in an orthonormal basis acquired through the diffusion maps algorithm. Our approach follows

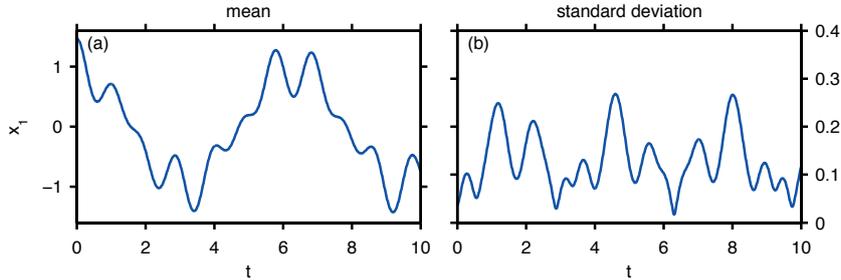


Figure 2: Statistical forecast of the component  $x^1$  of the torus embedding in  $\mathbb{R}^3$  for the irrational flow with frequencies  $(1, 30^{1/2})$ . The initial probability measure has the circular Gaussian density from (14) relative to the Haar measure with location and concentration parameters  $(0, 0)$  and  $30$ , respectively. (a) Mean forecast; (b) standard deviation.

closely [44], with the difference that we work in the Sobolev space  $H_1(M, \mu)$  appropriate for the weak formulation of the regularized eigenvalue problem (as opposed to the  $L^2(M, \mu)$  space employed in [44]). We also discuss the spectral properties of the regularized operator, including its asymptotic behavior in the weak-diffusion limit—this discussion will also motivate the time-change techniques of Section 6. We then describe the implementation of the techniques of Section 3 in the approximate eigenfunction basis acquired via the Galerkin scheme, and present numerical applications to a variable-speed ergodic flow on the 2-torus.

#### 4.1. Choice of Galerkin approximation space

Let  $g$  be the Riemannian metric on  $M$  inherited from the observation map (see Section 3.1), and  $\sigma$  the  $C^\infty$  density of the invariant measure of the dynamics relative to the Riemannian measure of  $g$ ; i.e.,  $\sigma = d\mu/d\text{vol}_g$ , where  $\sigma$  is bounded away from zero by compactness of  $M$ . For data generated by ergodic dynamical systems,  $\sigma$  is also the sampling density relative to the Riemannian measure, though the samples collected from a single time series (as is the case in (1)) are not independent. In general,  $\sigma$  will be a nonconstant function, and in what follows we work with the conformally transformed metric  $h = g\sigma^{2/m}$ . In particular, we will use Laplace-Beltrami eigenfunctions associated with  $h$  as a basis of our approximation space for the Koopman generator. Note that  $h$  contracts (expands) local distances with respect to the original metric  $g$  in regions of small (high) sampling density  $\sigma$ . As we will see below, due to this property the Laplace-Beltrami eigenfunctions associated with  $h$  acquire increased “resolution” in high- $\sigma$  regions, i.e., in regions where high resolution can be robustly attained using finite datasets. In Section 4.2, we will discuss how to approximate the Laplace-Beltrami eigenfunctions associated with  $h$  via diffusion maps.

**Proposition 10.** *Let  $F : M \mapsto \mathbb{R}^d$  and  $\tilde{F} : M \mapsto \mathbb{R}^{\tilde{d}}$  be embeddings of  $M$  with the corresponding induced Riemannian metrics  $g$  and  $\tilde{g}$ , respectively. Assume that  $g$  and  $\tilde{g}$  are conformally equivalent, i.e., that there exists a positive function  $r \in C^\infty(M)$  such that  $\tilde{g} = rg$ . Then,  $g\sigma^{2/m} = \tilde{g}\tilde{\sigma}^{2/m}$ , where  $\sigma = d\mu/d\text{vol}_g$  and  $\tilde{\sigma} = d\mu/d\text{vol}_{\tilde{g}}$  are the sampling densities associated with  $g$  and  $\tilde{g}$ , respectively.*

*Proof.* The claim follows immediately from the fact that  $\tilde{\sigma} = d\mu/d\text{vol}_{\tilde{g}} = r^{-m/2}d\mu/d\text{vol}_g = r^{-m/2}\sigma$ .  $\square$

**Corollary 11.** The metric  $h$  is unique for each equivalence class of observation maps associated with conformally equivalent induced metrics. Moreover,  $h$  has uniform volume form relative to the invariant measure of the dynamics since  $d\mu/d\text{vol}_h = (1/\sigma)d\mu/d\text{vol}_g = 1$ .

Consider now the Hilbert space  $L^2T(M, h)$  of square-integrable vector fields on  $M$  associated with  $h$ , equipped as usual with the Hodge inner product  $\langle u_1, u_2 \rangle = \int_M h(u_1, u_2) d\text{vol}_h = \int_M h(u_1, u_2) d\mu$  and norm  $\|u\| = \langle u, u \rangle^{1/2}$ . We denote the gradient of a function  $f \in C^\infty(M)$  with respect to  $h$  by  $\text{grad}_h f = h^{-1}(df, \cdot)$ , where  $h^{-1}$  is the “inverse metric”. Note that  $\text{grad}_h f = \sigma^{-2/m} \text{grad}_g f$ , and because  $d\mu/d\text{vol}_h$  is a constant,  $\text{grad}_h = -\text{div}_\mu^*$  where adjoints are taken on the Hilbert spaces associated with  $h$  and  $\mu$ . In local coordinates,

we have  $d\text{vol}_h = \sqrt{\det h} d\theta^1 \wedge \cdots \wedge d\theta^m$ ,  $\text{grad}_h = \sum_{i=1}^m (\text{grad}_h)^i \frac{\partial}{\partial \theta^i}$ , and

$$(\text{grad}_h f)^i = \sum_{j=1}^m h^{-1,ij} \frac{\partial f}{\partial \theta^j}, \quad \text{div}_\mu(u) = \frac{1}{\sqrt{\det h}} \sum_{i=1}^m \frac{\partial}{\partial \theta^i} \left( \sqrt{\det h} u^i \right). \quad (15)$$

We also introduce the order-1 Sobolev space  $H_1(M, h, \mu)$  associated with  $h$  and  $\mu$  (henceforth abbreviated as  $H_1(M, h)$  since  $\mu = \text{vol}_h$ ), which is equipped with the inner product  $\langle f_1, f_2 \rangle_1 = \langle f_1, f_2 \rangle + \langle \text{grad}_h f_1, \text{grad}_h f_2 \rangle$ .

The Dirichlet energy of functions in  $H_1(M, h)$  is given by the functional

$$E_h(f) = \langle \text{grad}_h f, \text{grad}_h f \rangle = \int_M \|\text{grad}_h f\|_h^2 d\mu = \int_M \|\text{grad}_g f\|_g^2 \sigma^{-2/m} d\mu, \quad (16)$$

where  $\|\text{grad}_h f\|_h^2 = h(\text{grad}_h f, \text{grad}_h f)$  and  $\|\text{grad}_g f\|_g^2 = g(\text{grad}_g f, \text{grad}_g f)$ . This functional provides a metric-dependent measure of roughness of functions which can be used to select generators for the spectrum of the Koopman group as described in Section 3.1. Note that due to the presence of the  $\sigma^{-2/m}$  term in the last integral in (16), functions with large gradient with respect to the ambient space metric  $g$  in regions of small sampling density will generally acquire large Dirichlet energy with respect to  $h$ .

Next, consider the Laplace-Beltrami operator  $\Delta_h = -\text{div}_\mu \text{grad}_h$  associated with the Riemannian metric  $h$ . This operator is positive-semidefinite and self-adjoint on  $L^2(M, \mu)$ , and there exists an orthonormal basis  $\{\phi_i\}_{i=0}^\infty$  consisting of the eigenfunctions

$$\Delta_h \phi_i = \eta_i \phi_i, \quad \phi_i \in C^\infty(M), \quad 0 = \eta_0 < \eta_1 \leq \eta_2 \leq \cdots \nearrow \infty, \quad (17)$$

with the corresponding eigenvalues  $\{\eta_i\}_{i=0}^\infty$ . Note that while the inner product of  $L^2(M, \mu)$  is metric-tensor-independent, the gradient operator, and hence  $\Delta_h$ ,  $\phi_i$ , and  $\eta_i$ , all depend on  $h$  (which depends in turn on the observation map  $F$ ). It is a standard result on Laplace-Beltrami operators that the eigenvalues are extrema of the Rayleigh quotient  $R_h(f) = E_h(f)/\|f\|^2$ , and the corresponding eigenfunctions are the extremizers; i.e.,  $\lambda_i = R_h(\phi_i)$ , and  $R_h(\phi_i) = E_h(\phi_i)$  for normalized eigenfunctions. Since the integrals in the evaluation of  $R_h$  are with respect to the invariant measure of the dynamics, we interpret  $R_h$  as a measure of “expected roughness” with respect to the Riemannian metric  $h$ . In particular, due to the  $\sigma$ -dependent term in (16), the leading extrema of  $R_h$  will generally correspond to eigenfunctions with weaker oscillatory behavior (as measured with respect to the ambient space metric) in regions of small sampling density and stronger oscillatory behavior (thus, higher resolution) in regions of large sampling density. Thus, we can interpret the finite collection  $\{\phi_0, \phi_1, \dots, \phi_{l-1}\}$  as the  $l$ -element orthonormal set on  $L^2(M, \mu)$  with the least expected roughness for the equivalence class of observation maps associated with  $h$ , in the sense that  $\sum_{i=0}^{l-1} R_h(\phi_i) \leq \sum_{i=0}^{l-1} R_h(f_i)$ , where  $\{f_0, f_1, \dots, f_{l-1}\}$  is any  $l$ -element orthonormal set on  $L^2(M, \mu)$ .

Intuitively, functions with small expected roughness can be robustly approximated from finite datasets in  $M$ , and in the case of i.i.d. samples this intuition can be rigorously verified through pointwise and spectral convergence results established for graph Laplacians [37, 43, 62, 70–73]. In particular, it has recently been shown [73] that a spectrum of the Laplacian-Beltrami operator for a Riemannian metric analogous to  $h$  can be consistently estimated by normalized graph Laplacians for a variable-bandwidth kernel of the same class as (20). Moreover, they show that for the class of conformally invariant metrics  $h$  the variance of the approximated eigenvalues has the leading-order behavior  $\text{var } \eta_i = \frac{\epsilon^{-(m/2+1)}}{N(N-1)} C \eta_i \langle \phi_i^2, \phi_i^2 \rangle / \langle \phi_i, \phi_i \rangle^2$ , where  $C$  is a constant independent of  $\eta_i$ ,  $\phi_i$ , and, importantly, the sampling density  $\sigma$ . In other words, the effect of the conformal change of metric  $g \mapsto h$  is to “undo” the effect of sampling density fluctuations and increase the robustness of the approximated spectrum of  $\Delta_h$ . While we are not aware of analogous spectral or pointwise convergence results in the case of correlated samples generated by ergodic dynamical systems, the favorable properties of  $h$  for robust data analysis should hold in that case too.

Notice now that the Laplace-Beltrami eigenfunctions from (17) are orthogonal on  $H_1(M, h)$  with  $\langle \phi_i, \phi_j \rangle_1 = (1 + \eta_i) \delta_{ij}$ , but because  $\|\phi_i\|_1 = (1 + \eta_i)^{1/2}$  exhibits unbounded growth as  $i \rightarrow \infty$ , functions of the form  $f = \sum_{i=0}^\infty c_i \phi_i$  with  $(c_0, c_1, \dots) \in \ell^2$  are not necessarily in  $H_1(M, h)$ . On the other hand, the rescaled eigenfunctions

$$\varphi_i = \begin{cases} \phi_0, & i = 0, \\ \eta_i^{-1/2} \phi_i, & i > 1, \end{cases} \quad (18)$$

with  $\langle \varphi_i, \varphi_j \rangle = (1 + \eta_i^{-1/2})\delta_{ij}$  for  $i > 0$  and  $(\varphi_0, \varphi_i) = \delta_{0i}$ , are orthogonal (but not orthonormal) on  $H_1(M, h)$  and have bounded  $H_1$  norm,  $\|\varphi_i\|_1 = (1 + \eta_i^{-1})$ . Therefore,  $\{\varphi_i\}_{i=0}^\infty$  is an orthogonal basis of  $H_1(M, h)$  with the property that every sequence  $(c_0, c_1, \dots) \in \ell^2$  corresponds to a function  $f = \sum_{i=0}^\infty c_i \varphi_i \in H_1(M, h)$ . Moreover, the Dirichlet energies of the basis elements are all equal to one for  $i > 0$ . Due to this property and the fact that the approximated spectrum of  $\Delta_h$  is robust against variations in the sampling density,  $\{\varphi_i\}$  will be our basis of choice for a well-conditioned Galerkin method for the eigenvalue problem of the Koopman generator. We also note that because  $E_h(\varphi_0) = 0$  and  $E_h(\varphi_i) = 1$  for  $i > 1$ , the Dirichlet energy from (16) can be conveniently computed from the  $\ell^2$  norm of the expansion coefficients with  $i \geq 1$ , i.e.,

$$E_h(f) = \sum_{i=1}^{\infty} |c_i|^2. \quad (19)$$

**Remark 12 (Weighted Laplacian).** An alternative elliptic operator to  $\Delta_h$ , whose eigenfunctions also provide an orthonormal basis of  $L^2(M, \mu)$  is the weighted Laplacian  $\Delta_{g, \mu} = -\operatorname{div}_\mu \operatorname{grad}_g$  associated with the invariant measure of the dynamics and the ambient-space metric  $g$ . This operator is the generator of a gradient flow on the Riemannian manifold  $(M, g)$  with potential  $-\log \sigma$ , and its eigenvalues and eigenfunctions can be obtained by extremizing the Rayleigh quotient  $R_{g, \mu}(f) = E_{g, \mu}(f)/\|f\|^2$ , where  $E_{g, \mu}(f) = \int_M \|\operatorname{grad}_g f\|_g^2 d\mu$ . Note that unlike  $E_h$  from (16), the Dirichlet energy  $E_{g, \mu}$  does not feature a  $\sigma$ -dependent term in the integral with respect to the invariant measure. Numerically, eigenfunctions of  $\Delta_{g, \mu}$  can be computed using variable-bandwidth kernels [43], or the canonical formulation of diffusion maps with radial Gaussian kernels and the “ $\alpha = 1/2$ ” normalization [37]. In [44], eigenfunctions of  $\Delta_{g, \mu}$  approximated via a variable bandwidth kernel were used to approximate Kolmogorov and Fokker-Planck operators of stochastic dynamical systems on manifolds. The Galerkin scheme for the Koopman operator developed here can be implemented using eigenfunctions of either  $\Delta_h$  or  $\Delta_{g, \mu}$  (computed via either standard diffusion maps, or variable-bandwidth kernels), though in practice we find that  $\Delta_h$  behaves more stably in applications with large variations of the sampling density  $\sigma$  (including the applications discussed in this paper).

#### 4.2. Data-driven orthonormal basis

To approximate the basis in (18) from data, we start from the variable-bandwidth kernel  $K_\epsilon : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$  given by [43]

$$K_\epsilon(x, y) = \exp\left(-\frac{\|x - y\|^2}{\epsilon \hat{\sigma}_\epsilon^{-1/m}(x) \hat{\sigma}_\epsilon^{-1/m}(y)}\right), \quad (20)$$

where  $\epsilon$  is a positive bandwidth parameter,  $\hat{\sigma}_\epsilon$  is a function approximating the sampling density  $\sigma$  at  $O(\epsilon)$  accuracy, and  $m$  is the dimension of  $M$ . The function  $\hat{\sigma}_\epsilon$  can be computed using any suitable density-estimation technique, and in what follows we employ the kernel method described in [43, 44]. This method uses an updated formulation of an automatic bandwidth-selection procedure originally developed in [74], and also provides an estimate of  $m$ . Alternatively,  $m$  can be estimated using one of the dimension estimation techniques available in the literature (e.g., [75, 76]). In the numerical experiments that follow we use a priori known values for  $m$ , though the estimates from the bandwidth selection algorithm are in good agreement with the true values.

Let  $q$  be the sampling density of the data relative to the invariant measure  $\mu$ . Throughout this section we consider that the sampling measure is equal to the invariant measure so that  $q = 1$ , but we carry out our analysis in the more general setting with nonuniform  $q$  in anticipation of the time-change techniques of Section 6. Using the same method to tune the kernel bandwidth parameter as in the density-estimation step, we perform the normalizations originally introduced in diffusion maps [37] and further developed in [65] to construct from the kernel in (20) an integral operator  $\mathcal{P}_\epsilon : L^2(M, \mu) \mapsto L^2(M, \mu)$  approximating the heat operator on the Riemannian manifold  $(M, h)$ . Specifically, we compute the action of  $\mathcal{P}_\epsilon$  on a function  $f \in L^2(M, \mu)$  through the operations,

$$\mathcal{G}_\epsilon f = \frac{1}{\epsilon^{m/2}} \int_M K_\epsilon(F(\cdot), F(a)) f(a) d\mu(a), \quad q_\epsilon = \mathcal{G}_\epsilon q, \quad \mathcal{H}_\epsilon f = \mathcal{G}_\epsilon(fq/q_\epsilon), \quad \mathcal{P}_\epsilon f = \frac{\mathcal{H}_\epsilon f}{\mathcal{H}_\epsilon 1}. \quad (21)$$

The operator  $\mathcal{P}_\epsilon$  preserves constant functions (i.e., it is an averaging operator with  $\mathcal{P}_\epsilon 1 = 1$ ), and we have  $\mathcal{P}_\epsilon f = \int_M p_\epsilon(\cdot, a) f(a) d\mu(a)$  for the Markov kernel

$$p_\epsilon(a, b) = \frac{K_\epsilon(F(a), F(b))q(b)}{\mathcal{H}_\epsilon 1(a)q_\epsilon(b)}, \quad a, b \in M.$$

The role of the normalizations in (21) is to remove biases due to curvature in  $h$  and the potentially nonuniform sampling density. In particular, taking Taylor expansions of  $\mathcal{P}_\epsilon$  about  $\epsilon = 0$  (e.g., [58, 65]), one can show that uniformly on  $M$ , and independently of  $q$ ,

$$\mathcal{P}_\epsilon f(a) = f(a) - \epsilon c \Delta_h f(a) + O(\epsilon^2) \quad (22)$$

for a constant  $c$ , so that  $(I - \mathcal{P}_\epsilon)/\epsilon$  can be used to approximate the desired eigenfunctions in (18) even when  $q$  is nonuniform.

Because the Markov kernel  $p_\epsilon$  can be nonsymmetric,  $\mathcal{P}_\epsilon$  is generally not self-adjoint on  $L^2(M, \mu)$ . Instead,  $\mathcal{P}_\epsilon$  is self-adjoint on the space  $L^2(M, \tilde{\mu}_\epsilon)$ , where

$$d\tilde{\mu}_\epsilon = \tilde{q}_\epsilon d\mu, \quad \tilde{q}_\epsilon(a) = \frac{\mathcal{H}_\epsilon 1(a)q(a)}{q_\epsilon(a)}. \quad (23)$$

Specifically, we have

$$\mathcal{P}_\epsilon f(a) = \int_M s_\epsilon(a, b) f(b) d\tilde{\mu}_\epsilon(b), \quad s_\epsilon(a, b) = \frac{K_\epsilon(F(a), F(b))}{\mathcal{H}_\epsilon 1(a)\mathcal{H}_\epsilon 1(b)},$$

and the self-adjointness of  $\mathcal{P}_\epsilon$  on  $L^2(M, \tilde{\mu}_\epsilon)$  follows from the fact that the kernel  $s_\epsilon$  is symmetric. Also, it is straightforward to verify that the density  $\tilde{q}_\epsilon$  is the eigenvector of the adjoint operator  $\mathcal{P}_\epsilon^*$  on  $L^2(M, \mu)$  corresponding to eigenvalue 1; i.e.,  $\tilde{q}_\epsilon$  is the invariant density relative to  $\mu$  for a discrete-time diffusion process on  $M$  generated by  $p_\epsilon$  [37]. Note that  $\tilde{q}_\epsilon = 1 + O(\epsilon)$  and the eigenfunctions of  $\mathcal{P}_\epsilon$  are orthogonal on  $L^2(M, \mu)$  at  $O(\epsilon)$ . In the algorithms described below, we find that including a discrete analog of  $\tilde{q}_\epsilon$  in inner products involving eigenfunctions from diffusion maps leads to a moderate increase in performance.

Turning to the discrete setting, we represent functions on the dataset consisting of the ergodic time series in (1) by  $N$ -dimensional vectors  $\vec{f} = (f_0, \dots, f_{N-1})$  with components  $f_i = f(a_i)$ , and the integral operator  $\mathcal{P}_\epsilon$  by an  $N \times N$  Markov matrix  $P$  such that  $\sum_{j=0}^{N-1} P_{ij} f_j$  approximates  $\mathcal{P}_\epsilon f(a_i)$  for  $\mu$ -a.e.  $a_i \in M$ . In the presence of nonuniform sampling, time averages of the form  $\sum_{i=0}^{N-1} f_i/N$  converge  $\mu$ -a.s. to the integrals  $\int_M f q d\mu$  [37, 65]. Therefore, up to unimportant proportionality constants, we approximate  $q_\epsilon(a_i)$  and  $\mathcal{H}_\epsilon f(a_i)$  by  $\hat{q}_i = \sum_{j=0}^{N-1} K_{ij}$  and  $\sum_{j=0}^{N-1} H_{ij} f_j$ , respectively, where  $K_{ij} = K_\epsilon(F(a_i), F(a_j))$ , and  $H_{ij} = K_{ij}/\hat{q}_j$ . The matrix elements of  $P$  are then given by  $P_{ij} = H_{ij}/\sum_{k=0}^{N-1} H_{ik}$ . The leading left eigenvector of  $P$ ,  $wP = w$ , has elements  $w_i = \sum_{j=0}^{N-1} H_{ij}/\hat{q}_i = K_{ij}/(\hat{q}_i \hat{q}_j)$  approximating (up to a proportionality constant) the continuous function  $\tilde{q}_\epsilon/q$ . Therefore, the weighted inner product

$$\langle \vec{f}_1, \vec{f}_2 \rangle_w = \sum_{i=0}^{N-1} w_i f_{1i} f_{2i} \quad (24)$$

converges  $\mu$ -a.s. to  $\int_M f_1 f_2 d\tilde{\mu}_\epsilon = \langle f_1, f_2 \rangle_{\tilde{\mu}_\epsilon}$  as  $N \rightarrow \infty$ , and moreover  $\langle f_1, f_2 \rangle_{\tilde{\mu}_\epsilon}$  converges to the ergodic inner product  $\langle f_1, f_2 \rangle$  as  $\epsilon \rightarrow 0$ . The limit  $N \rightarrow \infty$  and  $\epsilon \rightarrow 0$  with an appropriate scaling relationship  $\epsilon(N)$  corresponds to the limit of large data [43, 62].

With these ingredients, we approximate the eigenvalues and eigenfunctions of  $\Delta_h$  in (17) using the corresponding eigenvalues and eigenvectors of  $P$ ,

$$P \vec{\phi}_i = \kappa_i \vec{\phi}_i, \quad i \in \{0, 1, \dots, N-1\}, \quad 1 = \kappa_0 < \kappa_1 \leq \kappa_2 \leq \dots \leq \kappa_{N-1}, \quad \vec{\phi}_i = (\phi_{0i}, \dots, \phi_{N-1,i}) \in \mathbb{R}^N. \quad (25)$$

As stated in Section 4.1, to our knowledge there are no pointwise or spectral convergence results for diffusion maps and related kernel algorithms for correlated data generated by ergodic dynamical systems, but

pointwise [37, 43, 62, 65] and spectral [70–73] convergence results are available in the case of i.i.d. data. Hereafter, we will assume that spectral convergence of the operator  $(I - P)/\epsilon$  to  $\Delta_h$  holds in the ergodic case, so that as  $N \rightarrow \infty$  and for a suitable scaling  $\epsilon(N)$  we have  $\phi_{j_i} \rightarrow \phi_j(a_i)$  and  $\log \kappa_i \rightarrow \eta_i$ ,  $\mu$ -a.s. and up to proportionality constants. By convention, we will work with the normalized eigenvalues  $\hat{\eta}_i = \log \kappa_i / \log \kappa_1$ . Geometrically, this scaling is equivalent to a uniform scaling of the Riemannian metric  $h$ , which has no influence on the spectrum of  $\Delta_h$  other than an unimportant global scaling of the eigenvalues.

The discrete eigenvectors from (25) form an orthonormal basis for functions  $\vec{f}$  on the dataset with respect to the inner product in (24), and we have  $\vec{f} = \sum_{i=0}^{N-1} c_i \vec{\phi}_i$  with  $c_i = \langle \vec{\phi}_i, \vec{f} \rangle_w$ . This inner product is asymptotically equivalent as  $\epsilon \rightarrow 0$  to the uniformly-weighted ergodic sum from (4), but as mentioned earlier, in practice we find that the mode decomposition and forecasting algorithms presented below exhibit moderate performance gains by including the weights  $w$ . The analog of the Dirichlet energy in (16) for the discrete eigenfunctions is  $E(\vec{\phi}_i) = \hat{\eta}_i$  (whenever  $\kappa_i > 0$ ), and for the function  $\vec{f}$  above we have  $E(\vec{f}) = \sum_{i=1}^{N-1} \hat{\eta}_i |c_i|^2$ . By our spectral convergence assumption,  $E(\vec{f})$  converges up to a proportionality constant to its continuous counterpart,  $E_h(f)$ . Hereafter, whenever there is no risk of confusion with the continuous case we will omit hats and overarrows in our notation for quantities computed for the discrete dataset, such as  $\vec{\phi}_i$  and  $\hat{\eta}_i$ . The construction of our data-driven orthonormal basis of  $L^2(M, \mu)$  is summarized in Algorithm 1 in Appendix A.

#### 4.3. Spectral Galerkin method

We solve the eigenvalue problem for the Koopman generator  $v$  in weak form in the basis of  $H_1(M, \mu)$  in (18). Because the set of eigenvalues is dense on the imaginary line (see Remark 3), we first regularize the problem by adding a small amount of diffusion to  $v$  to form the operator

$$L_\epsilon = v - \epsilon \Delta_h. \quad (26)$$

Qualitatively, the effect of the diffusion term  $\epsilon \Delta_h$  is to suppress highly oscillatory eigenfunctions from the spectrum, which, as stated in Remark 9, can have arbitrarily small frequencies (eigenvalues). With these pathological modes suppressed from the spectrum, we identify the generating frequencies and eigenfunctions by ordering the computed eigenfunctions in order of increasing Dirichlet energy. We will return to a discussion of the effects of diffusion in the spectral properties of  $L_\epsilon$  in Section 4.4.

We solve the eigenvalue problem for  $L_\epsilon$  using spectral Galerkin methods for elliptic eigenvalue problems [77]. To pass from the strong form of the problem,

$$L_\epsilon u = \gamma u, \quad (27)$$

to the weak form, we multiply the eigenvalue equation by a test function  $\psi \in C^\infty(M)$ , integrate by parts with respect to the invariant measure, and require that the resulting integral equation is satisfied for all elements of appropriate trial and test spaces, which we both take to be  $H_1(M, \mu)$ .

**Definition 13 (Eigenvalue problem for  $L_\epsilon$ , weak form).** Find  $\gamma \in \mathbb{C}$  and  $u \in H_1(M, \mu)$  such that for any  $\psi \in H_1(M, \mu)$ ,

$$\mathcal{A}(\psi, u) = \gamma \mathcal{B}(\psi, u),$$

where  $\mathcal{A}$  and  $\mathcal{B}$  are sesquilinear forms on  $H_1(M, \mu) \times H_1(M, \mu)$  given by

$$\mathcal{A}(\psi, u) = \langle \psi, v(u) \rangle - \epsilon \langle \text{grad}_h \psi, \text{grad}_h u \rangle, \quad \mathcal{B}(\psi, u) = \langle \psi, u \rangle.$$

In the discrete formulation of the problem, we formally restrict the trial and test spaces to the  $n$ -dimensional subspaces  $H_{1,n} = \text{span}\{\varphi_0, \dots, \varphi_{n-1}\} \subset H_1(M, \mu)$  spanned by the basis functions in (18); i.e., we have  $\psi = \sum_{i=0}^{n-1} d_i \varphi_i$  and  $u = \sum_{i=0}^{n-1} c_i \varphi_i$ , where  $c = (c_0, \dots, c_{n-1})$  and  $d = (d_0, \dots, d_{n-1})$  are complex-valued expansion coefficients. However, because instead of the true Laplace-Beltrami eigenfunctions we only have access to the approximate eigenfunctions from Algorithm 1, and furthermore we do not have access to the exact generator  $v$ , we additionally make the following approximations for the evaluation of the sesquilinear forms in the continuous problem.

1. We approximate the basis function values on the dataset by the vectors  $\vec{\varphi}_i = (\varphi_{0i}, \dots, \varphi_{N-1,i}) \in \mathbb{R}^N$  with  $\varphi_{ji} = \phi_{ji}/\hat{\eta}_i \approx \varphi_i(a_j)$ , where  $\{\vec{\phi}_i\}$  and  $\{\hat{\eta}_i\}$  are the eigenvectors and eigenvalues from Algorithm 1.
2. In the case of  $\mathcal{D}$ , we put

$$\mathcal{D}(\psi, u) = \sum_{i,j=0}^{n-1} \int_M d_i^* c_j \text{grad}_h \varphi_i \cdot \text{grad}_h \varphi_j d\mu = d^\dagger D c, \quad (28a)$$

where  $D$  is the  $n \times n$  identity matrix since  $D_{ij} = \langle \text{grad}_h \varphi_i, \text{grad}_h \varphi_j \rangle = \delta_{ij}$ .

3. In the case of  $\mathcal{B}$ , we put

$$\mathcal{B}(\psi, u) = \sum_{i,j=0}^{n-1} \int_M d_i^* c_j \varphi_i \varphi_j d\mu \approx \sum_{i,j=0}^{n-1} d_i^* c_j \langle \vec{\varphi}_i, \vec{\varphi}_j \rangle_w = d^\dagger B c, \quad (28b)$$

where  $B$  is the  $n \times n$  diagonal matrix with the diagonal entries  $B_{ii} = 1 + \hat{\eta}_i^{-1/2}$ .

4. In the case of  $\mathcal{V}$ , we proceed similarly as with  $\mathcal{D}$  and  $\mathcal{B}$ , but we also approximate the action of the Koopman generator  $v(\varphi_i)$  on the basis elements using finite differences in time as stated in Remark 1. Hereafter, we will use a second-order central scheme for the sampling interval  $T$ , and make the approximation  $v(\varphi_i(a_j)) \approx (\varphi_{i,j+1} - \varphi_{i,j-1})/(2T)$ . We therefore set

$$\mathcal{V}(\psi, u) = \sum_{i,j=0}^{n-1} \int_M d_i^* c_j \varphi_i v(\varphi_j) d\mu \approx d^\dagger V c, \quad (28c)$$

where  $V$  is the  $n \times n$  matrix with elements  $V_{ij} = \sum_{k=1}^{N-2} \varphi_{ik} w_k (\varphi_{i,k+1} - \varphi_{i,k-1})/(2T)$ .

With these approximations, we define:

**Definition 14 (Eigenvalue problem for  $L_\epsilon$ , discrete approximation).** Find  $\lambda \in \mathbb{C}$  and  $c \in \mathbb{C}^n$  such that for any  $d \in \mathbb{C}^n$ ,

$$\hat{A}(d, c) = \gamma \hat{B}(d, c),$$

where  $\hat{A}$  and  $\hat{B}$  are sesquilinear forms on  $\mathbb{C}^n \times \mathbb{C}^n$  given by

$$\hat{A}(d, c) = d^\dagger A c, \quad A = V - \epsilon D, \quad \hat{B}(d, c) = d^\dagger B c.$$

The solution to the discrete problem is given by the generalized eigenvalue problem

$$A c = \lambda B c. \quad (29)$$

**Remark 15.** The  $\{\varphi_i\}$  basis from (18) and its discrete counterpart  $\{\vec{\varphi}_i\}$  are adapted to the  $H_1$  regularity of the eigenvalue problem for  $L_\epsilon$  in the sense that the highest-order sesquilinear form  $\mathcal{D}$  appearing in the weak formulation of the problem is represented by the identity matrix,  $D = I$ . This property ensures that the scheme remains well-conditioned at large spectral order of approximation  $n$ . In contrast, the condition number of the  $D$  matrix would exhibit unbounded growth with  $n$  if we were to work in the unscaled eigenfunction basis. This approach of tailoring the approximation basis to the Sobolev regularity of the continuous problem is sometimes used in spectral Galerkin methods with polynomial basis functions (e.g., [78, 79]).

The numerical solution of the generalized eigenvalue problem in (29) yields  $n' \leq n$  (depending on the numerical algorithm used) eigenvalue-eigenvector pairs  $\{(\gamma_k, c_k)\}_{k=0}^{n'-1}$  with  $c_k = (c_{0k}, \dots, c_{n-1,k}) \in \mathbb{C}^n$  and the corresponding discretely sampled eigenfunctions  $u_k = \sum_{i=0}^{n-1} \varphi_i c_{ik}$  with  $u_k = (u_{0k}, \dots, u_{N-1,k}) \in \mathbb{C}^N$ . Throughout, we work with the normalization  $\|u_k\|_w = 1$  from (24), which approximates the normalization  $\|u_k\| = 1$  on  $L^2(M, \mu)$  in the continuous case.

To identify a generating set  $\{\zeta_i\}_{i=1}^m$  for the eigenfunctions and the corresponding basic frequencies  $\{\Omega_i\}_{i=1}^m$  using the approach of Section 3.1, we first order the eigenfunctions in order of increasing Dirichlet energy, which we compute from the discrete analog of (19),  $E(z_k) = \sum_{i=1}^m |c_{ik}|^2 = \|c_k\|^2$ . Then, we select  $\{\zeta_i\}$  and  $\{\Omega_i\}$  from the first  $m$  nonconstant eigenfunctions in this set with “numerically rationally independent” eigenvalues up to some precision. Operationally, we declare  $\Omega_i$  and  $\Omega_j$  to be rationally independent at precision  $(\delta, \bar{q})$  if there exist no integers  $q_i$  and  $q_j$  with absolute value smaller than  $\bar{q}$  such that  $|q_j \Omega_i - q_i \Omega_j| \leq \delta$ . In practice, it is usually easy to identify rationally independent frequencies from the first few numerical eigenvalues manually. The numerical procedure to compute the generating frequencies and eigenfunctions is summarized in Algorithm 2 in Appendix A.

Using the identified generating sets, we form the product bases  $\{z_k\}$  and associated frequencies  $\{\omega_k\}$  as described in Definition 2. Note that the numerical generators  $\zeta_i = (\zeta_{0i}, \dots, \zeta_{N-1,i}) \in \mathbb{C}^N$  will not lie exactly on the unit circle, and this happens because of sampling errors (i.e., errors that vanish as  $N \rightarrow \infty$ ), but also bias errors in the Markov operator  $\mathcal{P}_\epsilon$  in diffusion maps and the regularized generator  $L_\epsilon$  occurring at nonzero  $\epsilon$ . We therefore rescale the generators pointwise to ensure that  $|\zeta_{ij}| = 1$  before taking powers. Experimentally, we find that generating the spectrum recursively from the generators produces more accurate eigenfunctions than the raw solutions of (29), presumably because the accuracy of the data-driven basis degrades more rapidly at high eigenfunctions than the loss of accuracy in the  $\{\zeta_i\}$  products.

#### 4.4. Spectral properties of the regularized generator

The spectral properties of the regularized generator  $L_\epsilon$  from (26) are compounded by the facts that (1) the limit  $\epsilon \rightarrow 0$  is a singular limit of the corresponding eigenvalue equation; (2) apart from special cases (e.g., irrational flows on flat tori), the operators  $v$  and  $\Delta_h$  do not commute, and as result  $L_\epsilon$  is nonnormal with

$$[L_\epsilon^*, L_\epsilon] = 2\epsilon[v, \Delta_h]. \quad (30)$$

If  $v$  has smooth eigenfunctions, one can confirm with asymptotic expansions that the influence of the diffusion term on these eigenfunctions is benign. In particular, writing  $u_k = z_k + \epsilon u'_k + O(\epsilon^2)$  and  $\gamma_k = \lambda_k + \epsilon \gamma'_k + O(\epsilon^2)$  with  $v(z_k) = \lambda_k z_k$ , and inserting these asymptotic series in the eigenvalue equation (27), we obtain the  $O(\epsilon)$  equation

$$(v - \lambda_k)u'_k = (\gamma'_k - \Delta_h)z_k.$$

This equation can be solved by requiring that  $u'_k$  is orthogonal to  $z_k$  as a solvability condition, i.e.,  $u'_k = \sum_{i \neq k} c_{ik} z_i$ , giving

$$c_{ik} = \frac{\langle \text{grad}_h z_i, \text{grad}_h z_k \rangle}{\lambda_i - \lambda_k}, \quad \gamma'_k = \langle \text{grad}_h z_k, \text{grad}_h z_k \rangle = E_h(z_k). \quad (31)$$

We therefore see that, at  $O(\epsilon)$ , the diffusion term perturbs the eigenvalues of  $v$  corresponding to smooth eigenfunctions by a purely real term equal to the Dirichlet energy of the unperturbed eigenfunctions. This provides a more quantitative estimate of the suppression of highly-oscillatory eigenfunctions of  $v$  from the spectrum of  $L_\epsilon$  claimed in Section 4.3. Note that the imaginary part of the perturbation to  $\lambda_k$ , which is important for nonparametric forecasting, occurs at  $O(\epsilon^2)$ . It also follows from these results that the non-orthogonality of the eigenfunctions of  $L_\epsilon$  can be estimated by  $\langle u_i, u_k \rangle = \epsilon c_{ik} + O(\epsilon^2)$ , and the coefficients  $c_{ik}$  vanish if  $z_i$  and  $z_k$  are eigenfunctions of both  $v$  and  $\Delta_h$ .

More generally, it follows from the spectral theorem and (30) that orthogonality and completeness of the eigenfunctions of  $L_\epsilon$  is guaranteed at all orders in  $\epsilon$  if  $v$  and  $\Delta_h$  are commuting operators. One situation that this occurs is when the dynamical flow  $\Phi_t$  preserves the Riemannian metric  $h$ , i.e.,  $v$  generates  $h$ -isometries. A necessary and sufficient condition for  $v$  to generate isometries is that  $v$  is a Killing vector field satisfying the equation  $\mathcal{L}_v h = 0$ , where  $\mathcal{L}_v$  is the Lie derivative on type  $(0, 2)$  tensors with respect to  $v$ . One can check that if this equation is satisfied then  $[v, \Delta_h] = 0$ . Moreover, when  $v$  generates isometries of a smooth Riemannian metric, then it must necessarily have a pure point spectrum [42, §7.1.c]. This is because the group of diffeomorphisms of a Riemannian manifold is a compact Lie group containing  $\{\Phi_t\}$  as an Abelian subgroup, and translations on compact Abelian groups have pure point spectra (see Section 2).

A particularly important property that holds if  $[v, \Delta_h] = 0$  is that  $v$  and  $\Delta_h$  have joint eigenfunctions. This means that (1) the eigenvalue problem for  $L_\epsilon$  yields exact eigenfunctions of  $v$ ; (2) the Galerkin method for approximating these eigenfunctions becomes highly efficient in the eigenbasis of  $\Delta_h$ ; (3) the eigenfunctions of  $v$  are extrema of the Rayleigh quotient  $R_h$  for the metric  $h$ . In particular, (3) implies that the leading generating eigenfunction  $\zeta_1$  identified with respect to the Dirichlet energy  $E_h(\zeta_1)$  lies entirely in the subspace of  $L^2(M, \mu)$  spanned by eigenfunctions of  $\Delta_h$  corresponding to its smallest nonzero eigenvalue. These facts suggest that for a system with pure point spectrum it would be preferable to regularize the generator with diffusion in a Riemannian metric  $h$  preserved by the dynamics. In Section 5, we will present an approach for approximating such a metric using delay-coordinate maps, but for the rest of this Section we will continue to work with  $h$  as defined in Section 4.1. As expected from the asymptotics (and verified in the experiments in Section 4.6 ahead), the effects of noncommutativity of  $v$  and  $\Delta_h$  should have minimal impact on the quality of the numerical eigenfunctions in this case. Moreover, it is important to establish that accurate Koopman eigenfunctions can be computed without having to perform delay-coordinate maps.

What about the behavior of  $L_\epsilon$  in more general ergodic systems where the generator also has a continuous spectrum, and in particular in weak-mixing and mixing systems where it has no nonconstant eigenfunctions? Mathematically,  $L_\epsilon$  has the same structure as a class of advection-diffusion operators arising in incompressible fluid dynamics on compact manifolds (recall that  $v$  generates an incompressible flow with respect to the invariant measure), for which theoretical results are available in the literature [80–82]. The latter references study the dynamical and spectral properties of operators of the form  $\tilde{L}_\alpha = \alpha v + \Delta$  in the advection dominated regime,  $\alpha \rightarrow \infty$ , and these operators are equivalent to  $\epsilon L_\epsilon$  with  $\epsilon = 1/\alpha$ . Franke et al. [82] show that  $\tilde{L}_\alpha$  (and hence  $L_\epsilon$ ) has no continuous spectrum irrespective of the mixing properties of  $v$ . For our purposes, this is a positive result as it eliminates an important source of numerical instability in the discrete problem. However, the physical significance of the eigenfunctions of  $L_\epsilon$  (and their utility in the dimension reduction and forecasting schemes of Sections 3 and 4) becomes questionable, especially in weak-mixing systems where  $v$  has no nonconstant eigenfunctions.

In [81], Constantin et al. establish necessary and sufficient conditions for  $\tilde{L}_\alpha$  to have the so-called relaxation-enhancing property as  $\alpha \rightarrow \infty$ , i.e., the property that the associated Kolmogorov semigroup (the advection-diffusion analog of the Koopman group) will produce relaxation of any function in  $L^2(M, \mu)$  to its mean value in arbitrarily small time. They show that  $\tilde{L}_\alpha$  is relaxation enhancing if and only if  $v$  has no eigenfunctions in  $H_1(M, \mu)$ ; a condition that includes but is not necessarily limited to weak-mixing systems. The spectral manifestation of the relaxation-enhancing property is that the spectral gap of  $\tilde{L}_\alpha$  diverges as  $\alpha \rightarrow \infty$  [82], meaning that  $L_\epsilon = \epsilon \tilde{L}_{1/\epsilon}$  will generally exhibit complicated spectral behavior, including the possibility of no eigenvalues with negligible real parts at arbitrarily small  $\epsilon$ . On the other hand, if  $v$  has eigenfunctions in  $H_1(M, \mu)$ , then  $L_\epsilon$  must necessarily have a vanishing spectral gap, and the asymptotic behavior described earlier applies. Due to these considerations, rather than attempting to regularize weak-mixing systems by diffusion alone, in Section 6 we will put forward an alternative approach which also involves a time change [55], implemented with variable bandwidth kernels of the same class as (20), attempting to reduce the mixing properties of the system while preserving its orbits.

#### 4.5. Numerical implementation of the dimension reduction and forecasting schemes

With the availability of the approximate generators and the corresponding basic frequencies from Algorithms 1 and 2, algorithms for dimension reduction, vector field decomposition, and forecasting of densities in systems with pure point spectra can be constructed following closely the continuous formulation in Sections 3.1 and 3.3, replacing inner products on  $L^2(M, \mu)$  with the  $\pi$ -weighted inner products from (24), and representing the action of the generator through finite differences of the diffusion maps basis as described in Section 4.3. An additional issue that needs to be taken into account is the non-orthogonality of the eigenfunctions of the regularized generator discussed in Section 4.4. In particular, the lack of orthogonality of the eigenfunctions leads to a modification of the inverse transforms to reconstruct observables and vector fields. For an observable  $f$  with expansion coefficients  $\hat{f}_k = \langle f, z_k \rangle_w$  we now have  $f = \sum_k \tilde{f}_k z_k$  (as opposed to  $f = \sum_k \hat{f}_k z_k$ ), where  $\tilde{f}_k = \sum_i G_{ki}^{-1} \hat{f}_i$ , and  $G^{-1}$  is the inverse of the Gramm matrix  $G_{ij} = \langle z_i, z_j \rangle_w$ . This matrix will generally fail to be an identity matrix due to both sampling and bias errors, but in applications

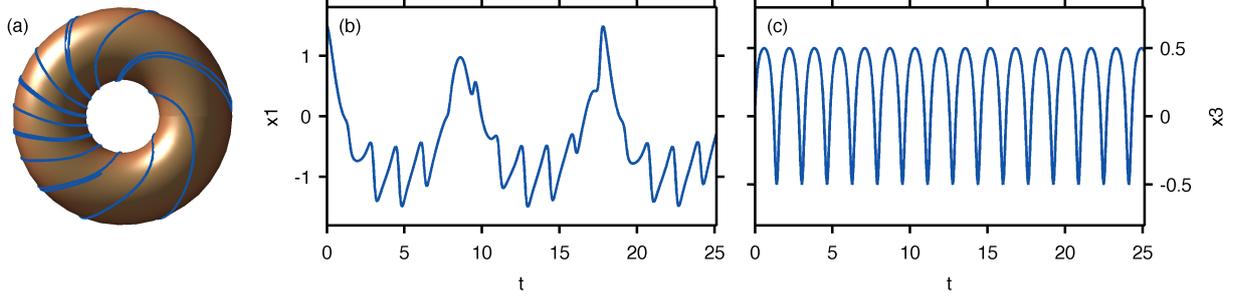


Figure 3: Observed time series for the variable-speed dynamical system on the 2-torus. (a) Embedding in  $\mathbb{R}^3$ ; (b, c) components  $x^1$  and  $x^3$  of the embedding, respectively.

we find that it is a well conditioned, sparse matrix. Our algorithms for vector field decomposition and nonparametric forecasting are listed in Algorithms 3 and 4 in Appendix A, respectively.

#### 4.6. Applications to variable-speed flows on the 2-torus

We apply the dimension reduction, vector field decomposition, and nonparametric forecasting techniques described above to the dynamical system on  $\mathbb{T}^2$  with the generator (cf. (12))

$$v = \sum_{\mu=1}^2 v^\mu \frac{\partial}{\partial \theta^\mu}, \quad v^1 = 1 + (1 - \beta)^{1/2} \cos \theta^1, \quad v^2 = \alpha(1 - (1 - \beta)^{1/2} \sin \theta^2), \quad (32)$$

where  $\theta^1, \theta^2 \in [0, 2\pi)$  are the two angles on the torus,  $\alpha$  is a frequency parameter set to an irrational number, and  $\beta \in (0, 1]$  is a parameter controlling the speed of the flow. This system, which was also studied in [58], is ergodic and has pure point spectrum as it can be transformed to an irrational flow with appropriate changes of coordinates. Its unique invariant measure has density

$$\sigma(\theta^1, \theta^2) \propto 1/[(1 + (1 - \beta)^{1/2} \cos(\theta^1))(1 - (1 - \beta)^{1/2} \sin \theta^2)] \quad (33)$$

relative to the Haar measure on  $\mathbb{T}^2$ , and the orbit  $(\dot{\theta}^1, \dot{\theta}^2) = (v^1, v^2)$  passing through  $(\theta^1, \theta^2) = (0, 0)$  at time  $t = 0$  is given by

$$\tan(\theta^1/2) = [1 + (1 - \beta)^{1/2}] \beta^{-1/2} \tan(\beta t/2), \quad \cot(\theta^2/2) = (1 - \beta)^{1/2} + \beta^{1/2} \cot(\beta^{1/2} \alpha t/2). \quad (34)$$

Here, we use the same frequency parameter,  $\alpha = 30^{1/2}$ , as in the irrational flow example of Section 3.4, and set the parameter  $\beta$  to  $1/2$ . We also use the same observation map as in Section 3.4, i.e., the standard embedding of the 2-torus into  $\mathbb{R}^3$  from (13) with radius parameter  $R = 1/2$ . As illustrated in Fig. 3, with this choice of parameters the system “slows down” at  $(\theta^1, \theta^2) \sim (\pi, \pi/2)$  and “speeds up” at  $(\theta^1, \theta^2) \sim (0, -\pi/2)$ , resulting in a large density contrast  $\sigma(\pi, \pi/2)/\sigma(0, -\pi/2) \simeq 34$  relative to the Haar measure. Moreover, the system exhibits two timescales which are mixed together in the components of the nonlinear observation map. Despite the apparent complexity of these time series, the underlying dynamical system is completely integrable and our method is able to detect this structure using no information other than time-ordered data.

In these experiments, we used as training data a time series consisting of 64,000 samples  $\{x_i\}$  taken from the orbit in (34) at a timestep  $T = 2\pi/500$ . Using this data, we computed 1000 diffusion eigenvalues and eigenfunctions  $\{(\eta_i, \varphi_i)\}$  through Algorithm 1, and constructed the generating set  $\{\zeta_1, \zeta_2\}$  and the corresponding basic frequencies  $\{\Omega_1, \Omega_2\}$  using Algorithm 2 with  $\epsilon = 3 \times 10^{-4}$ . We use the true manifold dimension,  $m = 2$ , as an input to these algorithms, but in this case an accurate dimension estimate can also be obtained from the kernel density estimation procedure in Algorithm 1. The eigenvalues of the approximate generator  $L_\epsilon$  corresponding to the basic frequencies are  $\gamma_1 = 0.0005 + 0.707i$  and  $\gamma_3 = 0.0016 + 3.871i$  ( $\gamma_2$  is

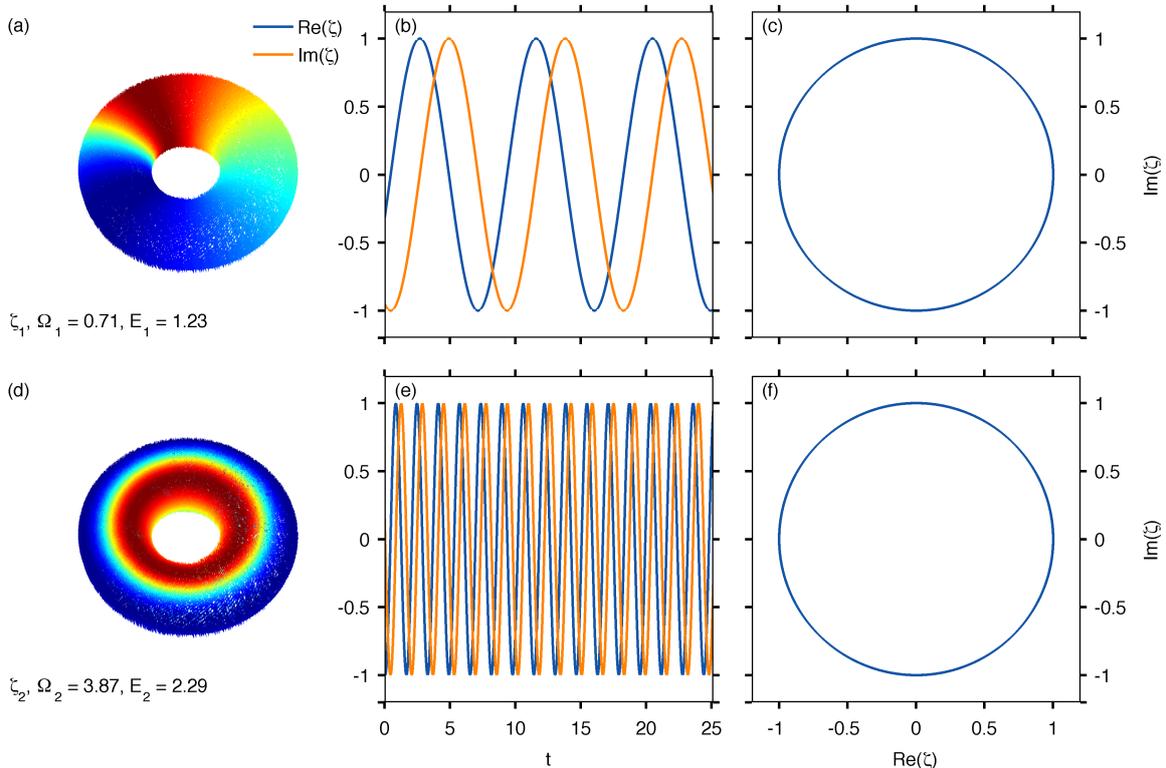


Figure 4: The generating Koopman eigenfunctions  $\zeta_1$  (a–c) and  $\zeta_2$  (d–f) for the variable-speed dynamical system on the 2-torus. (a, d) Scatterplots of  $\text{Re}(\zeta_i)$  on the torus; (b, e) time series of  $\text{Re} \zeta_i$  and  $\text{Im} \zeta_i$ ; (c, f) scatterplots of  $(\text{Re} \zeta_i, \text{Im} \zeta_i)$ . The eigenfunctions shown here have not been rescaled to the unit circle—they are the result of Step 5 of Algorithm 2.

the complex conjugate of  $\gamma_1$ ), and  $\Omega_i = \text{Im} \gamma_i$ . According to our convention, these eigenvalues are ordered in order of increasing Dirichlet energy; in this case  $E_1 = 1.52$  and  $E_3 = 5.34$ , and the computed eigenvalues have  $\text{Re} \gamma_1/\epsilon = 1.51 \approx E_1$  and  $\text{Re} \gamma_3/\epsilon = 5.34 \approx E_2$ , which is in good agreement with the asymptotic result in (31).

The generating eigenfunctions from these calculations are displayed in Fig. 4. There, it is evident that the projection maps from (6) based on the eigenfunctions send the data set to a near-exact circle in the complex plane, and the time series  $\zeta_i(t)$  from (7) describe simple harmonic oscillations. The frequencies of these oscillations (measured, e.g., through FFT) are in very good agreement with  $\Omega_i$ . As illustrated by the scatterplots in Fig. 4(a, b), eigenfunctions  $\zeta_1$  and  $\zeta_2$  vary purely along angles  $\theta^1$  and  $\theta^2$ , respectively. This is an outcome of the fact that in the induced Riemannian geometry from  $F$  these eigenfunctions have the smallest Dirichlet energies, but in other geometries we could have mixtures of the form  $\zeta_1^{q_1} \zeta_2^{q_2}$  appearing as the least oscillatory eigenfunctions. Compared to their counterparts for the irrational flow in Section 3.4 (which are pure sinusoids in the  $\theta^1$  and  $\theta^2$  angles), the generating eigenfunctions for the variable-speed flow exhibit “compressed” (“expanded”) waveforms in the regions of anomalously high (low) phase-space speed. As discussed in the SOM, the results in Fig. 4 are in good agreement with results obtained via EDMD using a dictionary of functions consisting of time-lagged components of the observation vector in  $\mathbb{R}^3$ .

With these results for the generating eigenfunctions and basic frequencies, we use Algorithm 3 to decompose  $v$  into the mutually commuting vector fields  $\{v_i\}_{i=1}^2$  from Theorem 5, and reconstruct these vector fields in data space. Figure 5 shows the reconstructions obtained with a spectral order parameter  $l = 30$ , corresponding to  $n = (2l + 1)^2 = 3721$  eigenfunctions. The vector fields exhibit the qualitative features expected from the corresponding generating eigenfunctions in Fig. 4, namely the reconstructed vector fields,  $V_1$  and  $V_2$ , describe flows purely along the  $\theta^1$  and  $\theta^2$  directions, respectively. As discussed in Section 3.2,

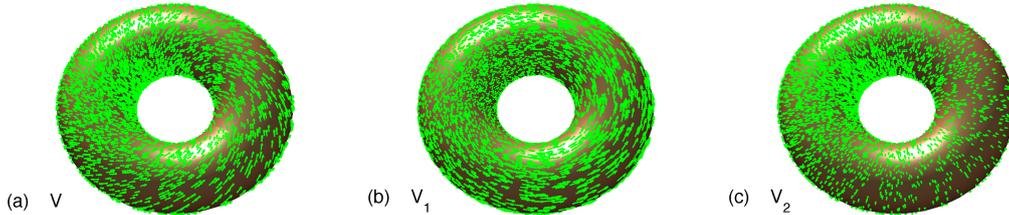


Figure 5: Vector field decomposition for the variable-speed dynamical system on the 2-torus. (a) Full vector field embedded in  $\mathbb{R}^3$ ; (b, c) mutually commuting components from Theorem 5 corresponding to the generators in Fig. 4, reconstructed in  $\mathbb{R}^3$  via the pushforward map in (8).

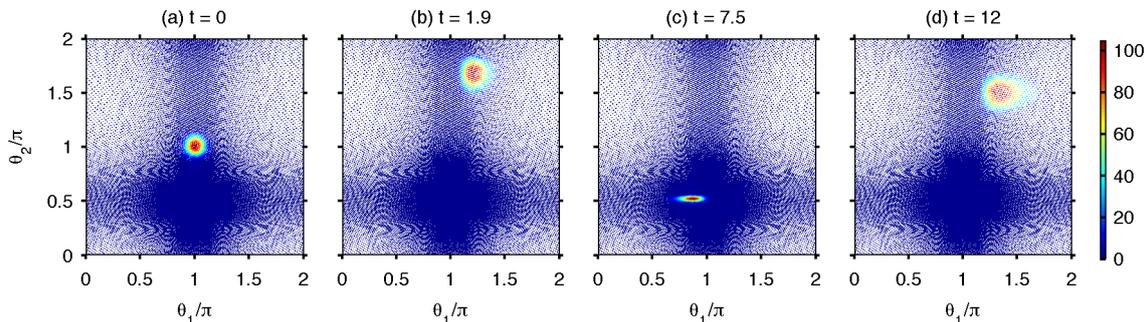


Figure 6: Time-dependent density  $\rho_t$  for the variable-speed dynamical system on the 2-torus relative to its equilibrium measure computed via the nonparametric method in Algorithm 4. (a) Initial von Mises density with location and concentration parameters  $(\pi, \pi)$  and 30, respectively; (b–d) snapshots of  $\rho_t$  illustrating phases of high uncertainty (b, d) and a phase of low uncertainty (c). The dynamic evolution of  $\rho_t$  is also shown in Movie 1.

these flows preserve the equilibrium measure of the dynamics (but not the Riemannian measure of the torus embedded in  $\mathbb{R}^3$ ), and are clearly not ergodic— $v_1$  ( $v_2$ ) has a non-trivial nullspace spanned by functions with no  $\theta^2$  ( $\theta^1$ ) dependence. With this large number of basis functions, the time-averaged (root mean square) reconstruction error in  $V_1 + V_2$  is 0.2% of the time-averaged norm of the full vector field  $V$  in  $\mathbb{R}^3$ . Note that good-quality reconstructions are possible with substantially fewer basis functions (e.g.,  $l = 5$ ,  $n = 121$  yields a 5.9% error). Moreover, the expansion coefficients of the  $V_i$  in the  $\{z_k\}$  basis exhibit strong sparsity, suggesting that the reconstructed patterns can be efficiently compressed using subsets of the  $\{z_k\}$  basis as dictionaries.

Next, we consider statistical forecasting of the components  $F^1(a) = x^1$  and  $F^3(a) = x^3$  of the state vector in  $\mathbb{R}^3$  using the nonparametric technique in Algorithm 4. As with the irrational-flow example in Section 3.4, we assign an initial probability measure  $\mu_0$  with a von Mises density relative to the Haar measure. In this case, we set the location and concentration parameters  $(\bar{\theta}^1, \bar{\theta}^2) = (\pi, \pi)$  and  $\kappa = 30$ , respectively. Note that the equilibrium measure  $\mu$  of the variable-speed system differs from the Haar measure on the torus, so the initial density  $\rho_0 = d\mu_0/d\mu$  is given by the density in (14) divided by  $\sigma$  from (33). We normalize the initial density on the dataset so that  $\sum_{i=0}^{N-1} \rho_{0,i}/N = 1$ , where  $\rho_{0,i} = \rho_0(a_i)$ . A scatterplot of the initial density on the torus is shown in Fig. 6(a). We advance this density and the corresponding expectation values and standard deviations for  $x^1$  and  $x^3$  for the lead times  $\{t_i\}_{i=1}^{1000}$  with  $t_i = iT$  using Algorithm 4. Figure 6 shows representative snapshots of the time-dependent density for these lead times, which is also visualized as a video in Movie 1. Figure 7 displays the means and standard deviations of  $x^1$  and  $x^3$  obtained via Algorithm 4 and an ensemble forecast with the perfect model using 10,000 samples drawn independently from  $\mu_0$ .

Following an initial transient stage, the probability density in Fig. 6 relaxes to an aperiodic pattern characterized by alternating phases of high (Fig. 6(b, d)) and low (Fig. 6(b)) uncertainty, corresponding to

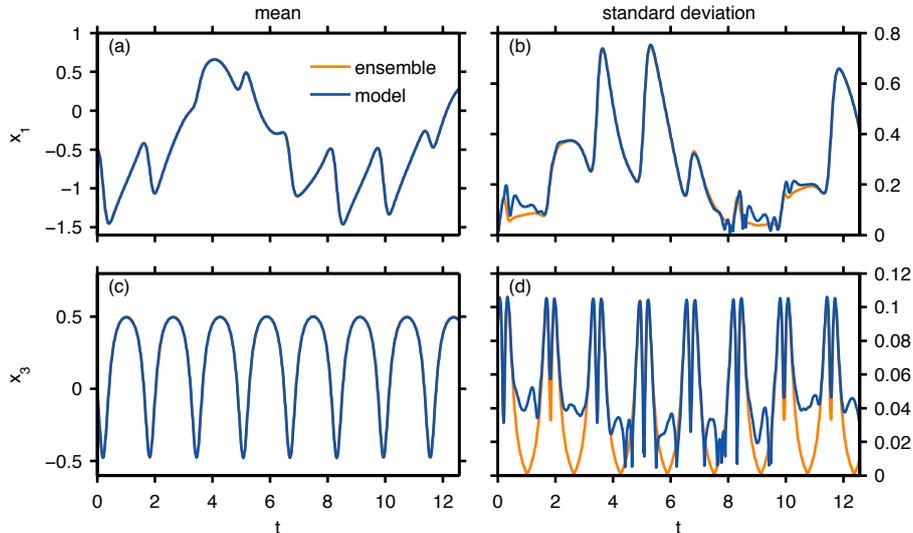


Figure 7: Statistical forecasts via the nonparametric method in Algorithm 4 and a 10,000-member ensemble from the perfect model for the variable-speed system on the 2-torus. The forecast observables are the components  $x^1$  (a, b) and  $x^2$  (c, d) of the embedding of the 2-torus into  $\mathbb{R}^3$  from (13) with radius parameter  $R = 1/2$ . The initial probability measure has the circular density in Fig. 6(a).

the regions with high speed (low density relative to the Haar measure) and low speed (high density relative to the Haar measure), respectively. Qualitatively, the time-dependent measure will eventually explore the full phase space on the torus by ergodicity, but because the system is not mixing, the density  $\rho_t$  retains its coherence and its Dirichlet energy is bounded in time. The results in Fig. 7(a, c) indicate that the mean forecast with the nonparametric model is in very good agreement with the ensemble forecast. The standard deviation forecast in Fig. 7(b, d) is also in good agreement with the ensemble, but in this case the nonparametric model overestimates the standard deviation in the low-uncertainty periods (especially for  $x^3$ ). A factor that may be contributing to this discrepancy is that during those periods the density  $\rho_t$  exhibits small-scale behavior and correspondingly large bandwidth in the  $\{z_k\}$  basis, limiting the accuracy of our finite- $l$  truncation.

## 5. Galerkin method with delay-coordinate maps for systems with pure point spectra

Despite the attractive numerical results in Section 4.6, in Section 4.4 we saw that due to non-commutativity of  $v$  and  $\Delta_h$ , the advection-diffusion operator  $L_\epsilon$  constructed via (26) may not be optimal for approximating Koopman eigenfunctions. In this Section, we present an approach for spectral decomposition of systems with pure point spectra that uses delay-coordinate maps to construct a diffusion operator that commutes with the Koopman group. This approach leads to a highly efficient Galerkin scheme for solving the Koopman eigenvalue problem, and provides a natural way of denoising data corrupted by i.i.d. observational noise.

### 5.1. Relationship between Koopman and Laplace-Beltrami operators in delay-coordinate space

Delay-coordinate maps [45–49] were originally developed as a state-space reconstruction technique that maps a time-ordered signal into a higher-dimensional space of sequences where, under mild assumptions, the attractor of the dynamical system generating the data is recovered. Fixing an integer parameter  $s$  (the number of delays), we replace the observed time series  $\{x_i\}_{i=0}^{N-1}$  in  $\mathbb{R}^d$  by the time series  $\{X_i\}_{i=s-1}^{N-1}$ , where  $X_i = (x_i, x_{i-1}, \dots, x_{i-s+1}) \in \mathbb{R}^{sd}$ . This procedure implicitly defines a new observation map,  $F_s : M \mapsto \mathbb{R}^{sd}$ , with

$$F_s(a_i) = (F(a_i), F(\hat{\Phi}_{i-1}(a_i)), \dots, F(\hat{\Phi}_{i-s+1}(a_i))), \quad (35)$$

where  $F$  and  $\hat{\Phi}_i$  are the observation map and discrete-time flow map introduced in Section 2, respectively. (Recall that the dynamical system is assumed to be invertible so we can evaluate  $\hat{\Phi}_k$  for both positive and negative  $k$ .) Here, we are interested in the behavior of diffusion maps for this class of observation maps at large numbers of delays  $s$ .

First, we examine the Riemannian metric  $g_s$  induced on  $M$  via the embedding in (35) when  $R^{sd}$  is equipped with the “time-average” inner product  $\langle X_i, X_j \rangle_s = \sum_{k=0}^{s-1} \langle x_{i-k}, x_{j-k} \rangle / s$ , where  $\langle \cdot, \cdot \rangle$  denotes the canonical inner product on  $\mathbb{R}^d$ . The Riemannian inner product of two tangent vectors  $u_1, u_2 \in T_a M$  with respect to this metric becomes

$$g_s(u_1, u_2) = \frac{1}{s} \sum_{k=0}^{s-1} g(\hat{\Phi}_{-k*} u_1, \hat{\Phi}_{-k*} u_2) = \frac{1}{s} \sum_{k=0}^{s-1} \langle F_* \hat{\Phi}_{-k*} u_1, F_* \hat{\Phi}_{-k*} u_2 \rangle, \quad (36)$$

where  $g$  is the induced Riemannian metric associated with  $F$ , and  $\hat{\Phi}_{k*} : TM \mapsto TM$  and  $F_* TM \mapsto T\mathbb{R}^d$  are the pushforward maps on tangent vectors associated with  $\hat{\Phi}_k$  and  $F$ , respectively. The limiting behavior of the sequence of smooth Riemannian metrics  $g_1, g_2, \dots$  is controlled by the spectrum of Lyapunov exponents of the system, and for systems with nonzero Lyapunov exponents  $g_s$  will generally fail to converge to a smooth tensor. In particular, according to Oseledets’ multiplicative ergodic theorem [83, 84], there exists a splitting of the tangent bundle into the direct sum  $TM = \bigoplus_{i=1}^m E_i$  where the subspaces  $E_i$  are invariant under the flow,  $\hat{\Phi}_{k*} E_i = E_i$ , and for  $\mu$ -a.e.  $a \in M$  and  $u \in E_i|_a \setminus 0$ ,

$$\lim_{k \rightarrow \infty} \frac{1}{2k} \log (g(\hat{\Phi}_{-k*} u, \hat{\Phi}_{-k*} u)) = -\Lambda_i,$$

where  $\Lambda_i$  is the Lyapunov exponent corresponding to  $E_i$ . This means that if  $\Lambda_i < 0$ , the inner product  $g_s(u, u)$  from (36) will exhibit exponential growth with  $s$ , and similarly if  $\Lambda_i > 0$   $g_s(u, u)$  will converge to zero. Thus, in systems with nonzero Lyapunov exponents the limit metric  $\lim_{s \rightarrow \infty} g_s$  will be non-smooth. Note that had we used forward-looking instead of backward-looking delays,  $g_s$  would expand (contract) the unstable (stable) subspaces (again in a measure-preserving manner). However, in the case of systems with pure point spectra,  $\bar{g} = \lim_{s \rightarrow \infty} g_s$  is a smooth, flow-invariant metric tensor.

**Theorem 16.** *In systems with pure point spectra and intrinsic dimension  $m$ ,*

$$\bar{g} = \sum_{i=1}^m B_{ij} \beta_i \otimes \beta_j,$$

where  $\beta_i$  are smooth dual vector fields to the vector fields  $v_i$  from Theorem 5, defined uniquely through the relations  $\beta_i(v_j) = \delta_{ij}$ , and  $B_{ij} = \int_M g(v_i, v_j) d\mu$  are the Hodge inner products of the  $v_i$  on  $L^2(TM, g, \mu)$ . Moreover,  $\bar{g}$  has the following properties:

(i) *It is invariant under each of the flows  $\Phi_{i,t}$  generated by  $v_i$ ; i.e.,  $\Phi_{i,t*} \bar{g} = \bar{g}$ , where  $\Phi_{i,t*}$  is the pullback map on  $(0, 2)$  tensors associated with  $\Phi_{i,t}$ .*

(ii) *It is flat.*

(iii) *Its associated volume form has uniform density relative to the invariant measure; i.e.,  $d\text{vol}_{\bar{g}} / d\mu = \Gamma$ , where  $\Gamma$  is a positive constant.*

A proof of Theorem 16 can be found in Appendix B.

**Corollary 17.** The inverse metric associated with  $\bar{g}$  is given by  $\bar{g}^{-1} = \sum_{i=1}^m B_{ij}^{-1} v_i \otimes v_j$ , where  $B_{ij}^{-1}$  are the elements of the  $m \times m$  inverse Gram matrix  $[B_{ij}]^{-1}$ .

According to Theorem 16, the  $v_i$  are Killing vector fields of  $\bar{g}$ , which in turn implies that  $v_i$  (and hence  $v$ ) commute with the Laplace-Beltrami operator  $\Delta_{\bar{g}}$  associated with  $\bar{g}$  (see Section 4.4); a result which can be also verified by explicit calculation of  $\Delta_{\bar{g}}$ .

**Lemma 18.** *In systems with pure point spectra, and in the limit of infinitely many delays, the Laplace-Beltrami operator  $\Delta_{\bar{g}}$  becomes*

$$\Delta_{\bar{g}} = - \sum_{i,j=1}^m B_{ij}^{-1} v_i \circ v_j.$$

*Proof.* It follows from Corollary 17 that the gradient of a function  $f \in L^2(M, \mu)$  with respect to  $\bar{g}$  is given by  $\text{grad}_{\bar{g}} f = \bar{g}^{-1}(\cdot, df) = \sum_{i=1}^m B_{ij}^{-1} v_i(f) v_j$ . Moreover, according to Theorem 5, the  $v_i$  have vanishing  $\mu$ -divergence, and since  $\text{vol}_{\bar{g}} = \Gamma \mu$  by Theorem 16(iii), the  $v_i$  have vanishing divergence with respect to  $\text{vol}_{\bar{g}}$  too (in fact,  $\text{div}_{\mu} = \text{div}_{\bar{g}}$  as can be seen from the expression for the divergence in local coordinates in (15)). Using the Leibniz rule for the divergence,  $\text{div}_{\bar{g}}(uf) = f \text{div}_{\bar{g}} u + u(f)$  where  $u$  is an arbitrary vector field, we obtain

$$\Delta_{\bar{g}} f = - \text{div}_{\bar{g}} \text{grad}_{\bar{g}} f = - \sum_{i,j=1}^m B_{ij}^{-1} \text{div}_{\bar{g}}(v_i(f) v_j) = - \sum_{i,j=1}^m B_{ij}^{-1} (v_i \circ v_j)(f). \quad \square$$

The expression for  $\Delta_{\bar{g}}$  in Lemma 18 in conjunction with the fact that  $[v_i, v_j] = 0$  manifestly shows that  $v$  and  $\Delta_{\bar{g}}$  are commuting operators. Thus, a Koopman eigenfunction  $z_k$  with corresponding eigenvalue  $\lambda_k = i \sum_{i=1}^m k_i \Omega_i$  is also an eigenfunction of  $\Delta_{\bar{g}}$ . In particular,  $\{z_k, z_k^*\}$ , or, equivalently  $\{\text{Re } z_k, \text{Im } z_k\}$ , are orthogonal eigenfunctions of  $\Delta_{\bar{g}}$  at the corresponding eigenvalue  $\sum_{i=1}^m B_{ij}^{-1} \Omega_i \Omega_j k_i k_j$ . As stated in Section 4.4, the leading generating eigenfunction  $\zeta_1$  selected on the basis of the Dirichlet energy  $E_{\bar{g}}(\zeta_1)$  lies entirely in the eigenspace of  $\Delta_{\bar{g}}$  corresponding to its smallest nonzero eigenvalue,  $B_{11}^{-1} \Omega_1^2$ . Moreover, the Galerkin approximation space associated with the eigenfunctions of  $\Delta_{\bar{g}}$  is efficient for approximating the eigenfunctions of  $v$  since each  $z_k$  is expressible as a finite linear combination of eigenfunctions of  $\Delta_{\bar{g}}$ .

To approximate eigenfunctions of  $\Delta_{\bar{g}}$  using a finite number of delays we apply diffusion maps with a variable-bandwidth kernel on  $\mathbb{R}^{sd} \times \mathbb{R}^{sd}$  analogous to (20), viz.

$$K_{\epsilon}(X_i, X_j) = \exp \left( - \frac{\|X_i - X_j\|^2}{\epsilon \hat{\sigma}_{s,\epsilon}^{-1/m}(X_i) \hat{\sigma}_{s,\epsilon}^{-1/m}(X_j)} \right), \quad (37)$$

where  $\hat{\sigma}_{s,\epsilon}$  are estimates of the sampling density  $\sigma_s = d\mu / \text{dvol}_{g_s}$ , accurate at  $O(\epsilon)$ . Note that as  $s \rightarrow \infty$ ,  $\sigma_s$  tends to a constant  $\bar{\sigma} = 1/\Gamma$  in accordance with Theorem 16(iii), but at finite  $s$  it will exhibit fluctuations, and the variable-bandwidth kernel ensures that the numerical eigenfunctions are orthogonal with respect to the invariant measure despite fluctuations in the sampling density. In particular, at finite  $s$ , diffusion maps approximates the Laplace-Beltrami operator  $\Delta_{h_s}$  associated with the conformally transformed metric  $h_s = g_s \sigma_s^{2/m}$ , which has analogous invariance properties and stable behavior as the metric  $h$  introduced in Section 4.1.

The procedure for tuning the kernel bandwidth and computing  $\hat{\sigma}_{s,\epsilon}$  and the Laplace-Beltrami eigenfunctions follows Algorithm 1 with the kernel in (20) replaced by (37). The computational cost of pairwise-kernel evaluations associated with delay-coordinate maps scales linearly with  $s$ . In applications, we sometimes found that the automatic tuning procedure somewhat underestimates appropriate values for  $\epsilon$  at large  $s$ , and a modest bandwidth inflation was necessary to produce stable solutions to the diffusion maps eigenvalue problem. However, the results of the Galerkin method in Algorithm 2 were not too sensitive after  $\epsilon$  exceeded a threshold. We have checked that the numerical results presented in this Section and in Section 5.2 are robust for different values of bandwidth inflation in the range  $4\times$  to  $10\times$ . In practice, there are obvious limitations on how large  $s$  can be since delay embedding reduces the  $N$  samples originally available for analysis to  $N - s$  (the first  $s$  samples are “used up” to create  $\tilde{X}_1$ ), and  $N - s$  must be significantly larger than  $N$  to avoid the risk of spurious correlations in the data.

To illustrate the behavior of the diffusion maps basis obtained via this approach, in Fig. 8 we compare the moduli  $|\langle \phi_k, \zeta_i \rangle|$  of the expansion coefficients of the Koopman eigenfunctions computed via Algorithm 2 in the diffusion maps basis  $\{\phi_k\}$  for  $s = 1$  (i.e., the case studied in Section 4.6) and  $s = 800$ . In the latter case, the ambient space dimension is  $sd = 2400$ , but the dataset lies in a set  $F_s(M)$  of intrinsic dimension  $m = 2$  which has the manifold structure of a 2-torus equipped (according to Theorem 16) with an approximately

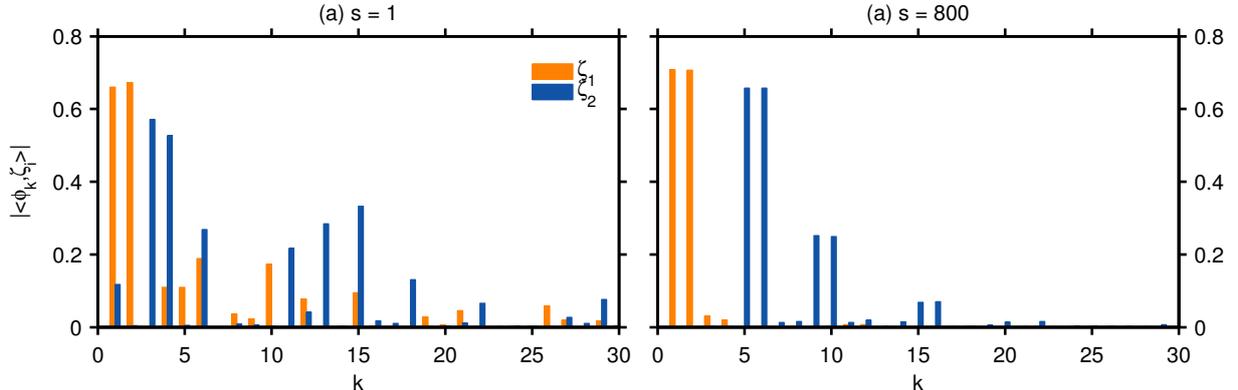


Figure 8: Moduli  $|\langle \phi_k, \zeta_i \rangle|$  of the expansion coefficients of the generating Koopman eigenfunctions  $\zeta_1$  and  $\zeta_2$  in the Laplace-Beltrami eigenfunction basis  $\{\phi_k\}$  from diffusion maps for the variable-speed flow on the 2-torus. (a) No delay coordinate maps,  $s = 1$ ; (b) delay-coordinate maps with  $s = 800$  delays.

flat metric. To compute Koopman eigenvalues and eigenfunctions, we used the same dataset (prior to delay-coordinate mapping) as in Section 4.6, and executed Algorithm 2 using the same parameters as in that Section. While the spectra of the  $\zeta_i$  in Fig. 8 are concentrated in the leading ( $\lesssim 30$ ) diffusion eigenfunctions in both cases, it is evident that in the  $s = 800$  case the spectra are significantly sparser, especially for  $\zeta_1$  which projects almost entirely onto the diffusion eigenfunctions in the leading two-dimensional eigenspace with corresponding eigenvalue  $\eta_1$ , as expected theoretically. Eigenfunction  $\zeta_2$  also exhibits a tight spectral expansion for  $s = 800$  but includes significant contributions from three pairs of diffusion eigenfunctions.

**Remark 19 (Diffusion maps and timescale separation).** Delay-coordinate maps have previously been employed in conjunction with diffusion maps for extraction of spatiotemporal patterns in complex systems ([53, 54], and citing references). In these works, a behavior that has typically been observed is that as the number of delays increases, the time series formed by the diffusion eigenfunctions  $\phi_k$  become increasingly monochromatic, and are able to isolate distinct frequencies from broadband input signals. In [54], it was argued that this behavior can be explained under the assumption that (for a suitable normalization) the diffusion operator approximated by diffusion maps approximates the long-time relaxation to equilibrium of the true (deterministic) system. Under that assumption, the time series of the leading  $\phi_k$  should evolve independently at the characteristic timescales determined by the corresponding diffusion eigenvalues, and the fact that the diffusion eigenvalues have finite spacings leads to timescale separation. Here, we have shown that in the case of systems with pure point spectra the timescale separation in diffusion maps can be explained from the connection between the generator of the Koopman group and the Laplace-Beltrami operator established in Lemma 18. That is, in pure point systems and in the limit of infinitely many delays the diffusion eigenfunctions converge to Koopman eigenfunctions, which provide “maximal” timescale separation due to the fact that they evolve at a single frequency. Of course, the analysis presented here strictly applies only in the restrictive case of pure point spectrum systems. Nevertheless, our results should provide a useful reference point to study connections between diffusion maps and Koopman operators in more complex systems.

As stated below (36), in systems with nonzero Lyapunov exponents the induced metric  $g_s$  becomes ill-behaved as  $s \rightarrow \infty$  since it exponentially expands the stable Oseledets subspaces while shrinking the stable subspaces (in a manner that preserves the invariant measure). The consequence of this behavior in the context of diffusion maps is that in the limit  $s \rightarrow \infty$  and  $\epsilon \rightarrow 0$  the operator  $(I - \mathcal{P}_\epsilon)/\epsilon$  (see Section 4.2) fails to converge to a Laplace-Beltrami operator associated with a smooth metric. Operationally, this means that as  $s$  increases the eigenfunctions from diffusion maps become increasingly biased towards the subspace of  $L^2(M, \mu)$  consisting of functions with vanishing directional derivatives along all but the most

stable subspace (in other words, eigenfunctions with appreciable directional derivatives along the unstable subspaces become increasingly rough in the kernel induced geometry). This effect was first identified in [54], who used a weighted version of the delay coordinate map in (35) to regularize the induced metric and establish connections with Lyapunov metrics of dynamical systems [84]. While biasing the eigenfunctions towards stable subspaces may actually be desirable in certain cases (e.g., by providing an effective means for intrinsic dimension reduction in systems with many positive Lyapunov exponents), in other cases, delay-coordinate maps with many delays may hinder the performance of approximation methods for Koopman eigenfunctions (e.g., if approximate eigenfunctions associated with unstable directions are desired). For these reasons, the availability of schemes such as those presented in Section 4 and Section 6 ahead which are able to accurately approximate Koopman eigenfunctions without performing delay-coordinate maps is important.

### 5.2. Koopman eigenfunctions from noisy data

In this section, we demonstrate that besides being useful for improving the efficiency of spectral Galerkin schemes for the Koopman eigenvalue problem, delay-coordinate maps are also effective when dealing with data generated by dynamical systems with pure point spectra which are corrupted by i.i.d. observational noise. Specifically, we consider that instead of the noise-free time series  $\{x_i\}$  we observe a noisy time series  $\{\tilde{x}_i\}$ , where  $\tilde{x}_i = x_i + \xi_i$ , and the  $\xi_i$  are i.i.d. random variables in  $\mathbb{R}^d$  such that  $\mathbb{E}(\xi_i) = 0$ ,  $\mathbb{E}(\|\xi_i\|^2) = R^2 < \infty$ , and the third and fourth moments of  $\xi_i$  are finite (note that the  $\xi_i$  should not to be confused with the time change function  $\xi$  introduced in Section 6). Here, the assumption that the  $\xi_i$  have vanishing expectation leads to no loss of generality since a nonzero constant expectation can be absorbed by a shift of the data, which leaves the pairwise distances in the diffusion maps kernel unchanged. Besides the above requirements, we do not make specific assumptions about the distribution of the  $\xi_i$ , but in the numerical experiments below we use Gaussian noise.

Following the approach presented in Section 5.1, we fix an integer parameter  $s$  and replace the observed time series  $\{\tilde{x}_i\}_{i=0}^{N-1}$  by the time series  $\{\tilde{X}_i\}_{i=s-1}^{N-1}$ , where  $\tilde{X}_i = (\tilde{x}_i, \tilde{x}_{i-1}, \dots, \tilde{x}_{i-s+1}) \in \mathbb{R}^{sd}$ . Note that we can also write  $\tilde{X}_i = X_i + \Xi_i$ , where  $X_i = F_s(a_i) = (x_i, x_{i-1}, \dots, x_{i-s+1})$  are the noise-free samples in delay-coordinate space, and  $\Xi_i = (\xi_i, \xi_{i-1}, \dots, \xi_{i-s+1})$ . To compute diffusion eigenfunctions from noisy data, we replace the kernel in (37) by

$$K_\epsilon(\tilde{X}_i, \tilde{X}_j) = \exp\left(-\frac{\|\tilde{X}_i - \tilde{X}_j\|^2}{\epsilon \tau_{s,\epsilon}^{-1/m}(\tilde{X}_i) \tau_{s,\epsilon}^{-1/m}(\tilde{X}_j)}\right), \quad (38)$$

where the functions  $\tau_{s,\epsilon}(\tilde{X}_i)$  are modified bandwidth functions which estimate the sampling density up to a proportionality constant in the presence of noise. Details on these functions and the asymptotic properties of diffusion maps with the kernel in (38) are included in Appendix C.

Kernels in delay-embedding space have previously been used in conjunction with diffusion maps for correction of timing uncertainties in time-ordered data [85]. Here, we demonstrate that such kernels are also useful for denoising data corrupted by i.i.d. noise. As discussed in Appendix C, the effect of i.i.d. noise is to introduce a random bias with positive expectation to the pairwise squared distances  $\|\tilde{X}_i - \tilde{X}_j\|^2$ . In particular, by the law of large numbers, as the number of delays  $s$  increases that bias converges to a constant number equal to  $2R^2$ . This bias affects the kernel by introducing (1) a multiplicative bias in the density estimates  $\hat{\sigma}_{s,\epsilon}$  used in (37), and (2) a positive offset in  $\|\tilde{X}_i - \tilde{X}_j\|^2$  relative to the true value  $\|X_i - X_j\|^2$ . However, both of these effects can be removed through appropriate normalizations; in particular, the influence on the bandwidth functions can be removed by a rescaling of  $\hat{\sigma}_{s,\epsilon}$  to construct  $R$ -independent functions  $\tau_{s,\epsilon}$ , and the influence of the distance offset is canceled out in the diffusion maps normalization. As a result, the diffusion maps matrix  $\tilde{P}$  constructed from the noisy data provides a consistent approximation of the integral operator  $\mathcal{P}_\epsilon$  (see Section 4.2) up to an  $O(\epsilon^2)$  correction which does not affect the consistency of the Laplace-Beltrami operator approximated through  $(I - \tilde{P})/\epsilon$ . We can therefore employ our Galerkin scheme for the Koopman eigenvalue problem in Algorithm 2 using clean basis functions determined through the eigenfunctions of  $\tilde{P}$ . Theoretically, this approach allows for the removal of i.i.d. noise of arbitrarily large

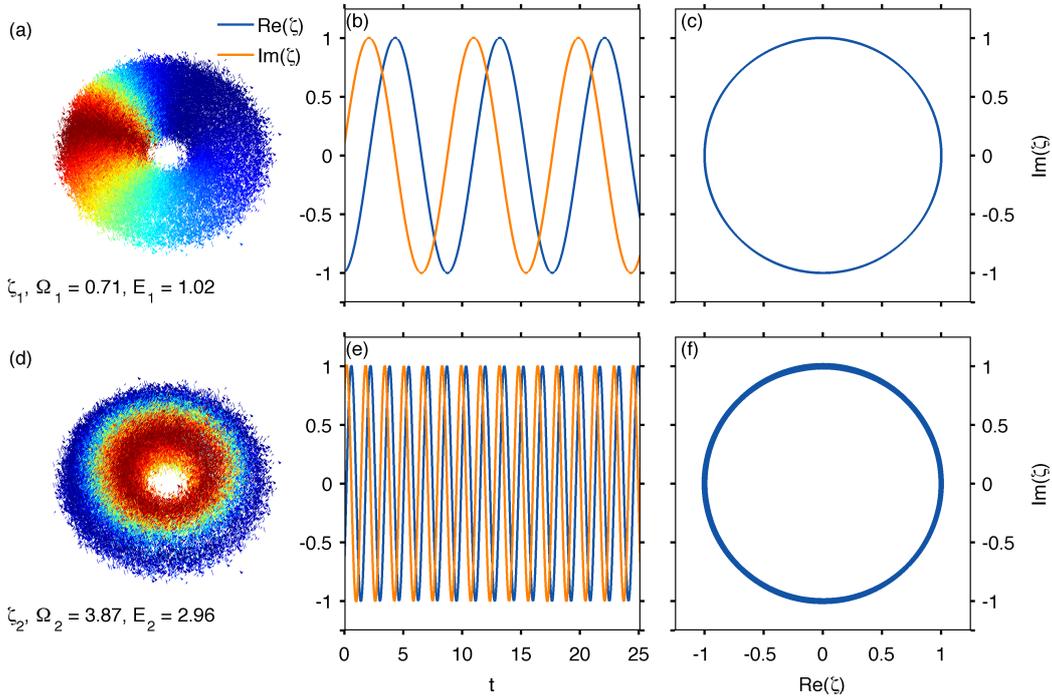


Figure 9: The generating Koopman eigenfunctions  $\zeta_1$  (a–c) and  $\zeta_2$  (d–f) for the variable-speed dynamical system on the 2-torus computed using data corrupted by i.i.d. Gaussian noise with standard deviation 0.1 in each component of the observation vector in  $\mathbb{R}^3$ . (a, d) Scatterplots of  $\text{Re}(\zeta_i)$  on the noisy torus; (b, e) time series of  $\text{Re} \zeta_i$  and  $\text{Im} \zeta_i$ ; (c, f) scatterplots of  $(\text{Re} \zeta_i, \text{Im} \zeta_i)$ . The eigenfunctions shown here have not been rescaled to the unit circle—they are the result of Step 5 of Algorithm 2 executed with the modified kernel in (38) for  $s = 100$  lags. The number of samples in this dataset is 64,000.

variance provided that arbitrarily many delays  $s$  can be used. In practice, the amount of variance that can be feasibly handled is limited by the total number of samples  $N$  (as we must have  $s \ll N$ ). Furthermore, if the system has nonzero Lyapunov exponents (a case that we are not treating here) adding delays leads to the same eigenfunction bias issues as described in Section 5.1.

Figures 9 and 10 show the generating Koopman eigenfunctions computed via the approach described above and in Appendix C for the same variable-speed torus system as in Section 4.6 but with Gaussian noise added to the observed data in  $\mathbb{R}^3$ . We consider two cases; one with moderate noise of standard deviation 0.1 added to each component of the observation vector (Fig. 9), and one with strong noise of standard deviation 1 (Fig. 9). In the moderate-noise case, we used a dataset of 64,000 samples sampled at the same time interval as in Section 4.6. In the strong-noise case the number of samples was 128,000 though reasonably good results can also be obtained using 64,000 samples. For ease of comparison, we used  $s = 800$  delays in both cases, though in the moderate-noise case as few as 20 delays are sufficient for denoising. Using the diffusion eigenfunctions from each case, we computed the generating Koopman eigenfunctions using Algorithm 2 with the same parameters as in Section 4.6 except that in the strong-noise case we used  $n = 401$  (as opposed to 1001) diffusion eigenfunctions to build the Galerkin approximation space.

As is evident from Figs. 9 and 10, with the modifications for noisy data described above, Algorithms 1 and 2 successfully recover the generating Koopman eigenfunctions shown for the noise-free data in Fig. 4, in both the moderate- and strong-noise cases. In both cases, the Koopman frequencies  $\Omega_i$  agree to their counterparts computed from the noise-free data to within 3 significant figures. As expected from its small Dirichlet energy, eigenfunction  $\zeta_1$  is particularly well recovered in both cases. Eigenfunction  $\zeta_2$  exhibits a moderate amount of amplitude modulation in the strong-noise case, but it retains a nearly-monochromatic oscillation at the correct theoretical frequency. Koopman eigenfunction results analogous to Figs. 9 and 10

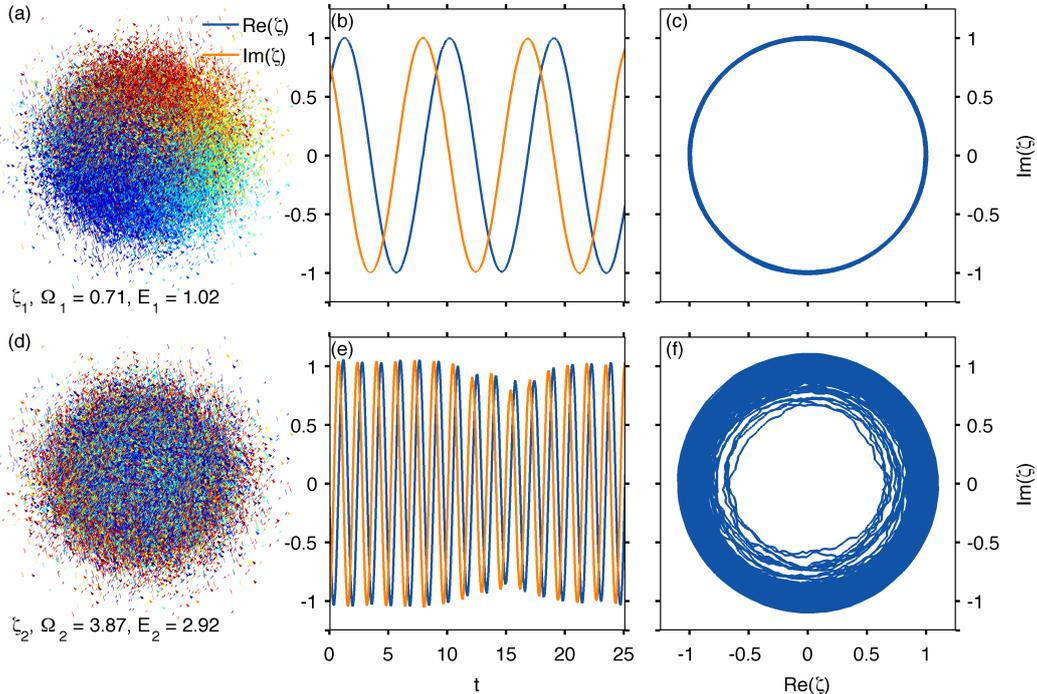


Figure 10: Same as Fig. 9 but for noise standard deviation equal to 1 and number of observed samples equal to 128,000.

obtained via EDMD are shown in Figs. 2 and 3 in the SOM, respectively. There, it can be seen that while our approach and EDMD perform comparably in the moderate-noise case, in the strong-noise case the quality of the numerical Koopman eigenfunctions obtained via our approach is higher.

## 6. Regularization by time change

### 6.1. Time change in dynamical systems

Time change (e.g., [55]) is a technique in continuous-time dynamical systems which involves scaling the generator  $v$  by a smooth positive function  $\psi$  to create the vector field  $w = \psi^{-1}v$ . This new vector field generates a flow  $\Psi_t$  on  $M$  having the same orbits as the original flow  $\Phi_t$  (since  $w$  and  $v$  are parallel), but time flows differently with respect to  $\Psi_t$  than it does with respect to  $\Phi_t$ . In particular, for  $\Psi_\tau a = \Phi_t a$  we have  $\tau = \int_0^t \psi(\Phi_{t'} a) dt'$ . Moreover, if  $(M, \mathcal{B}, \mu, \Phi_t)$  is ergodic, then  $(M, \mathcal{B}, \nu, \Psi_t)$  is also ergodic for the invariant probability measure  $\nu$  with density

$$d\nu = \frac{\psi}{\bar{\psi}} d\mu, \quad \bar{\psi} = \int_M \psi d\mu. \quad (39)$$

While time change preserves ergodicity, the same is not true for mixing—a theorem due to Kočergin [86] states that for any ergodic flow there exists a time change by a smooth function which renders it mixing. Heuristically, we can think of mixing by time change in the following way. Suppose that  $A \subset M$  is a measurable set that we picture as a coherent globular object. In the original, non-mixing system,  $\Phi_t(A)$  maintains its coherence in the course of dynamical evolution, but in the time-changed system parts of  $A$  will lag behind others. For a suitable time-change function,  $\Psi_t(A)$  will become increasingly stretched and twisted, and its volume will become asymptotically equidistributed in  $M$  in accordance with the condition for mixing in Section 2. As a concrete example, which we will study with numerical experiments in Section 6.4.1 ahead,

Fayad [50] shows that an irrational flow on the 3-torus becomes mixing through the analytic time-change function

$$\psi(\theta^2, \theta^2, \theta^3) = 1 + \operatorname{Re} \sum_{k=1}^{\infty} \sum_{|l| \leq k} \frac{e^{-k}}{k} \left( e^{ik\theta^1} + e^{ik\theta^2} \right) e^{il\theta^3}, \quad (40)$$

where  $\theta^\mu \in [0, 2\pi)$  are the three angles on  $\mathbb{T}^3$ . Figure 11 shows  $1/\psi$  as a time series for the underlying irrational flow  $v = \sum_{\mu=1}^3 v^\mu \frac{\partial^\mu}{\partial \theta^\mu}$  with frequencies  $(\alpha_1, \alpha_2, \alpha_3) = (v^1, v^2, v^3) = (1, 5^{1/2}, 10^{1/2})$ .

## 6.2. Dimension reduction and nonparametric forecasting with time change

Traditionally, time change is used as a technique to improve mixing and stochasticity of dynamical systems. Here, our approach is to use time change as a regularization tool to reduce mixing and improve the spectral properties of the Koopman operators for dimension reduction and nonparametric forecasting. In essence, we will use the eigenfunctions of the time-changed system to decompose the original dynamical system into a collection of non-autonomous oscillators with variable frequency (as opposed to simple harmonic oscillators). We will also use these eigenfunctions to perform a vector field decomposition analogous to Theorem 5, but the vector field components in this case will be non-commuting. Finally, we will generalize the forecasting methods of Section 3.3 to the time-changed framework.

In what follows,  $(M, \mathcal{B}, \mu, \Phi_t)$  will be an ergodic dynamical system on an  $m$ -dimensional manifold  $M$  with vector field  $v$  generating the observed data, and  $(M, \mathcal{B}, \nu, \Psi_t)$  will be a time-changed system with vector field  $w = \psi^{-1}v$  and invariant measure  $\nu$  with the density in (39).

**Proposition 20** (Dimension reduction with time change). *Assume that  $(M, \mathcal{B}, \nu, \Psi_t)$  has pure point spectrum with basic frequencies  $\{\Omega_i\}_{i=1}^m$  and corresponding orthonormal eigenfunctions  $\{\zeta_i\}_{i=1}^m$  on  $L^2(M, \nu)$ , and define the projection maps  $\pi_i : M \mapsto \mathbb{C}$  as in (6), i.e.,  $\pi_i(a) = \zeta_i(a)$ . Then, the time series  $\tilde{\zeta}_i(t) = \zeta_i(\Phi_t a)$  evolves as a non-autonomous variable-frequency oscillator with*

$$\frac{d\tilde{\zeta}_i}{dt} = i\Omega_i \tilde{\psi}(t) \tilde{\zeta}_i(t), \quad \tilde{\psi}(t) = \psi(\Phi_t a).$$

*Proof.* This equation follows directly from the facts that  $d\zeta_i(\Psi_\tau a)/d\tau = i\Omega_i \zeta_i(\Psi_\tau a)$  and  $d\tau = \tilde{\psi}(t) dt$  when  $\Phi_t a = \Psi_\tau a$ .  $\square$

Proposition 20 indicates that if the dynamical system generating the data is related to a system with pure point spectrum by a time change, we can use the spectral properties of that system to perform dimension reduction and forecasting at the expense of introducing additional complexity (in the case of Proposition 20, non-autonomous dynamics) due to the time change function. The following Lemma, which we state without proof, is the analog of Theorem 5 for time-changed systems.

**Lemma 21** (Vector field decomposition with time change). *Let  $(M, \mathcal{B}, \nu, \Psi_\tau)$  be as in Proposition 20 and let  $w = \sum_{i=1}^m w_i$  be the vector field decomposition of  $w$  from Theorem 5. Then,  $v$  admits the decomposition  $v = \sum_{i=1}^m v_i$  into the nowhere-vanishing, linearly independent,  $\mu$ -preserving vector fields  $v_i = \psi w_i$ .*

Unlike Theorem 5, in this case the vector fields  $v_i$  can be non-commuting due to the presence of  $\psi$ , which means that in general we cannot view them as generators of independent dynamical processes. Nevertheless, the  $v_i$  can still be reconstructed in data space using a pushforward operation analogous to (8), with the difference that the observation map  $F$  is now decomposed in the orthonormal basis of  $L^2(M, \nu)$  associated with  $w$ . That is, we write  $F = \sum_k \hat{F}_k z_k$ , where  $\hat{F}_k = \langle z_k, F \rangle_\nu$ , and  $z_k = \prod_{i=1}^m \zeta_i^{k_i}$  is the eigenfunction of  $w$  corresponding to eigenvalue  $\lambda_k = i \sum_{i=1}^m k_i \Omega_i$ ,  $k = (k_1, \dots, k_m) \in \mathbb{Z}^m$ . In this basis, the pushforward operation for  $v_i$  becomes

$$V_i := F_* v_i = v_i(F) = \sum_k i k_i \Omega_i \psi \hat{F}_k z_k, \quad V := F_* v = \sum_{i=1}^m V_i. \quad (41)$$

We proceed similarly to modify the forecasting scheme of Section 3.3. As before, we are interested in forecasting the density  $\rho_t = d\mu_t/d\mu$  of a time-dependent measure  $\mu_t$  relative to  $\mu$  given the initial data  $\rho_0$ . The evolution of the density is still governed by the Perron-Frobenius operator  $U_t^* = U_{-t}$  for the dynamical system  $(M, \mathcal{B}, \mu, \Phi_t)$ , but in this case we treat  $\rho_t$  as a function in  $L^2(M, \nu)$  and expand it in the orthonormal basis  $\{z_k\}$ . That is, substituting the expansion  $\rho_t = \sum_i \hat{\rho}_i(t) z_i$  with  $\hat{\rho}_i(t) = \langle z_i, \rho_t \rangle_\nu$  into the equation

$$\frac{d\rho_t}{dt} = \lim_{\delta t \rightarrow 0} \frac{U_{\delta t}^* \rho_t - \rho_t}{\delta t} = -v(\rho_t),$$

we obtain

$$\sum_j \frac{d\hat{\rho}_j}{dt} z_j = - \sum_j \hat{\rho}_j(t) v(z_j) = - \sum_j i\omega_j \hat{\rho}_j(t) \psi z_j. \quad (42)$$

Taking now  $L^2(M, \nu)$  inner products of the last equation with the basis elements  $\{z_i\}$ , it follows that

$$\frac{d\hat{\rho}_i}{dt} = - \sum_j iH_{ij} \omega_j \hat{\rho}_j(t), \quad H_{ij} = \langle z_i, \psi z_j \rangle_\nu = \bar{\psi}^{-1} \langle z_i, \psi^2 z_j \rangle_\mu. \quad (43)$$

Thus, the uncoupled simple harmonic oscillators of Section 3.1 are replaced by coupled oscillators in the time-changed setting with coupling coefficients given by the  $H_{ij}$  coefficients. These coefficients can be viewed as the elements of an operator on sequences in  $\ell^2$ , and writing  $\hat{\rho}(t) = (\dots, \hat{\rho}_{i-1}, \hat{\rho}_{i_0}, \hat{\rho}_{i_1}, \dots)$  we have

$$\frac{d\hat{\rho}}{dt} = -iH\omega\hat{\rho}(t),$$

where  $\omega$  is the diagonal operator on  $\ell^2$  in (10). This last equation has the formal solution (cf. (9))

$$\hat{\rho}(t) = e^{-iH\omega t} \hat{\rho}(0).$$

With this solution, the time-dependent expectation value of an observable  $f = \sum_j \hat{f}_j z_j$  can be computed from an analogous expression to (11), but with additional coupling terms due to time change, viz.

$$\bar{f}_t = \mathbb{E}_{\mu_t} f = \int_M f \rho_t d\mu = \sum_{i,j} \hat{\rho}_i(t) H'_{ij} \hat{f}_j, \quad (44)$$

where the coefficients  $H'_{ij}$  are given by

$$H'_{ij} = \langle z_i, z_j \rangle_\mu = \bar{\psi} \langle z_i, \psi^{-1} z_j \rangle_\nu. \quad (45)$$

Note that as with (43), we can compute  $H'_{ij}$  using the inner products of either of  $L^2(M, \mu)$  or  $L^2(M, \nu)$ .

**Remark 22 (Loss of predictability).** The coupling of the expansion coefficients  $\hat{\rho}_i(t)$  can potentially lead to barriers in predictability. In particular, if  $\psi$  is a polynomial in  $\{\zeta_i\}$  of degree  $N < \infty$  then knowledge of a finite number of expansion coefficients is sufficient to forecast the expectation values of polynomial observables at arbitrarily long times. On the other hand, if the number of coefficients needed to represent  $\psi$  exactly is infinite (e.g., the time-change function in (40)), then prediction of polynomial observables at arbitrarily long times may not be possible with finite numbers of  $\hat{\rho}_i$  coefficients known at forecast initialization. Note that if the system is mixing then in the course of dynamical evolution  $\rho_t$  develops arbitrarily small lengthscales on  $M$  and arbitrarily large bandwidth in the eigenfunction basis. However, the corresponding measures  $\mu_t$  converge to  $\mu$  in the sense that  $\int_M f d\mu_t \rightarrow \int_M f d\mu$  for all  $f \in L^2(M, \mu)$  [41].

### 6.3. Data-driven implementation

The analysis in Section 6.2 shows that time-change transformations can extend the applicability of Koopman eigenfunction techniques to certain systems that do not possess nonconstant eigenfunctions. Of course, the systems in question have special structure, namely that they are related to pure point systems via time change. Even in such cases, taking advantage of this special structure from data is challenging without prior information about the time-change function  $\psi$ . Nevertheless, motivated from the results of Section 6.2, in this section, we put forward a time-change scheme whose objective is to improve the spectral properties of the generator for eigendecomposition. This method employs a time-change function which is empirically accessible from time-ordered data, and recovers  $\psi$  in special cases.

Let  $\xi = \|v\|_g = \sqrt{g(v, v)}$  be the norm of the dynamical vector field with respect to the ambient space Riemannian metric  $g$ . This quantity is nonnegative and measurable, and thus can be used as a time change function to obtain the vector field  $\hat{w} = \xi^{-1}v$ . Note that  $\xi$  vanishes only at fixed points of  $v$ , which have measure zero by ergodicity. In the special case that  $v$  is indeed related to a system with pure point spectrum by a time change with function  $\psi$ , then it follows from Lipschitz equivalence of Riemannian metrics on compact manifolds that there exists a constant  $0 < C < \infty$  such that  $\xi/C \leq \psi \leq C\xi$ . Moreover, there exists a metric  $g$  (e.g., the flat metric on the torus) such that  $\xi = \psi$ .

A key property of  $\xi$  is that it can be approximated from time-ordered data using finite differences. In particular, we have  $\xi^2 = V \cdot V$ , where  $V = F_*v$  is the pushforward of  $v$  in data space, and we can approximate  $V$  via finite differences as described in Section 2. Our approach is to use  $\xi$  as an empirically accessible time-change function, and consider  $\hat{w}$  as a candidate vector field with well-behaved eigenfunctions on the Hilbert space  $L^2(M, \hat{\nu})$  with  $d\hat{\nu} = \xi d\mu/\bar{\xi}$ ,  $\bar{\xi} = \int_M \xi d\mu$ . In what follows, we will carry out the vector field decomposition in Theorem 5 and the statistical forecasts in (42) and (44) using  $\xi$  as the time-change function in place of  $\psi$ .

Proceeding in direct analogy with the methods of Section 4.3, we compute approximate eigenfunctions of  $\hat{w}$  by approximating this operator in a basis of eigenfunctions of a Laplace-Beltrami operator for a suitable Riemannian metric  $\hat{h}$ . In this case, the metric should have volume form equivalent to  $d\hat{\nu}$ , so we consider the conformal transformation  $\hat{h} = (\xi\sigma)^{2/m}g$ , where, as before,  $\sigma$  is the density of the invariant measure  $\mu$  relative to the Riemannian measure  $\text{vol}_g$  of the ambient space metric. We approximate the Laplace-Beltrami operator for this geometry using diffusion maps with the modified kernel (cf. (20))

$$\hat{K}_\epsilon(x, y) = \exp\left(-\frac{\|x - y\|^2}{\epsilon(\xi(x)\hat{\sigma}_\epsilon(x))^{-1/m}(\xi(y)\hat{\sigma}_\epsilon(y))^{-1/m}(y)}\right) \quad (46)$$

and the same normalization as Algorithm 1. Note that in this case the sampling density  $q$  relative to the Riemannian measure of  $\hat{h}$  is nonuniform, but the effects of  $q$  are canceled out as  $\epsilon \rightarrow 0$  via normalization as described in Section 4.2. Applying Algorithm 1 with the kernel in (46), we obtain an orthonormal basis of  $L^2(M, \hat{\nu})$  from the Laplace-Beltrami eigenfunctions  $\{\hat{\phi}_i\}$  associated with  $\hat{h}$ . We then construct the associated basis  $\{\hat{\varphi}_i\}$  of  $H_1(M, \hat{\nu})$  following (18), and compute approximate eigenvalues and eigenfunctions of  $\hat{w}$  from the eigenvalue problem for the advection-diffusion operator

$$\hat{L}_\epsilon = \hat{w} + \epsilon\Delta_{\hat{h}} \quad (47)$$

with an analogous algorithm to Algorithm 2.

Our vector field decomposition and statistical forecasting techniques with time change can then be carried out using Algorithms 3 and 4 with appropriate modifications to take into account the change of basis and the coupling coefficients in (42) and (45). Instead of giving full listings for these modified algorithms, we indicate below the required changes.

- The eigenfunctions  $\{u_i\}$  and eigenvalues  $\{\gamma_i\}$  are for the time-changed operator  $\hat{L}_\epsilon$ . Similarly, the discrete inner product weights  $w_i$  are for the diffusion eigenfunction basis  $\{\hat{\phi}_i\}$ .
- In Step 3 of Algorithm 3, the vector components  $V_i$  are now given by  $V_i = i\hat{F}G^{-1}\bar{\Omega}_i z^\top \text{diag } \xi$ , where  $\text{diag } \xi$  is an  $N \times N$  diagonal matrix whose  $i$ -th diagonal entry is set to the value of the time-change function at the  $i$ -th data sample,  $\xi_i$ .

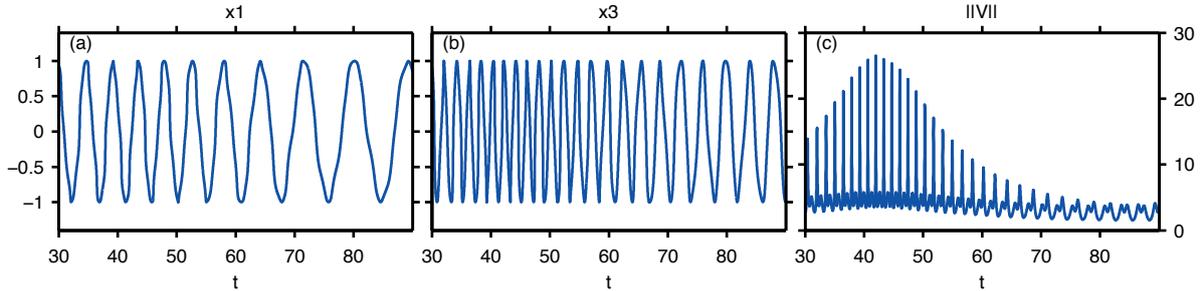


Figure 11: Time series from the mixing system on the 3-torus. (a, b) State vector components  $x^1$  and  $x^3$ , respectively, in the flat embedding of the 3-torus in  $\mathbb{R}^6$ ; (c) magnitude  $\|V\| = \|v\|_g$  of the dynamical vector field with respect to the ambient-space metric. In this case,  $\|V\|$  is inversely proportional to the time-change function  $\psi$  producing mixing.

- In Algorithm 4, we also compute the  $n \times n$  matrices  $H = z^\dagger \text{diag } w \text{diag } \xi z$  and  $H' = z^\dagger z/N$  approximating the coupling coefficients in (43) and (45), respectively. Note that we avoid division by  $\xi_i$  in the expression for  $H'$  using ergodic sums to approximate the  $L^2(M, \mu)$  inner product in (45).
- In Step 3(a) of Algorithm 4, the expansion coefficients for the density are now given by  $\tilde{\rho}(t_i) = \exp(-iG^{-1}H \text{diag } \omega t_i) \tilde{\rho}_0$ . Similarly, in Step 3(c), the expectation value of the observable now becomes  $\bar{f}(t_i) = \hat{f}(G^{-1}H' \tilde{\rho}(t_i))^\dagger$ .

## 6.4. Applications

### 6.4.1. Mixing flow on the 3-torus

As a first application of our dimension reduction and nonparametric forecasting techniques with time change, we study the mixing system on  $M = \mathbb{T}^3$  in [50] with the time-change function  $\psi$  in (40), using the underlying irrational flow with angular frequencies  $(\alpha_1, \alpha_2, \alpha_3) = (1, 5^{1/2}, 10^{1/2})$ . For this set of experiments, we generated a time series consisting of  $N = 512,000$  samples in  $\mathbb{R}^6$  taken at a timestep  $T = 0.01$  with the canonical embedding  $F(a) = (x^1, \dots, x^6) = (\cos \theta^1, \sin \theta^1, \cos \theta^2, \sin \theta^2, \cos \theta^3, \sin \theta^3)$ , where  $(\theta^1, \theta^2, \theta^3)$  are the angles corresponding to the point  $a \in M$ . With this embedding, the torus inherits the flat metric  $g$ , and we have  $\|v\|_g = \psi^{-1}(\alpha_1^2 + \alpha_2^2 + \alpha_3^2)^{1/2}$ . Therefore, in this case,  $\|v\|_g = \xi$  is inversely proportional to the true time-change function  $\psi$ , and the method described in Section 6.3 is guaranteed to recover an irrational flow on  $\mathbb{T}^3$ . This application is therefore a “best-case scenario” for the techniques of Section 6.3 to perform well, but nevertheless it remains challenging for data-driven eigendecomposition and forecasting techniques due to mixing dynamics.

Figure 11 shows representative time series from this system together with the vector field norm  $\|v\|_g = \xi \propto \psi^{-1}$ . Qualitatively,  $\xi$  exhibits a series of intermittent spikes whose amplitude is modulated by a low-frequency envelope. These variations in the speed of the flow, whose maxima and minima differ by more than a factor of 10, produce phase-modulated wavetrains in the observed time series (see Fig. 11(a, b)). As a result, the invariant measure  $\mu$  has highly nonuniform density relative to the Haar measure on the 3-torus.

For comparison, we begin by solving the eigenvalue problem for the generator using the method in Section 4.5 which does not involve time change. Because the system is mixing and  $v$  has no eigenfunctions, the spectrum of the regularized operator  $L_\epsilon$  is expected from theory [81, 82] to deviate significantly from the simple harmonic oscillator patterns arising in systems with pure point spectra, and as shown in Fig. 11 this is indeed the case. The eigenfunctions in Fig. 11, which were computed for the regularization parameter  $\epsilon = 3 \times 10^{-4}$ , include one with a quasiperiodic time series (Fig. 12(a)) that appears to capture the low-frequency modulating envelope of the time change function (Fig. 11(c)). However, other eigenfunctions (Fig. 12(b, c)) exhibit strong amplitude modulation—a manifestation of the fact that advection and diffusion interact non-trivially in  $L_\epsilon$  in the presence of mixing and even for small  $\epsilon$ . Note that applying no regularization ( $\epsilon = 0$ ) results in significantly less coherent patterns than those in Fig. 12.

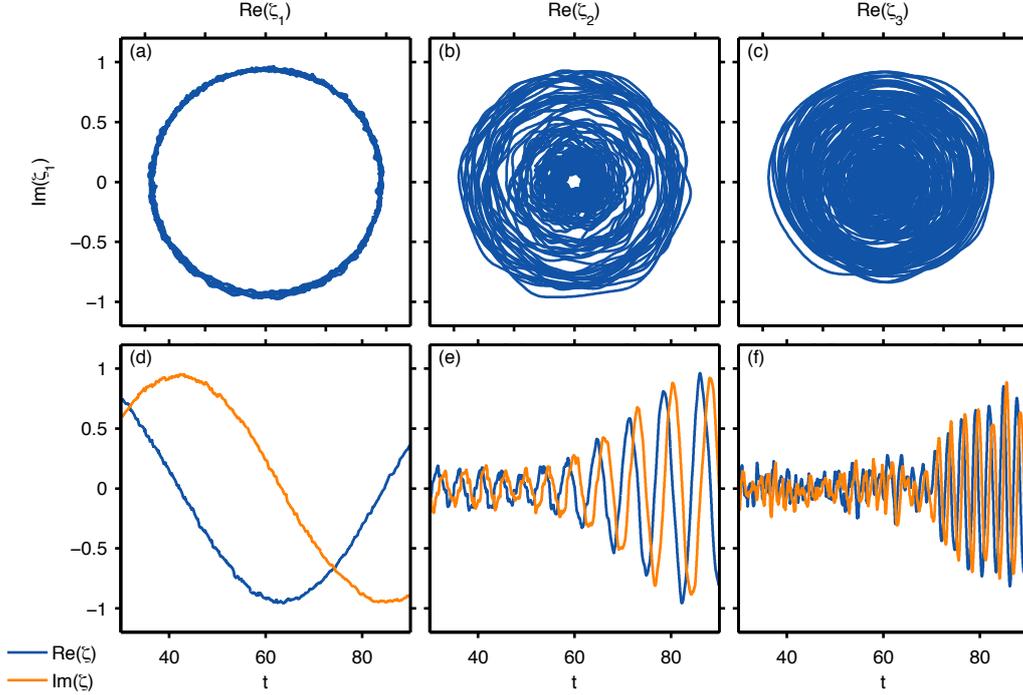


Figure 12: Generating eigenfunctions for the mixing system on the 3-torus without time change. (a) Quasiperiodic pattern associated with the low-frequency envelope in Fig. 11(c); (b, c) amplitude-modulated patterns.

Next, we discuss the results obtained via the time-change technique of Section 6.3. As shown in Fig. 13, the generating eigenfunctions  $\{\zeta_1, \zeta_2, \zeta_3\}$  from this method lie on the unit circle to a good approximation (though with somewhat less fidelity than the eigenfunctions for the 2-torus system in Fig. 4), and, as expected from Proposition 20, the corresponding time series have the structure of phase-modulated wavetrains. The basic frequencies associated with the generating eigenfunctions are  $\{\Omega_1, \Omega_2, \Omega_3\} = \{0.5587, 0.2498, 0.7902\}$ ; these values agree to within three significant figures with the frequencies  $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3) = (1, 5^{1/2}, 10^{1/2})/4 \approx (0.2500, 0.5590, 0.7906)$ , of the irrational flow  $\hat{w} = \|v\|_g^{-1}v = \sum_{\mu=1}^3 \hat{\alpha}_\mu \frac{\partial}{\partial \theta^\mu}$ . Based on this identification, we expect the eigenfunctions corresponding to the  $\Omega_i$  to have the structure  $\zeta_i = e^{i\theta_i}$  and to have equal Dirichlet energies  $E_{\zeta_i}$  in the flat metric. Indeed, the numerical Dirichlet energies for the eigenfunctions in Fig. 6.3 are  $E_{\zeta_1} = 1.264$ ,  $E_{\zeta_2} = 1.266$ , and  $E_{\zeta_3} = 1.28$ . The real parts of the numerical eigenvalues are  $\text{Re } \gamma_i \approx 5 \times 10^{-4}$  in all three cases, indicating that the effects of diffusion in the regularized operator  $\hat{L}_\epsilon$  are minimal after time change.

Figure 14 displays the vector field decomposition from Lemma 21 for this system using a moderately small spectral order parameter  $l = 5$  (see Algorithm 3), i.e.,  $n = (2l+1)^3 = 1331$  eigenfunctions in total. The vector field components are reconstructed in the six-dimensional data space, and then projected to three-dimensional periodic boxes for visualization. At this spectral order of approximation, the reconstruction error in  $\sum_{i=1}^3 V_i$  is 1% in  $\mathbb{R}^6$ , but note that the visualizations in Fig. 14 are subject to additional projection errors, and therefore appear noisier than the native vector fields in  $\mathbb{R}^6$ . Nevertheless, the plots in Fig. 14(b–d) illustrate clearly that  $V_1, V_2,$  and  $V_3$  generate flows along the  $\theta^2, \theta^1,$  and  $\theta^3$  directions, respectively, which is consistent with the identification with the  $\hat{w}$  system made earlier.

Next, we turn to nonparametric forecasting of this system. We set the initial probability measure  $\mu_0$  to a measure with isotropic circular Gaussian density  $\sigma$  relative to the Haar measure on  $\mathbb{T}^3$ , viz.  $\sigma(\theta^1, \theta^2, \theta^3) = \exp(\kappa \sum_{i=1}^3 \cos(\theta^i - \hat{\theta}^i)) / (I_0(\kappa))^3$  which is the three-dimensional analog of (14). Taking into account the density of the invariant measure relative to the Haar measure, our initial probability measure has density  $\rho_0 = d\mu_0/d\mu = \sigma\hat{\psi}/\psi$  relative to  $\mu$ , where  $\psi$  and  $\hat{\psi}$  are the time change function and its ergodic mean

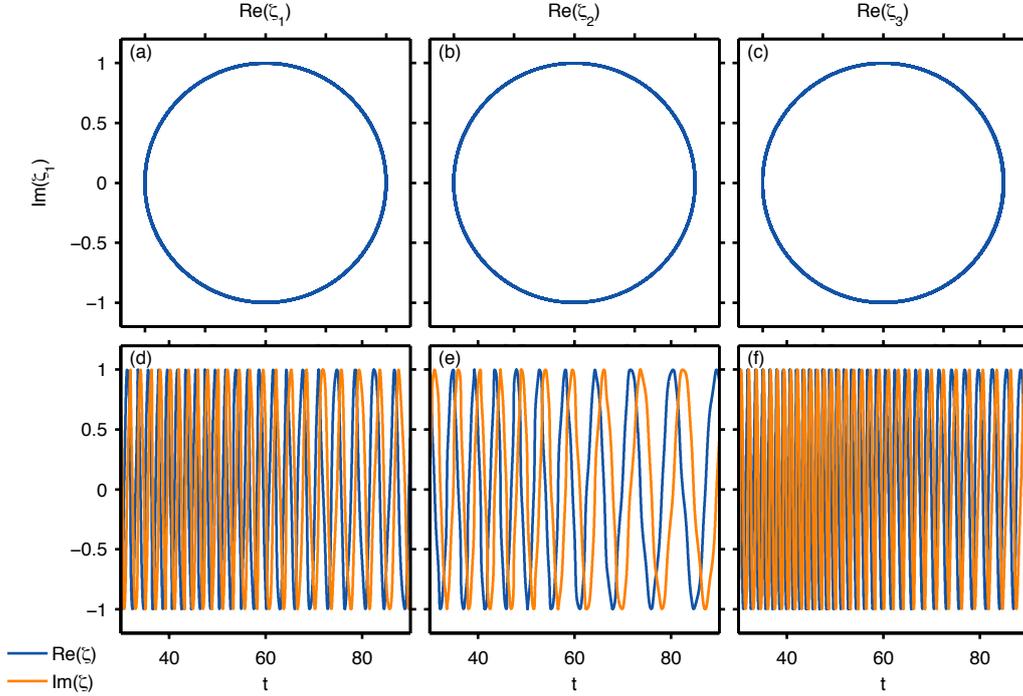


Figure 13: Generating eigenfunctions  $\zeta_i$  for the mixing system on the 3-torus with time change. (a–c) Mapping of the 3-torus to the circles ( $\text{Re}\zeta_i, \text{Im}\zeta_i$ ) on the complex plane; (d–f) time series of the real and imaginary parts of the eigenfunctions exhibiting time-dependent frequency (cf. the eigenfunctions in Fig. 4 for the 2-torus system with pure point spectrum).

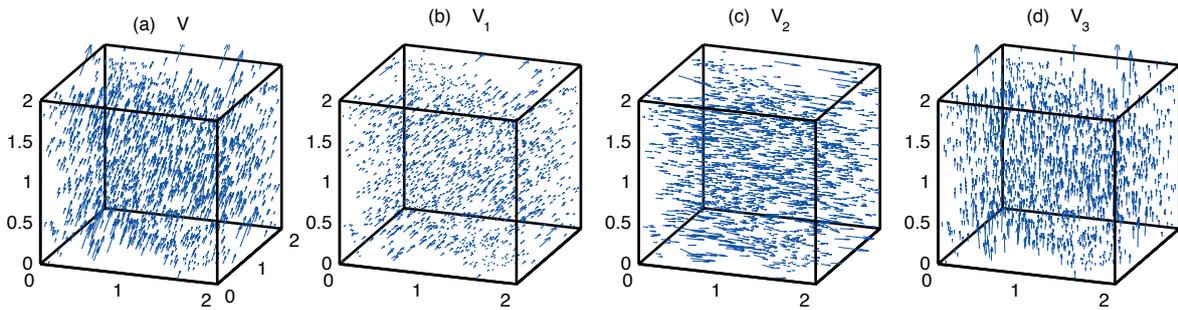


Figure 14: Vector field decomposition for the mixing dynamical system on the 3-torus. (a) Full vector field; (b–d) vector field components from Lemma 21. All vector fields are visualized in a periodic box with the  $x$ ,  $y$ , and  $z$  axes corresponding to the normalized angles  $\theta^1/\pi$ ,  $\theta^2/\pi$ , and  $\theta^3/\pi$ , respectively, on  $\mathbb{T}^3$ . To create these plots, the vector fields were first reconstructed in data space,  $\mathbb{R}^6$ , via (41) and then projected to the periodic box via the inverse derivative map  $\partial\theta^\mu/\partial x^\nu$ . This projection step is needed because (41) cannot be applied to reconstruct the vector fields directly in the periodic box (in that case, the map  $F : \mathbb{T}^3 \mapsto \mathbb{R}^3$  would be discontinuous). Because  $\partial\theta^\mu/\partial x^\nu$  has singularities, the reconstructed arrow plots are more noisy than the natively reconstructed vector fields in  $\mathbb{R}^6$ .

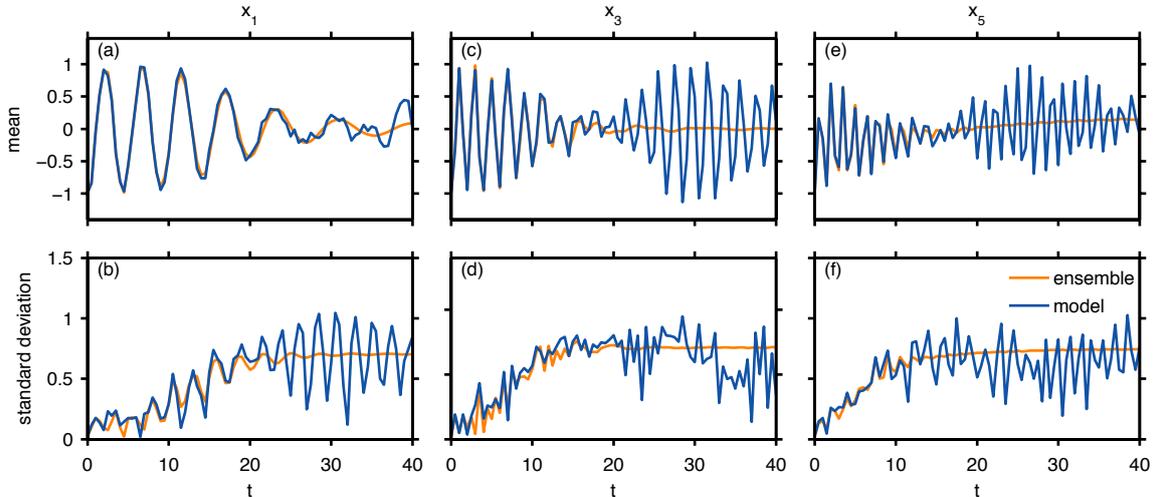


Figure 15: Nonparametric and ensemble forecasts of the components  $x^1$  (a, b),  $x^3$  (c, d), and  $x^5$  (e, f) of the flat embedding into  $\mathbb{R}^6$  for the mixing dynamical system on the 3-torus. The initial probability measure has circular Gaussian density relative to the Haar measure with location and concentration parameters  $(\bar{\theta}^1, \bar{\theta}^2, \bar{\theta}^3) = (\pi, \pi, \pi)$  and  $\kappa = 30$ , respectively. The nonparametric forecast is performed via Algorithm 4, modified for time change as described in Section 6.3. The ensemble forecast is based on 10,000 independent samples drawn from the initial probability measure and evolved using the perfect model.

from (40). Here, we use the location and concentration parameters  $\bar{\theta}^\mu = \pi$  and  $\kappa = 30$ . Also, we take the components  $\{x^1, x^3, x^5\}$  of the observation map as our forecast observables.

Figure 15 displays forecast results for the mean and standard deviation of these observables obtained via the method in Section 6.3 for  $l = 13$  ( $n = 19,683$ ) and an ensemble forecast with 10,000 particles using the perfect model. Because this system is mixing, expectation values with respect to the time-dependent probability measures  $\mu_t$  converge to expectation values with respect to  $\mu$  (see Remark 22). This late-time behavior is clearly evident in the ensemble forecast results, but as it turns out it is particularly challenging to emulate with the nonparametric model consisting of a finite collection of coupled oscillators. That is, the stochastic behavior of this mixing deterministic system involves the generation of arbitrarily small lengthscales in the time-dependent densities  $\rho_t$ , and as a result any diffusion-free algorithm that attempts to replicate this behavior via a finite collection of deterministic systems will ultimately fail at late times. In Fig. 15, the nonparametric model is able to accurately track the mean and standard deviation of the ensemble forecast over times less than the equilibrium relaxation time of the perfect model, but eventually develops spurious oscillations. The latter are possibly due to aliasing effects in the finite-bandwidth representation of  $\rho_t$ , and could be suppressed at a fixed time  $t$  by increasing  $l$ . However, failure to relax to the correct equilibrium should occur at long-enough times for fixed  $l$ . Note that the consequences of this shortcoming may not be particularly detrimental if the timescale at which the nonparametric model deviates from the perfect model is comparable or exceeds the intrinsic predictability limits for the given initial distribution and forecast observable (as is the case in Fig. 15).

### 6.5. Ergodic flow on the 2-torus with a fixed point

In this set of experiments, we consider the dynamical system on the 2-torus with the vector field  $v = \sum_{\mu=1}^2 v^\mu \frac{\partial}{\partial \theta^\mu}$ , where

$$v^1 = v^2 + (1 - \alpha)(1 - \cos \theta^2), \quad v^2 = \alpha(1 - \cos(\theta^1 - \theta^2)), \quad (48)$$

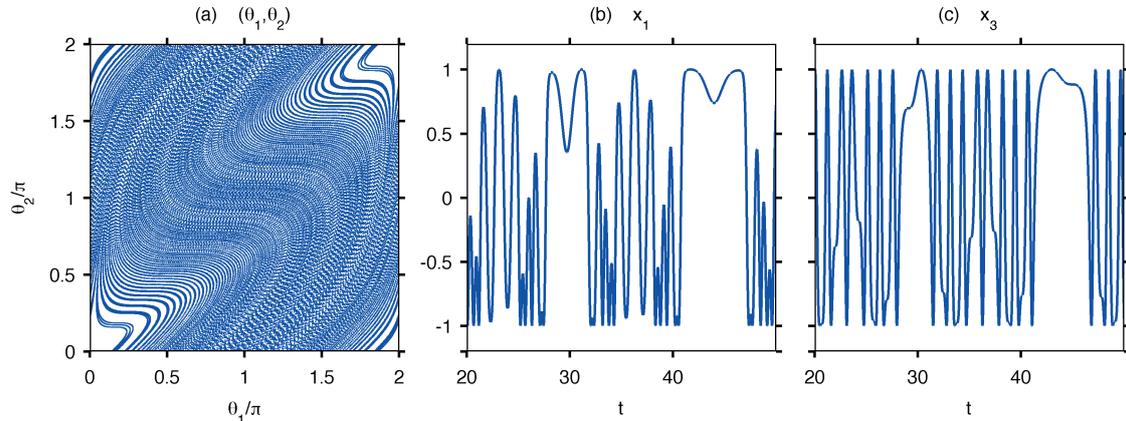


Figure 16: Time series for the fixed-point system on the 2-torus. (a) Phase space diagram on a periodic box for 64,000 samples; (b, c) components  $x^1 = \cos \theta^1$  and  $x^3 = \cos \theta^2$  of the standard (flat) embedding  $F : \mathbb{T}^2 \mapsto \mathbb{R}^4$ .

and  $\alpha$  is an irrational frequency parameter. Originally introduced by Oxtoby [51], flows of this class have a fixed point  $b$  with coordinates  $\theta^1 = \theta^2 = 0$ , and are ergodic for the Haar measure on the 2-torus.<sup>1</sup> Indeed, it is straightforward to check that  $\operatorname{div}_\mu v$  vanishes everywhere on  $\mathbb{T}^2$  for the Haar measure, which is a necessary and sufficient condition for the flow to be  $\mu$ -preserving. Due to the presence of the fixed point, this system is not uniquely ergodic, but the Haar measure is its only invariant Borel probability measure for which  $\mu(b) = 0$ . Besides the Haar measure, its only other ergodic Borel probability measure is the trivial measure with  $\mu(b) = 1$ . The system is topologically equivalent (i.e., equivalent up to homeomorphism) to a class of dynamical systems on the 2-torus called Stepanoff flows. The latter have the structure of a time-changed linear flow as in Section 6.1, but contain a single fixed point where the time change function diverges. Systems of this type are known to be topologically mixing, but we have not been able to find results on their measure-theoretic mixing properties apart from a conjecture in [51] that a class of Stepanoff flows that includes (48) is mixing. In summary, the system specified in (48) does not in itself have the structure of a time changed flow, but is connected to a (singular) time changed-system system via a continuous transformation. Our objective in this section is to demonstrate how the time-change methods of Section 6.3 behave in this more general context.

Qualitatively, the orbit of a point  $a \neq b$  under (48) will pass by the fixed point at arbitrarily small distances, but by incompressibility, the trajectory develops “bumps” and circumvents the fixed point when it comes close to it. Because the speed of the flow can become arbitrarily slow near the fixed point, the time series of observables of this system exhibit complex behavior over a broad range of timescales, with intervals of rapid evolution separated by quasi-stationary periods. This behavior is illustrated in Figure 16 for the frequency  $\alpha = 20^{1/2}$  and the standard (flat) embedding of the 2-torus in  $\mathbb{R}^4$ ,  $F(a) = (\cos \theta^1, \sin \theta^1, \cos \theta^2, \sin \theta^2)$ . In what follows, we discuss dimension reduction and forecasting results for this system and observation map using a time series of 128,000 samples initialized at  $(\theta^1, \theta^2) = (\pi/2, \pi/2)$  and sampled at a timestep  $T = 0.01$ .

First, we note that the methods of Section 4 with no time change fail in the initial diffusion maps stage, as the Laplace-Beltrami eigenfunctions computed via Algorithm 1 are corrupted by series of spikes (a hallmark of ill conditioning of the heat kernel matrix  $P$ ). We experimented with different kernels, tuning procedures, and normalizations (including the standard  $\alpha = 1/2$  normalization of diffusion maps that requires no density estimation), but in all cases the quality of the eigenfunctions was poor. This ill-conditioning is likely caused by the behavior of the system near the fixed point, where the sampling density through finite-time trajectories has a singular, “one-dimensional” structure (see Fig. 16(a)), even though the asymptotic sampling density

<sup>1</sup>The convention in [51] is that the frequency parameter lies in the unit interval; the cases with  $\alpha > 1$  are equivalent to systems with frequency  $1/\alpha$  up to an unimportant change of sign in  $v^1$  and a constant time change.

is uniform with respect to the Haar measure. On the other hand, after time change by the empirically accessible phase-space speed function  $\xi$ , the quality of the eigenfunctions from diffusion maps improves markedly. We attribute this improvement to the modified Riemannian metric  $\hat{h}$  from Section 6.3. This metric becomes degenerate near the fixed point where  $\xi$  vanishes, assigning arbitrarily small norm to all tangent vectors (this can also be seen from the fact that the kernel in (46) assigns near-maximal affinity to all pairs of points with small corresponding  $\xi$ ). Therefore, the heat kernel associated with  $\hat{h}$  produces stronger averaging (smoothing) near the fixed point resulting in a well-behaved eigenfunction basis which is crucial to the success of the techniques of Section 6.3.

In what follows, we work with the approximate eigenfunctions for the time-changed vector field computed using the advection-diffusion operator  $\hat{L}_\epsilon$  from (47) for the regularization parameter  $\epsilon = 0.02$ . We selected this value as a reasonable compromise between bias error and smoothness of the computed eigenfunctions after testing various candidate values of  $\epsilon$  in the interval  $10^{-4}$  to  $10^{-1}$ . As shown in Fig. 17, the generating eigenfunctions  $\{\zeta_1, \zeta_2\}$  for this value of  $\epsilon$  do not lie on the unit circle with the same accuracy as the earlier results in Figs. 4 and 13. Nevertheless, the eigenfunctions lie on a narrow annulus about the unit circle, and the corresponding time series have the structure of phase-modulated waves with weak amplitude modulation. The basic frequencies and Dirichlet energies are  $\{\Omega_1, \Omega_2\} = \{0.735, 0.165\}$  and  $\{E_{\zeta_1}, E_{\zeta_2}\} = \{1.54, 2.42\}$ . The eigenfunction time series exhibit timescale separation, with  $\zeta_1$  evolving at faster timescales than  $\zeta_2$ , but this timescale separation has a time-dependent nature in the sense that both time series evolve slowly near the fixed point. The timescale separation between  $\zeta_1$  and  $\zeta_2$  is also evident from the scatterplots on the torus in Fig. 17(a, d). There, it can be seen that the level sets of  $\zeta_2$  are aligned with the orbits of the dynamics, whereas the level sets of  $\zeta_1$  are transverse to the dynamics resulting to rapid oscillations due to frequent level-set crossings. As discussed in the SOM, EDMD implemented with a dictionary consisting of lags of the state vector fails to recover Koopman eigenfunctions of comparable quality to those in Fig. 17. In particular, as shown in Fig. 4 in the SOM, the EDMD spectrum contains an eigenfunction that somewhat resembles eigenfunction  $\zeta_2$  in Fig. 17, but is significantly more noisy. Moreover, we did not find evidence of an EDMD eigenfunction analogous to  $\zeta_1$  which varies predominantly in the direction along the flow as opposed to directions transverse to the flow.

Figure 17 shows the vector field decomposition from Lemma 21 associated with the generating eigenfunctions from Fig. 18. This decomposition was performed using the bandwidth parameter  $l = 30$ , and similarly to the 3-torus example of Section 6.4.1, the results were projected to a periodic box on the plane for visualization. With this choice of bandwidth, the time-averaged reconstruction error of the full vector field was 7%. Qualitatively, the vector field component  $V_1$  corresponding to  $\zeta_1$  describes a flow which is primarily directed along the  $\theta^2$  direction, apart from a band centered around the  $\theta^1 = \theta^2$  line where the flow turns towards the negative  $\theta^1$  direction and its magnitude is diminished. On the other hand, the flow described by the component  $V_2$  has significant magnitude along that band, together with inflow (outflow) from the  $\theta_2 > \theta_1$  ( $\theta_2 < \theta_1$ ) portions of the torus.

Turning now to nonparametric forecasting, we use an initial probability measure  $\mu_0$  from the same von-Mises family as the irrational flow example of Section 3.4. That is,  $\mu_0$  has the density function in (14) relative to the Haar measure, and in this case we set the location and concentration parameters to  $(\bar{\theta}^1, \bar{\theta}^2) = (\pi, \pi)$  and  $\kappa = 30$ , respectively. The dynamic evolution of the density for the time interval  $[0, 10]$  is illustrated with snapshots in Fig. 19 and as a video in Movie 2. With the chosen mean and concentration parameter, the initial density function is concentrated away from the fixed point at  $(0, 0)$ . In the course of dynamical evolution, the density function wanders close to the fixed point and becomes increasingly stretched. Eventually (around  $t = 350$ ), the density function appears to break up into disconnected components, but it is possible that these are artifacts caused by truncation to a finite number of basis functions and/or errors in the generating eigenfunctions and basic frequencies.

To assess the forecast skill of this nonparametric model, we compare its predictions against ensemble forecast obtained from 10,000 independent samples drawn from  $\rho_0$  and evolved using the true model from (48), taking the observation-map components  $x^1 = \cos \theta^1$  and  $x^3 = \cos \theta^2$  as our forecast observables. As shown in Fig. 20, over the time interval  $[0, 10]$  the mean forecast from the nonparametric model is in good agreement with the ensemble forecast. The nonparametric model also provides a reasonable uncertainty quantification

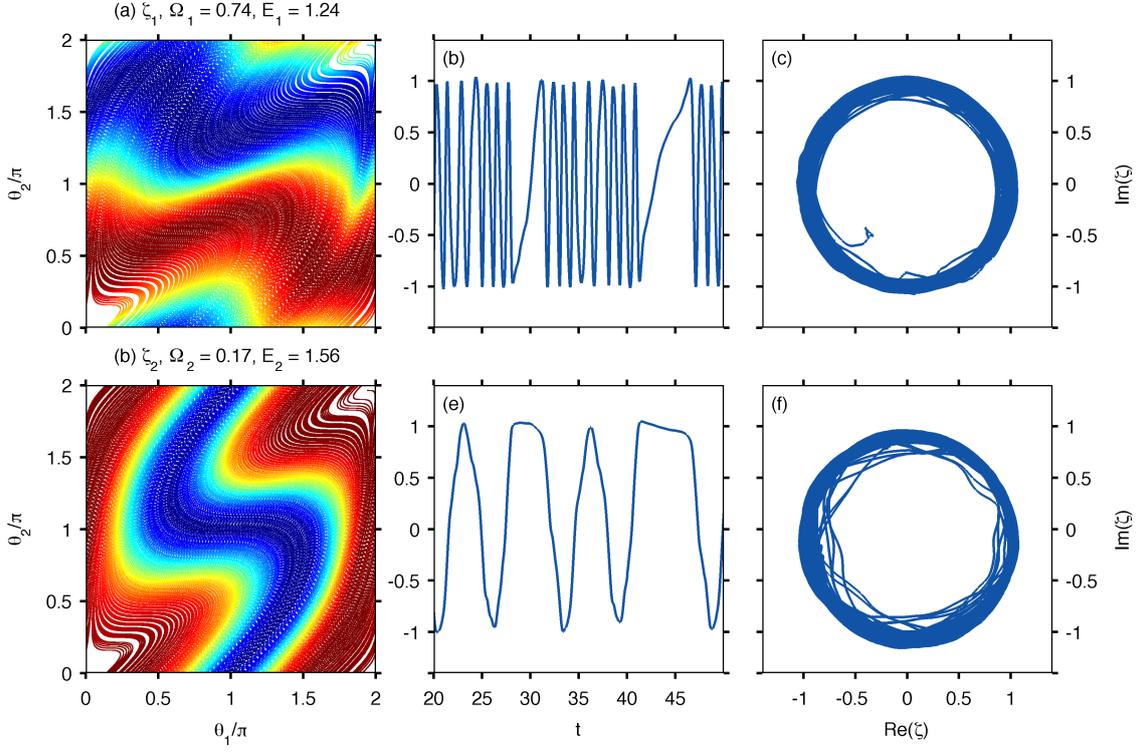


Figure 17: Generating eigenfunctions  $\zeta_1$  (a-c) and  $\zeta_2$  (d-f) for the fixed point system on the 2-torus obtained via the time-change technique. (a, d) Scatterplots of  $\text{Re}(\zeta_i)$ ; (b, e) time series of  $\text{Re}(\zeta_i)$ ; (c, f) scatterplots of  $(\text{Re}(\zeta_i), \text{Im}(\zeta_i))$ . The imaginary parts of  $\zeta_i$  are not shown in (b, e) for clarity, but to a good approximation they are  $90^\circ$  phase-shifted versions of  $\text{Re} \zeta_i$  apart from near fixed points where that phase relationship holds less accurately.

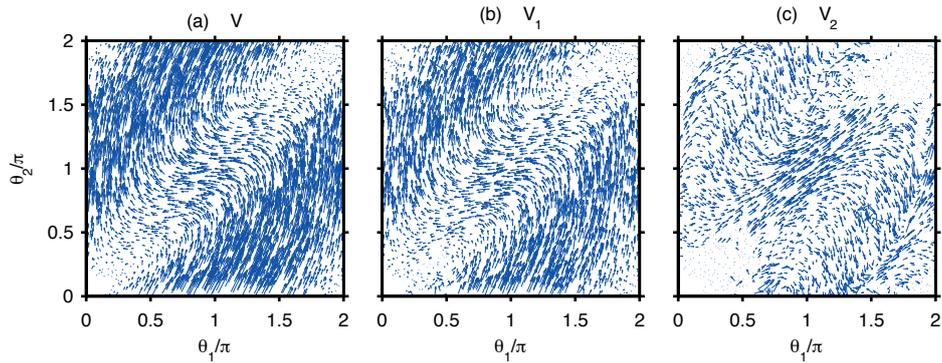


Figure 18: Vector field decomposition for the fixed-point system on the 2-torus. (a) Full vector field; (b, c) vector field components from Lemma 21. All vector fields are visualized in a periodic box for the angles  $(\theta^1, \theta^2)$ . As with Fig. 14, the vector fields were first reconstructed in data space,  $\mathbb{R}^4$ , and subsequently projected to the periodic domain. As a result, the arrow plots in (b, c) are noisier than the native vector fields in  $\mathbb{R}^4$ .

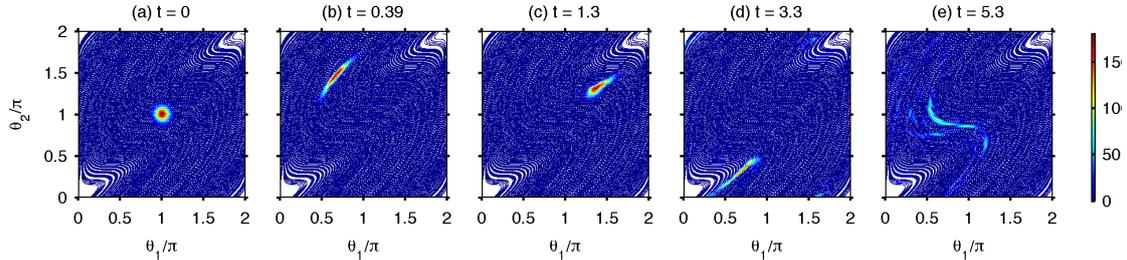


Figure 19: Snapshots of the time-dependent probability density  $\rho_t$  relative to the equilibrium (Haar) measure for the fixed-point system on the 2-torus. (a) At initialization time, the density is strongly concentrated around the point  $(\pi, \pi)$ ; (b, c) at early times, most of the probability mass is located away from the fixed point at  $(0, 0)$ ; (d) the peak of the density function approaches the fixed point and becomes highly stretched; (e) eventually, the numerical density function breaks up into disconnected components.

through the predicted standard deviation, but as with the previous experiments the deviations from the ensemble forecast are greater for the standard deviation than the mean. In summary, the time-change approach is particularly effective in this class of systems both in terms of the quality of the diffusion maps basis (which is useful in other contexts besides the Koopman operators studied here), but also as a regularization tool that transforms the dynamical system to an orbit-equivalent system with improved spectral properties.

## 7. Concluding remarks

In this work, we developed a family of data analysis techniques for dimension reduction, mode decomposition, and nonparametric forecasting of data generated by ergodic dynamical systems. Our approach is based on the Koopman and Perron-Frobenius formalisms for nonlinear dynamical systems, where the central objects of study are groups of unitary operators governing the evolution of observables and probability measures. In certain classes of systems (in particular, systems with pure point spectra), the spectral properties of these operators naturally lead to algorithms for dimension reduction and mode decomposition of spatiotemporal data, featuring timescale separation and strong invariance properties under changes of observation modality. The Perron-Frobenius operators also provide algorithms for equation-free forecasting of probability measures and expectation values of observables. Here, we develop these algorithms using a recent technique [44] for approximating the generator of the Koopman group in a complete orthonormal basis of the  $L^2$  space of the dynamical system, acquired from time-ordered data through the diffusion maps algorithm [37]. Placed in context, this work has connections with methods for mode decomposition and model reduction based on both the Koopman [1, 9–22] and Perron-Frobenius [23–34] perspectives, since the complete orthonormal basis learned from the data via diffusion maps allows us to pass between the two representations using standard linear algebra operations.

The ability to work in a complete orthonormal basis for the  $L^2$  space of the dynamical system with a well-defined notion of smoothness has a number of advantages which enable us to (1) construct nonlinear dimension reduction maps from Koopman eigenfunctions with projectible dynamics and small roughness on the data manifold; (2) decompose the generator of the Koopman group into a sum of mutually commuting vector fields, which we reconstruct in data space through a spectral representation of the pushforward map for vector fields on manifolds; (3) improve the efficiency and noise robustness of Galerkin methods for the Koopman eigenvalue problem through delay-coordinate maps; (4) predict the time-evolution of arbitrary probability densities and the expectation values of observables. These techniques perform best in the setting of systems with pure point spectra, where they lead to a decomposition of nonlinear dynamical systems into uncoupled simple harmonic oscillators. We demonstrated the efficacy of these methods in numerical experiments with variable-speed flows on the torus with multiple timescales, large contrasts of the sampling density in ambient data space, and strong i.i.d. observational noise. We also established an explicit connection between Koopman operators for systems with pure point spectra and the Laplace-

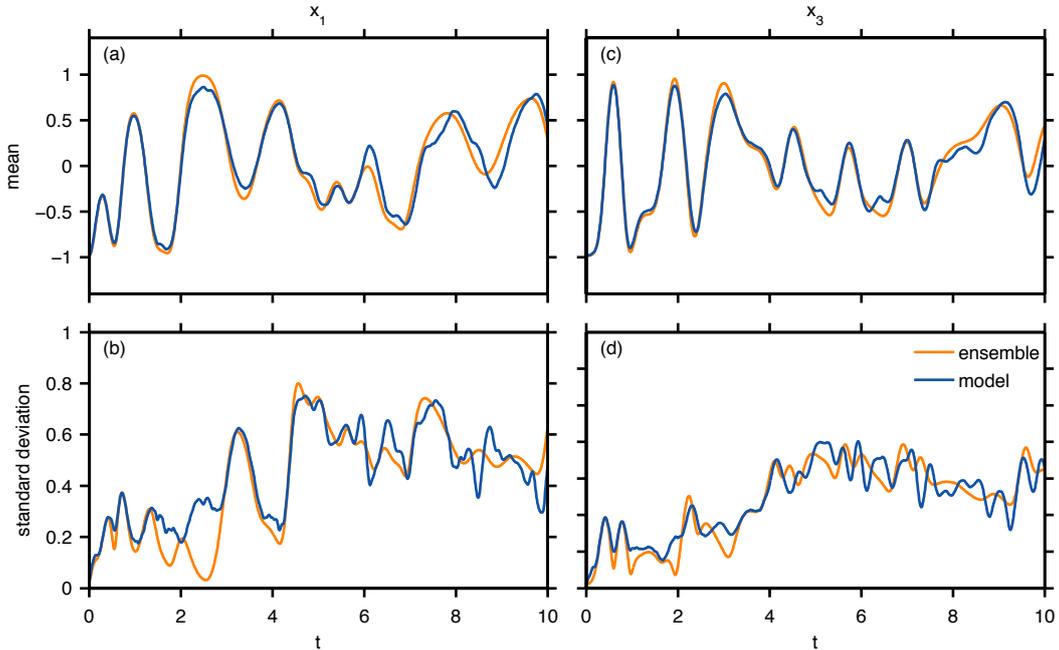


Figure 20: Nonparametric and ensemble forecasts of the mean and standard deviation of the components  $x^1$  (a, b) and  $x^4$  (c, d) of the canonical (flat) embedding of the 2-torus in  $\mathbb{R}^4$  for the fixed-point system. The initial probability measure has the circular Gaussian density in Fig. 19(a). The nonparametric forecast is performed via Algorithm 4, modified for time change as described in Section 6.3. The ensemble forecast is based on 10,000 independent samples drawn from the initial probability measure and evolved using the perfect model.

Beltrami operators approximated by diffusion maps applied to delay-coordinate mapped data, providing a rigorous justification of the timescale separation seen in diffusion coordinates [54] for this class of systems.

Another objective of this work has been to study and improve the regularity of numerically approximated eigenvalues and eigenfunctions of the Koopman group, particularly for mixing dynamical systems where the generator has no nonconstant eigenfunctions. In systems with pure point spectra, we demonstrated that adding a small amount of diffusion to the generator in an appropriate basis tailored to  $H_1$  regularity eliminates oscillatory eigenfunctions with large Dirichlet energy for the given observation modality. This type of regularization is important even in simple systems such as irrational flows, since the spectrum of the generator is dense in the imaginary line, and contains highly oscillatory eigenfunctions near (in the sense of the corresponding eigenvalues) eigenfunctions with low roughness.

In mixing systems, rather than regularizing the spectrum of the generator by diffusion alone (which is known to impart singular changes [81, 82] that are difficult to analyze in generality), we followed a different approach inspired by time-change methods in dynamical systems [50, 55]. In particular, we developed a strategy that involves rescaling the generator by its norm in the ambient data space and using the spectrum of the rescaled generator for dimension reduction and forecasting. This transformation is empirically computable from time-ordered data, and preserves the orbits of the dynamics while changing the flow of time along the orbits. In special cases, the transformation formally recovers a system with pure point spectrum from a class of time-changed mixing systems, but can be broadly used as an ad hoc regularization tool. We constructed analogs of our vector field decomposition and nonparametric forecasting techniques in the time-changed setting, where non-commuting vector fields and coupled oscillators encode the time-change function producing mixing. We demonstrated this approach in applications to a mixing flow on the 3-torus [50], and a challenging ergodic flow on the 2-torus with a fixed point [51] where diffusion maps fails to produce a well-conditioned basis if no time change is applied.

There is a number of areas for future research stemming from this work. First, the methods and applica-

tions discussed in the paper are heavily focused on dynamical systems on tori, so they should be investigated in more general classes of systems. In particular, we expect the techniques developed here to be useful in dynamical systems with Kronecker factors [55], i.e., systems possessing invariant subspaces of their full  $L^2$  space spanned by a (non-complete) set of eigenfunctions of the Koopman group. Also, it would be useful to study connections between Koopman operators and diffusion maps applied in delay-coordinate space for systems with nonzero Lyapunov exponents. Similarly, the time-change regularization strategy based on the empirical norm of the generator could be extended to other transformations having the goal of mapping the system under study to a system with improved spectral properties. At a more operational level, the algorithms formulated in this paper are all based on spectral expansions in global bases of  $L^2$  spaces, and these bases are generally inefficient in representing localized objects such as probability densities. It would therefore be fruitful to explore applications of multiscale bases (e.g., [87, 88]) as alternatives to the global  $L^2$  bases used here and in [44]. We plan to study these topics in future work.

## Acknowledgments

The author would like to thank Tyrus Berry, Peter Constantin, John Harlim, Fanghua Lin, Andrew Majda, Lai-Sang Young, and Zhizhen Zhao for stimulating conversations on this work. This research was supported by ONR grant N00014-14-1-0150, ONR MURI grant 25-74200-F7112, and NSF grant DMS-1521775.

## Appendix A. Algorithms

In this Appendix, we list the spectral decomposition and forecasting algorithms developed in Sections 4 and 6.

Algorithm 1 summarizes the construction of our data-driven orthonormal basis of  $L^2(M, \mu)$  via diffusion maps. This algorithm also includes a summary of the kernel density estimation and bandwidth-parameter tuning procedures. These procedures require a set  $\{\epsilon_l\}$  of candidate bandwidth parameters and a nearest-neighbor truncation parameter  $k_{\text{nn}}$  as additional inputs; see [43, 44] for complete descriptions. We note that the content of Algorithm 1 and the other algorithms in the paper are intended to be at a conceptual level, and some of the steps would be implemented differently in practice to ensure efficiency. In particular, in applications with  $N \gg 1$ , we can truncate  $P$  to  $k$  nearest neighbors (incurring small loss of accuracy due to exponential decay of  $K_\epsilon$ ). Also, it is customary for efficiency and stability to obtain the eigenvalues and eigenvectors of  $P$  from the eigenproblem for the symmetric matrix  $S = D^{1/2}PD^{-1/2}$ , where  $D$  is an  $N \times N$  diagonal matrix with  $D_{ii} = \sum_{j=1}^N H_{ij}$ . The eigenfunctions of  $P$  and the inner product weights  $w$  can be computed from the eigenvectors of  $S$  through the relationships in the last step of Algorithm 1. We refer the reader to [37, 43, 44, 58, 65] for further details on the numerical implementation and error estimates for this class of kernel algorithms.

### Algorithm 1 (Data-driven orthonormal basis).

- Inputs
  - Observed time series  $\{x_i\}_{i=0}^{N-1}$ ,  $x_i \in \mathbb{R}^d$ , at sampling interval  $T$
  - Candidate bandwidth parameter values  $\{\epsilon_l\}$  with  $\epsilon_l = 2^l$
  - Number of nearest neighbors  $k_{\text{nn}}$  for kernel density estimation
  - Manifold dimension  $m$
  - Desired number of eigenvalues and eigenfunctions  $n$
- Outputs
  - Sampling densities  $\{\hat{\sigma}_\epsilon(x_i)\}_{i=0}^{N-1}$  relative to the Riemannian measure

- Estimated manifold dimension  $\tilde{m}$
- Laplace-Beltrami eigenfunctions  $\{\phi_0, \phi_1, \dots, \phi_{n-1}\}$ ,  $\phi_i \in \mathbb{R}^N$ , and the corresponding eigenvalues  $\{\eta_0, \eta_1, \dots, \eta_{n-1}\}$
- Inner product weights  $w \in \mathbb{R}^N$

- Density estimation phase

1. For each  $x_i$ , compute the ad-hoc bandwidth function  $r_i^2 = \sum_{j=2}^{k_{\text{nn}}} \|x_i - x_{I(i,j)}\|^2 / (k_{\text{nn}} - 1)$ , where  $I(i, j)$  is the index of the  $j$ -th nearest neighbor of  $x_i$  in the dataset.
2. For each  $\epsilon_l$ , compute the sum  $\Sigma_l = \sum_{i,j=0}^{N-1} \tilde{K}_\epsilon(x_i, x_j) / N^2$  for the kernel

$$\tilde{K}_\epsilon(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (\epsilon_l r_i r_j)).$$

3. Choose the bandwidth parameter  $\epsilon \in \{\epsilon_l\}$  that maximizes  $\Sigma'_l = (\log \Sigma_{l+1} - \log \Sigma_l) / (\log \epsilon_{l+1} - \log \epsilon_l)$ . The estimated manifold dimension is  $\tilde{m} = 2\Sigma'_l$ .
4. With the bandwidth parameter from Step 2, compute the sampling density

$$\hat{\sigma}_\epsilon(x_i) = \sum_{j=0}^{N-1} \tilde{K}_\epsilon(x_i, x_j) / [N(w\epsilon r_i^2)^{m/2}].$$

- Main calculation phase

1. Select the bandwidth parameter  $\epsilon$  for the kernel  $K_\epsilon$  in (20) using the same method as Steps 2 and 3 of the density estimation phase.
2. Compute the vector  $q \in \mathbb{R}^N$  with  $q_i = \sum_{j=0}^{N-1} K_\epsilon(x_i, x_j)$  and the  $N \times N$  matrix  $H$  with  $H_{ij} = K_\epsilon(x_i, x_j) / q_j$ .
3. Compute the vector  $D \in \mathbb{R}^N$  with  $D_i = \sum_{j=0}^{N-1} H_{ij}$  and the  $N \times N$  symmetric matrix  $S$  with  $S_{ij} = H_{ij} / (D_i D_j)^{1/2}$ .
4. Solve the eigenvalue problem  $S\tilde{\phi}_i = \kappa_i \tilde{\phi}_i$  with  $i \in \{0, 1, \dots\}$  and  $\tilde{\phi}_i = (\tilde{\phi}_{0i}, \dots, \tilde{\phi}_{N-1,i}) \in \mathbb{R}^N$ .
5. Set the eigenvalues to  $\eta_i = \log \kappa_i / \log \kappa_1$ , the inner product weights to  $w = (w_0, \dots, w_{N-1}) \in \mathbb{R}^N$  with  $w_i = \tilde{\phi}_{i0}^2$ , and the eigenfunctions to  $\phi_i = (\phi_{0i}, \dots, \phi_{N-1,i})$  with  $\phi_{ji} = \tilde{\phi}_{ji} / \tilde{\phi}_{0i}$ .

Note that Algorithm 1 is not too sensitive to the choice of the nearest neighbor parameter  $k_{\text{nn}}$  and the candidate bandwidth parameter values. Hereafter, we will always work with the values  $k_{\text{nn}} = 8$  and  $\epsilon_l = 2^l$  with  $l \in \{-30, -29.9, \dots, 9.9, 10\}$ , which were also used in [44].

Algorithm 2 summarizes the numerical procedure to compute the generating frequencies and eigenfunctions of the Koopman group.

### Algorithm 2 (Generating frequencies and eigenfunctions).

- Inputs

- Diffusion eigenvalues  $\{\eta_i\}_{i=0}^n$ , eigenfunctions  $\{\varphi_i\}_{i=0}^n$ , and invariant measure  $w$  from Algorithm 1
- Manifold dimension  $m$
- Regularization parameter  $\epsilon$

- Outputs

- Generating frequencies  $\{\Omega_i\}_{i=1}^m$ ,  $\Omega_i \in \mathbb{C}$ , and eigenfunctions  $\{\zeta_i\}_{i=1}^n$ ,  $\zeta_i \in \mathbb{C}^N$

- Execution steps

1. Compute the  $n \times n$  matrices  $D$ ,  $B$ , and  $V$  from (28), and form the matrix  $A = V + \epsilon D$ .

2. Solve the generalized eigenvalue problem in (29); obtain the eigenvalues  $\{\gamma_i\}_{i=1}^{n'}$ ,  $\lambda_i \in \mathbb{C}$ , and the eigenvectors  $\{u_i\}_{i=1}^{n'}$ ,  $u_i \in \mathbb{C}^N$ , for  $n' \leq n$ .
3. Normalize each eigenvector to unit norm  $\|u_i\|_w$ , where  $\|u_i\|_w^2 = \sum_{j=1}^N w_j |u_{ji}|^2 / \eta_i$ .
4. Compute the Dirichlet energies of the eigenvectors,  $E(u_i) = \sum_{j=1}^n |u_{ji}|^2$ , and order the eigenvectors and eigenvalues in order of increasing  $E(u_i)$ .
5. Set  $\{\Omega_i\}_{i=1}^m$  to the first  $m$  rationally independent frequencies  $\text{Im } \gamma_i$  and  $\{\zeta_i\}_{i=1}^m$  to the corresponding eigenvectors.

Next, we summarize the main steps in the numerical implementation of the vector field decomposition from Theorem 5 and the statistical forecasting scheme in Section 3.3 in Algorithms 3 and 4, respectively. In these algorithms, the notation  $\text{diag } w$  refers to an  $N \times N$  diagonal matrix with diagonal elements equal to the components  $w_i$  of the inner product weight vector  $w$ . Moreover, in Algorithm 4 we present our nonparametric forecast scheme for general vector-valued observables in  $\mathbb{R}^s$ , instead of the scalar-valued observables discussed in the main text.

**Algorithm 3 (Vector field decomposition).**

- Inputs
  - Observed time series  $\{x_i\}_{i=0}^{N-1}$ ,  $x_i \in \mathbb{R}^d$ , at sampling interval  $T$
  - Invariant measure  $w \in \mathbb{R}^N$  from Algorithm 1
  - Basic frequencies  $\{\Omega_i\}_{i=1}^m$  and generating eigenfunctions  $\{\zeta_i\}_{i=1}^m$  from Algorithm 2
  - Spectral order parameter  $l$
- Outputs
  - Vector field components  $\{V_i\}_{i=1}^m$ , where  $V_i = \{V_{i0}, \dots, V_{i,N-1}\}$ , and  $V_{ij} \in \mathbb{R}^d$  is a tangent vector at  $x_j$
- Preparatory steps
  1. Arrange  $\{x_i\}_{i=0}^{N-1}$  into a  $d \times N$  data matrix  $x = (x_1 \cdots x_N)$ .
  2. Rescale the generators  $\zeta_i = (\zeta_{0i}, \dots, \zeta_{N-1,i}) \in \mathbb{C}^N$  to lie on the unit circle,  $\zeta_{ji} \leftarrow \zeta_{ji} / |\zeta_{ji}|$ .
  3. Construct the index set  $K = \{k_1, \dots, k_n\}$  with  $k_i = (q_1, \dots, q_m)$ ,  $|q_i| \leq l$  and  $n = (2l + 1)^m$ .
  4. For each  $i \in \{1, \dots, n\}$  form the basis vector  $z_i = \prod_{j=1}^m \zeta_j^{q_j}$  with  $(q_1, \dots, q_m) = k_i \in K$ . Arrange the  $z_k$  in the  $N \times n$  matrix  $z = (z_1 \cdots z_n)$ .
  5. Compute the Gramm matrix  $G \in \mathbb{C}^{n \times n}$  with  $G_{ij} = \langle z_i, z_j \rangle_w$ ; in matrix notation,  $G = z^\dagger \text{diag } w z$ .
- Execution steps
  1. Compute the expansion coefficient matrix  $\hat{F} \in \mathbb{C}^{N \times n}$  of the observation map s.t.  $\hat{F} = (\hat{F}_1 \cdots \hat{F}_n)$ ,  $\hat{F}_i = \langle z_i, F \rangle_w = \sum_{j=1}^N w_j x_j z_{ji}^*$ . In matrix notation,  $\hat{F} = x \text{diag } w z^*$ .
  2. For  $i \in \{1, \dots, m\}$  compute the diagonal matrix  $\bar{\Omega}_i \in \mathbb{R}^{n \times n}$ , where  $\bar{\Omega}_{i,jj} = q_i \Omega_i$  and  $q_i$  is the  $i$ -th element of the index vector  $k_j \in K$ .
  3. Reconstruct the vector fields in (8) by forming the  $d \times N$  matrices  $V_i = i \hat{F} G^{-1 \top} \bar{\Omega}_i z^\top$ . Set  $V_{ij}$  to the  $j$ -th column of  $V_i$ .

**Algorithm 4 (Nonparametric forecast).**

- Inputs
  - Observable time series  $\{f_i\}_{i=0}^{N-1}$ ,  $f_i \in \mathbb{R}^s$  with sampling interval  $T$
  - Invariant measure  $w \in \mathbb{R}^N$  from Algorithm 1

- Discretely sampled initial probability density  $\rho_0 = (\rho_{00}, \dots, \rho_{0,N-1}) \in \mathbb{R}^N$ , normalized such that  $\sum_{i=1}^N \rho_{0i} = 1$
  - Forecast times  $\{t_i\}_{i=1}^{N'}$  in multiples of  $T$
  - Basic frequencies  $\{\Omega_i\}_{i=1}^m$  and generating eigenfunctions  $\{\zeta_i\}_{i=1}^m$  from Algorithm 2
  - Spectral order parameter  $l$
- Outputs
    - Forecast densities  $\{\rho(t_i)\}_{i=1}^{N'}$ ,  $\rho(t_i) \in \mathbb{R}^N$
    - Expectation values  $\{\bar{f}(t_i)\}_{i=1}^{N'}$ ,  $\bar{f}(t_i) \in \mathbb{R}^s$ , of the observable
  - Preparatory steps
    1. Repeat the preparatory steps of Algorithm 3.
    2. Compute the frequency  $\omega_i = \sum_{j=1}^m q_j \Omega_j$  corresponding to each index vector  $k_i = (q_1, \dots, q_m) \in K$ .
  - Execution steps
    1. Compute the expansion coefficient matrix  $\hat{f} \in \mathbb{C}^{s \times n}$  of the observable s.t.  $\hat{f} = f \text{diag } w z^*$ .
    2. Compute the expansion coefficients  $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_n)$  of the initial density with  $\hat{\rho} = \rho_0 \text{diag } w z^*$ ; pre-multiply with the inverse Gramm matrix to form the coefficients  $\tilde{\rho}_0 = G^{-1} \hat{\rho}$ .
    3. For each forecast time  $t_i$ 
      - (a) Advance the expansion coefficients of the density using  $\tilde{\rho}_j(t_i) = e^{-i\omega_j t_i} \tilde{\rho}_j$ . Arrange the result in a row vector  $\tilde{\rho}(t_i) = (\tilde{\rho}_1(t_i), \dots, \tilde{\rho}_n(t_i)) \in \mathbb{C}^n$ .
      - (b) Reconstruct the density  $\rho(t_i)$  using  $\rho(t_i) = \tilde{\rho}(t_i) z^\top$ .
      - (c) Compute the expectation value of the observable  $\bar{f}(t_i) = \hat{f} \tilde{\rho}^\dagger(t_i)$ .

## Appendix B. Proof of Theorem 16

According to Theorem 5, the vector fields  $v_i$  are nowhere-vanishing and linearly independent. Therefore, to verify the desired expression for  $\bar{g}$  it suffices to show that  $\bar{g}(v_i, v_j) = B_{ij}$ . Indeed, using (8), (36), and Theorem 5, we find

$$\begin{aligned}
\bar{g}(v_i, v_j) &= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \langle F_* \hat{\Phi}_{k*} v_i, F_* \hat{\Phi}_{k*} v_j \rangle \\
&= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \langle \hat{\Phi}_{k*} v_i(F), \hat{\Phi}_{k*} v_j(F) \rangle \\
&= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \langle v_i(F \circ \hat{\Phi}_k), \hat{v}_j(F \circ \hat{\Phi}_k) \rangle \\
&= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \left\langle v_i \left( \sum_p \hat{F}_p^* (\zeta_1^{-p_1} \dots \zeta_m^{-p_m}) \circ \hat{\Phi}_k \right), v_j \left( \sum_q \hat{F}_q (\zeta_1^{q_1} \dots \zeta_m^{q_m}) \circ \hat{\Phi}_k \right) \right\rangle \\
&= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \left\langle v_i \left( \sum_p \hat{F}_p^* \zeta_1^{-p_1} \dots \zeta_m^{-p_m} e^{-\sum_{l=1}^m ik \Omega_l T p_l} \right), v_j \left( \sum_q \hat{F}_q \zeta_1^{q_1} \dots \zeta_m^{q_m} e^{-\sum_{n=1}^m ik \Omega_n T q_n} \right) \right\rangle \\
&= \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} \left\langle \left( \sum_p \hat{F}_p^* \Omega_i p_i \zeta_1^{-p_1} \dots \zeta_m^{-p_m} e^{-\sum_{l=1}^m ik \Omega_l T p_l} \right), \left( \sum_q \hat{F}_q \Omega_j q_j \zeta_1^{q_1} \dots \zeta_m^{q_m} e^{-\sum_{n=1}^m ik \Omega_n T q_n} \right) \right\rangle
\end{aligned}$$

$$\begin{aligned}
&= \sum_{p,q} \langle \hat{F}_p^*, \hat{F}_q \rangle \Omega_i \Omega_j p_i q_j \zeta_1^{q_1 - p_1} \dots \zeta_m^{q_m - p_m} \left( \lim_{s \rightarrow \infty} \frac{1}{s} \sum_{k=0}^{s-1} e^{\sum_{l=1}^m i k \Omega_l T (q_l - p_l)} \right) \\
&= \sum_{p,q} \langle \hat{F}_p^*, \hat{F}_q \rangle \Omega_i \Omega_j p_i q_j \zeta_1^{q_1 - p_1} \dots \zeta_m^{q_m - p_m} \delta_{q_1 p_1} \dots \delta_{q_m p_m} \\
&= \sum_p \|\hat{F}_p\|^2 \Omega_i \Omega_j p_i p_j.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
B_{ij} &= \int_M g(v_i, v_j) d\mu \\
&= \int_M \langle F_* v_i, F_* v_j \rangle d\mu \\
&= \int_M \langle v_i(F), v_j(F) \rangle d\mu \\
&= \int_M \left\langle v_i \left( \sum_p \hat{F}_p^* \zeta_1^{-p_1} \dots \zeta_m^{-p_m} \right), v_j \left( \sum_q \hat{F}_q^* \zeta_1^{q_1} \dots \zeta_m^{q_m} \right) \right\rangle d\mu \\
&= \int_M \left\langle \left( \sum_p \hat{F}_p^* \Omega_i p_i \zeta_1^{-p_1} \dots \zeta_m^{-p_m} \right), \left( \sum_q \hat{F}_q^* \Omega_j q_j \zeta_1^{q_1} \dots \zeta_m^{q_m} \right) \right\rangle d\mu \\
&= \sum_{p,q} \langle \hat{F}_p^*, \hat{F}_q \rangle \Omega_i \Omega_j p_i q_j \langle \zeta_1^{p_1} \dots \zeta_m^{p_m}, \zeta_1^{q_1} \dots \zeta_m^{q_m} \rangle \\
&= \sum_{p,q} \langle \hat{F}_p^*, \hat{F}_q \rangle \Omega_i \Omega_j p_i q_j \delta_{p_1 q_1} \dots \delta_{p_m q_m} \\
&= \sum_p \|\hat{F}_p\|^2 \Omega_i \Omega_j p_i p_j.
\end{aligned}$$

We therefore have  $\bar{g}(v_i, v_j) = B_{ij}$  as claimed above.

Next, to verify that  $\hat{\Phi}_{n,t*}$  is an isometry of  $\bar{g}$  as claimed in part (i), is sufficient to show that

$$\bar{g}(v_i|_a, v_j|_a) = \bar{g}(\hat{\Phi}_{n,t*} v_i|_a, \hat{\Phi}_{n,t*} v_j|_a)$$

for all  $i, j \in \{1, \dots, m\}$  and  $a \in M$ . This can be confirmed using Theorem 5(iv), according to which

$$\begin{aligned}
\bar{g}(\hat{\Phi}_{n,t*} v_i|_a, \hat{\Phi}_{n,t*} v_j|_a) &= \bar{g}(v_i|_{\hat{\Phi}_{n,t}(a)}, v_j|_{\hat{\Phi}_{n,t}(a)}) \\
&= \sum_{k,l=1}^m B_{kl}(\beta_k(v_i))|_{\hat{\Phi}_{n,t}(a)} (\beta_l(v_j))|_{\hat{\Phi}_{n,t}(a)} \\
&= B_{ij} \\
&= \bar{g}(v_i|_a, v_j|_a).
\end{aligned}$$

Moreover, to show that  $\bar{g}$  is flat as stated in part (ii) we note that each  $v_i$  is a Killing vector field since it generates a one-parameter group of isometries  $\Phi_{i,t}$ , and therefore by Theorem 5  $\{v_i\}_{i=1}^m$  is a set of  $m$  linearly independent, mutually commuting, Killing vector fields. It is a standard result from differential geometry that whenever such vector fields exist in an open neighborhood of a point in  $M$  the metric is flat at that point, and the global flatness of  $\bar{g}$  follows from the fact that the  $v_i$  are globally defined on  $M$ .

Finally, we check that  $\text{dvol}_{\bar{g}}/d\mu$  is a constant as stated in part (iii) working in local coordinates  $\{\theta^1, \dots, \theta^m\}$  such that  $v_i = \frac{\partial}{\partial \theta^i}$ . (Note that such coordinates exist at every point in  $M$  since the  $v_i$  are linearly independent and mutually commuting.) In these coordinates  $\bar{g}$  has constant components  $\bar{g}_{ij} = B_{ij}$ ,

and we have  $d\text{vol}_{\bar{g}} = \sqrt{\det B} d\theta^1 \wedge \cdots \wedge d\theta^m$  where  $B$  is the  $m \times m$  positive matrix with elements  $B_{ij}$ . Similarly, we can expand  $d\mu = \gamma d\theta^1 \wedge \cdots \wedge d\theta^m$ , where  $\gamma$  is a smooth positive function. Now, by Theorem 5(iii), the  $v_i$  have vanishing  $\mu$ -divergence, and using the local-coordinate expression for the divergence in (15) we obtain  $0 = \text{div}_{\mu} v_i = \gamma^{-1} \partial \gamma / \partial \theta^i$ . Thus,  $d\text{vol}_{\bar{g}} / d\mu = \sqrt{\det B} / \gamma = \Gamma$  is locally a constant, and since the  $v_i$  are smooth and defined at every point in  $M$ ,  $d\text{vol}_{\bar{g}} / d\mu$  is also globally constant.  $\square$

### Appendix C. Treatment of i.i.d. noise through delay embeddings and diffusion maps

In this Appendix, we study the asymptotic properties of the procedure employed in Section 5.2 to remove the effects of i.i.d. observational noise from the diffusion eigenfunctions computed through Algorithm 1. In particular, we are interested in the behavior of diffusion maps at small kernel bandwidth  $\epsilon$  and in the limit of infinitely many delays,  $s \rightarrow \infty$ , with an appropriate scaling relationship between  $s$  and the number of samples  $N$ .

**Theorem 23.** *Let  $f$  be an observable in  $L^2(M, \mu)$  with the values  $(f(a_{s-1}), \dots, f(a_{N-1})) = \vec{f}$  on the sampled states  $\{a_i\}_{i=s-1}^{N-1}$ . Under the assumptions on the noise stated in Section 5.2 and for an ergodic dynamical system with a pure point spectrum, as  $s, N \rightarrow \infty$  and  $s = o(N)$ , with high probability,*

$$(\tilde{P}\vec{f})_i = \mathcal{P}_{\epsilon} f(a_i) + O(\epsilon^2),$$

where  $\mathcal{P}_{\epsilon}$  is the averaging operator on  $L^2(M, \mu)$  constructed through (21) for the kernel in (38),  $\tilde{P}$  is the corresponding Markov matrix constructed from the noisy data via (C.3), and the estimate holds uniformly on  $M$ .

To prove Theorem 23, note that the noise variables in delay embedding space are sequences  $\Xi_i = (\xi_i, \xi_{i-1}, \dots, \xi_{i-s+1})$  of the i.i.d. noise variables  $\xi_i$ , and for  $s = o(N)$ , the probability (with respect to the sampling measure from the dynamics) that  $\Xi_i$  and  $\Xi_j$  have at least one overlapping element is  $O(N^{-1})$ . Since later on we will be taking  $N \rightarrow \infty$  limits, it suffices to consider the case  $|i - j| > s$  where  $\Xi_i$  and  $\Xi_j$  are independent. In that case,

$$\mathbb{E}(\|\Xi_i - \Xi_j\|^2) = \frac{1}{s} \sum_{k=0}^{s-1} \mathbb{E}(\|\xi_{i-k} - \xi_{j-k}\|^2) = \frac{1}{s} \sum_{k=0}^{s-1} (\mathbb{E}(\|\xi_{i-k}\|^2) + \mathbb{E}(\|\xi_{j-k}\|^2)) = 2R^2,$$

where we have used the facts that  $\mathbb{E}(\xi_i) = 0$  and  $\mathbb{E}(\|\xi_i\|^2) = R^2$ . Since the random variables  $\xi_i$  have finite moments up to order 4,  $\|\xi_i\|^2$  have finite variance and by the law of large numbers, as  $s \rightarrow \infty$ ,  $\|\Xi_i - \Xi_j\|^2$  converges almost surely to  $2R^2$ . Similarly, writing  $\tilde{X}_i - \tilde{X}_j = X_i - X_j + \Xi_i - \Xi_j$ , it follows that the squared pairwise distances  $\|\tilde{X}_i - \tilde{X}_j\|^2$  between the noisy data have the expectation

$$\mathbb{E}(\|\tilde{X}_i - \tilde{X}_j\|^2) = \|X_i - X_j\|^2 + \mathbb{E}(\|\Xi_i - \Xi_j\|^2) = \|X_i - X_j\|^2 + 2R^2.$$

Therefore, as  $s \rightarrow \infty$ ,  $\|\tilde{X}_i - \tilde{X}_j\|^2$  converges to  $\|X_i - X_j\|^2 + 2R^2$ ; i.e., the noisy pairwise distances  $\|\tilde{X}_i - \tilde{X}_j\|^2$  acquire a shift which is constant on the dataset. In particular, one can check that the variance of  $\|X_i - X_j\|^2$  is  $O(s^{-1})$ .

As stated in Section 5.2, one of the effects of the bias in the pairwise distances  $\|\tilde{X}_i - \tilde{X}_j\|^2$  is to introduce a bias in the density estimates  $\hat{\sigma}_{s,\epsilon}(\tilde{X}_i)$ . In particular, in the case of noisy data and as  $s \rightarrow \infty$ , the kernel  $\tilde{K}_{\epsilon}$  used in Algorithm 1 for density estimation takes the values  $\tilde{K}_{\epsilon}(\tilde{X}_i, \tilde{X}_j) = e^{-2R^2/(\epsilon r_i r_j)} K_{\epsilon}(X_i, X_j)$  with high probability. To remove the non-constant multiplicative bias term  $e^{-2R^2/(\epsilon r_i r_j)}$ , we replace  $\tilde{K}_{\epsilon}$  with the fixed bandwidth kernel  $\bar{K}_{\epsilon}(\tilde{X}_i, \tilde{X}_j) = e^{-\|\tilde{X}_i - \tilde{X}_j\|^2/\epsilon}$ , and perform density estimation with the normalized kernel

$$\kappa_{\epsilon}(\tilde{X}_i, \tilde{X}_j) = \frac{\bar{K}_{\epsilon}(\tilde{X}_i, \tilde{X}_j)}{(N - s + 1)^{-1} \sum_{k,l=s-1}^{N-1} \bar{K}_{\epsilon}(\tilde{X}_k, \tilde{X}_l)}. \quad (\text{C.1})$$

Note that the sum in the denominator in (C.1) runs from  $s - 1$  to  $N - 1$  (instead of 0 to  $N - 1$ ) since the first  $s - 1$  samples  $\tilde{x}_i$  are used for delay embedding.

As  $s \rightarrow \infty$  and for non-overlapping sequences  $\tilde{X}_i$  and  $X_j$ ,

$$\bar{K}_\epsilon(\tilde{X}_i, \tilde{X}_j) \xrightarrow{\text{a.s.}} e^{-2R^2/\epsilon} \bar{K}_\epsilon(X_i, X_j), \quad (\text{C.2})$$

where we used the fact that  $\bar{K}_\epsilon$  is continuous and bounded. We therefore see that  $\bar{K}_\epsilon(\tilde{X}_i, \tilde{X}_j)$  acquires a constant multiplicative bias  $e^{-2R^2/\epsilon}$ , but this bias appears in both the numerator and denominator in (C.1) and is therefore canceled. (At finite  $N$ , the cancellation is approximate since the sum in the denominator contains kernel values at potentially overlapping sequences  $\tilde{X}_i, \tilde{X}_j$  whose pairwise distance bias is smaller than  $2R^2$ , but the effect of these terms vanishes as  $N \rightarrow \infty$  and  $s = o(N)$ .)

Defining

$$\tau_{s,\epsilon}(\tilde{X}_i) = \frac{1}{(N - s + 1)} \sum_{j=s-1}^{N-1} \kappa_\epsilon(\tilde{X}_i, \tilde{X}_j),$$

it follows from the above arguments and the pointwise ergodic theorem that as  $s \rightarrow \infty$  and  $s = o(N)$ ,

$$\tau_{s,\epsilon}(\tilde{X}_i) \xrightarrow{\text{a.s.}} \bar{\tau}_\epsilon(X_i) = \lim_{s \rightarrow \infty} \int_M \kappa_\epsilon(F_s(a_i), F_s(a_j)) d\mu(a_j) = \lim_{s \rightarrow \infty} \frac{\int_M \bar{K}_\epsilon(F_s(a_i), F_s(a_j)) d\mu(a_j)}{\int_M (\int_M \bar{K}_\epsilon(F_s(a_k), F_s(a_l)) d\mu(a_k)) d\mu(a_l)},$$

where  $a_i$  the unique state in  $M$  underlying the noisy observation  $\tilde{X}_i$  (i.e.,  $\tilde{X}_i = X_i + \Xi_i$  with  $X_i = F_s(a_i)$ ). Using small-bandwidth asymptotics for Gaussian integrals on compact manifolds (e.g., [37, 65]), we obtain, uniformly on  $M$ ,

$$\begin{aligned} \frac{1}{\epsilon^{m/2}} \int_M \bar{K}_\epsilon(F_s(a_i), F_s(a_j)) d\mu(a_j) &= \frac{1}{\epsilon^{m/2}} \int_M \bar{K}_\epsilon(F_s(a_i), F_s(a_j)) \sigma_s(a_j) d\text{vol}_{g_s}(a_j) \\ &= c_s \sigma_s(a_i) + \sigma'_s(a_i) \epsilon + O(\epsilon^2), \end{aligned}$$

where  $\sigma_s = d\mu/d\text{vol}_{g_s}$  is the density of the invariant measure relative to the Riemannian measure at  $s$  delays,  $c_s$  is a constant that does not depend on  $\epsilon$ , and  $\sigma'_s$  is a function that vanishes if  $g_s$  is flat. Moreover, according to Theorem 16,  $\bar{g}$  is flat and has uniform volume form relative to  $\mu$ , so that  $\lim_{s \rightarrow \infty} \sigma_s = 1/\Gamma$ ,  $\lim_{s \rightarrow \infty} \sigma'_s = 0$ , and

$$\bar{\tau}_\epsilon(X_i) = \lim_{s \rightarrow \infty} \frac{\sigma_s(a_i) + \sigma'_s(a_i) \epsilon + O(\epsilon^2)}{\int_M [(\sigma_s(a_l))^2 + \sigma_s(a_l) \sigma'_s(a_l) \epsilon] d\text{vol}_{g_s}(a_l) + O(\epsilon^2)} = C^m + O(\epsilon^2),$$

where  $C$  is a positive constant. We therefore conclude that in the limit of large data and infinitely many delays  $\tau_\epsilon(\tilde{X}_i)$  converges up to  $O(\epsilon^2)$  to a constant on  $M$ . Since in this limit the sampling density relative to the Riemannian measure is also a constant,  $\tau_\epsilon(\tilde{X}_i)$  provides an  $O(\epsilon^2)$  estimate of the true sampling density up to a proportionality constant, which is of sufficient accuracy for the theory of variable-bandwidth kernels to apply [43].

Next, consider the diffusion maps normalization performed on the kernel  $K_\epsilon$  in (38) for the noisy data. Following the approach described in Section 4.2, we normalize  $K_\epsilon(\tilde{X}_i, \tilde{X}_j)$  to construct a Markov matrix  $\tilde{P}$  such that

$$\tilde{P}_{ij} = \frac{\tilde{H}_{ij}}{\sum_{k=s-1}^{N-1} \tilde{H}_{ik}}, \quad \tilde{H}_{ij} = \frac{K_\epsilon(\tilde{X}_i, \tilde{X}_j)}{\sum_{k=s-1}^{N-1} K_\epsilon(\tilde{X}_j, \tilde{X}_k)}. \quad (\text{C.3})$$

Then, we use  $\tilde{P}$  to approximate the action  $\mathcal{P}_\epsilon f$  of the averaging operator  $\mathcal{P}_\epsilon$  associated with the kernel on the observable  $f$  through the matrix-vector product  $\tilde{P}\vec{f}$ . By the same arguments used to write down (C.2) and the results on  $\tau_{s,\epsilon}$  derived above, as  $s \rightarrow \infty$ ,  $s = o(N)$ , and non-overlapping  $\tilde{X}_i, \tilde{X}_j$ ,

$$K_\epsilon(\tilde{X}_i, \tilde{X}_j) = \exp\left(-\frac{\|\tilde{X}_i - \tilde{X}_j\|^2}{\epsilon \tau_{s,\epsilon}^{-1/m}(\tilde{X}_i) \tau_{s,\epsilon}^{-1/m}(\tilde{X}_j)}\right) \xrightarrow{\text{a.s.}} \exp\left(-\frac{2R^2}{\epsilon(C^2 + O(\epsilon^2))}\right) K_\epsilon(X_i, X_j).$$

Using this result together with the pointwise ergodic theorem to convert sums in (C.3) to integrals with respect to  $\mu$ , we conclude that as  $s \rightarrow \infty$ ,  $(\tilde{P}f)_i = \mathcal{P}_\epsilon f(a_i) + O(\epsilon^2)$  with high probability, proving the theorem.  $\square$

We therefore see that provided that sufficiently many delays are used, the effect of i.i.d. noise is to produce an  $O(\epsilon^2)$  bias in the pointwise approximation of the action of  $\mathcal{P}_\epsilon$  on functions. However, that bias is of the same order as the error in approximating  $\Delta_{\tilde{g}}$  through  $\mathcal{P}_\epsilon$  (see (22)); that is, in the limit of infinitely many delays,  $(I - \tilde{P}f)_i/\epsilon$  approximates  $\Delta_{\tilde{g}}f(a_i)$  at the same accuracy as  $(I - \mathcal{P}_\epsilon)f(a_i)/\epsilon$ . We note again that in systems with nonzero Lyapunov exponents this type of noise removal is generally accompanied with biasing of the eigenfunctions extracted from  $\tilde{P}$  towards subspaces of the full  $L^2$  space of the dynamical system associated with stable or unstable Oseledec subspaces of the tangent bundle (for backward- and forward-looking delays, respectively).

## References

### References

- [1] M. Budisić, R. Mohr, I. Mezić, Applied Koopmanism, *Chaos* 22 (2012) 047510. doi:10.1063/1.4772195.
- [2] J. D. Farmer, J. J. Sidorowich, Predicting chaotic time series, *Phys. Rev. Lett.* 59 (8) (1987) 845–848. doi:http://dx.doi.org/10.1103/PhysRevLett.59.845.
- [3] T. Sauer, Time series prediction by using delay coordinate embedding, in: A. S. Weigend, N. A. Gerhsenfeld (Eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Vol. 15 of SFI Studies in the Sciences of Complexity, Addison-Wesley, 1003, pp. 175–193.
- [4] D. Kugiumtzis, O. C. Lingjaerde, N. Christophersen, Regularized local linear prediction of chaotic time series, *Phys. D.* 112 (1998) 344–360. doi:10.1016/S0167-2789(97)00171-1.
- [5] D. S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, *Complex Systems* 2 (1988) 321–355.
- [6] I. G. Kevrekides, C. W. Gear, G. Hummer, Equation-free: The computer-aided analysis of complex multiscale systems, *AIChE J.* 50 (7) (2004) 1346–1355. doi:10.1002/aic.10106.
- [7] R. Talmon, R. R. Coifman, Empirical intrinsic geometry for nonlinear modeling and time series filtering, *Proc. Natl. Acad. Sci.* 110 (31) (2013) 12535–12540.
- [8] M. Maggioni, M. Crosskey, ATLAS: A geometric approach to learning high-dimensional stochastic systems near manifolds (2014). arXiv:1404.0667.
- [9] I. Mezić, A. Banaszuk, Comparison of systems with complex behavior, *Phys. D.* 197 (2004) 101–133. doi:10.1016/j.physd.2004.06.015.
- [10] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, *Nonlinear Dyn.* 41 (2005) 309–325. doi:10.1007/s11071-005-2824-x.
- [11] I. Mezić, Analysis of fluid flows via spectral properties of the koopman operator, *Annu. Rev. Fluid Mech.* 45 (2013) 357–378. doi:10.1146/annurev-fluid-011212-140652.
- [12] P. J. Schmid, J. L. Sesterhenn, Dynamic mode decomposition of numerical and experimental data, in: *Bull. Amer. Phys. Soc.*, 61st APS meeting, San Antonio, 2008, p. 208.
- [13] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. S. Henningson, Spectral analysis of nonlinear flows, *J. Fluid Mech.* 641 (2009) 115–127. doi:10.1017/S0022112009992059.
- [14] K. K. Chen, J. H. Tu, C. W. Rowley, Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses, *J. Nonlinear Sci.* 22 (6) (2012) 887–915. doi:10.1007/s00332-012-9130-9.
- [15] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28. doi:10.1017/S0022112010001217.
- [16] M. R. Jovanović, P. J. Schmid, J. W. Nichols, Sparsity-promoting dynamic mode decomposition, *Phys. Fluids* 26 (2014) 024103. doi:10.1063/1.4863670.
- [17] J. H. Tu, C. W. Rowley, C. M. Luchtenburg, S. L. Brunton, J. N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comput. Dyn.* 1 (2) (2014) 391–421. doi:10.3934/jcd.2014.1.391.
- [18] M. S. Hemati, C. W. Rowley, E. A. Deem, L. N. Cattafesta, De-biasing the dynamic mode decomposition for applied Koopman spectral analysis (2015). arXiv:1502.03854.
- [19] M. O. Williams, I. G. Kevrekidis, C. W. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, *J. Nonlinear Sci.* (2015). doi:10.1007/s00332-015-9258-5.
- [20] J. N. Kutz, X. Fu, S. L. Brunton, Multiresolution dynamic mode decomposition, *SIAM J. Appl. Dyn. Sys.* 15 (2) (2016) 713–735. doi:10.1137/15M1023543.
- [21] J. L. Proctor, S. L. Brunton, J. N. Kutz, Dynamic mode decomposition with control, *SIAM J. Appl. Dyn. Sys.* 15 (1) (2016) 142–161.
- [22] M. Budisić, I. Mezić, Geometry of the ergodic quotient reveals coherent structures in flows, *Phys. D* 241 (2012) 1255–1269. doi:10.1016/j.physd.2012.04.006.

- [23] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, *SIAM J. Numer. Anal.* 36 (1999) 491. doi:10.1137/S0036142996313002.
- [24] M. Dellnitz, G. Froyland, On the isolated spectrum of the PerronFrobenius operator, *Nonlinearity* (2000) 1171–1188doi:10.1088/0951-7715/13/4/310.
- [25] G. Froyland, M. Dellnitz, Detecting and locating near-optimal invariant sets and cycles, *SIAM J. Sci. Comput.* 24 (6) (2003) 1839–1863. doi:10.1137/S106482750238911X.
- [26] G. Froyland, On Ulam approximation of the isolated spectrum and eigenfunctions of hyperbolic maps, *Discrete Cont. Dyn. S.* 17 (3) (2007) 671–689. doi:10.3934/dcds.2007.17.671.
- [27] G. Froyland, Unwrapping eigenfunctions to discover the geometry of almost-invariant sets in hyperbolic maps, *Phys. D.* 237 (2008) 840–853. doi:10.1016/j.physd.2007.11.004.
- [28] G. Froyland, C. González-Tokman, A. Quas, Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools, *J. Comput. Dyn.* 1 (2) (2014) 249–278. doi:10.3934/jcd.2014.1.249.
- [29] G. Froyland, Statistically optimal almost-invariant sets, *Phys. D.* 200 (2005) 205–219. doi:10.1016/j.physd.2004.11.008.
- [30] G. Froyland, K. Padberg, Almost-invariant sets and invariant manifolds – Connecting probabilistic and geometric descriptions of coherent structures in flows, *Phys. D* 238 (2009) 1507–1523. doi:10.1016/j.physd.2009.03.002.
- [31] G. Froyland, N. Santitissadeekorn, A. Monahan, Transport in time-dependent dynamical systems: Finite-time coherent sets, *Chaos* 20 (2010) 0431116. doi:10.1063/1.3502450.
- [32] G. Froyland, S. Lloyd, N. Santitissadeekorn, Coherent sets for nonautonomous dynamical systems, *Phys. D* 239 (2010) 1527–1541. doi:10.1016/j.physd.2010.03.009.
- [33] C. Schütte, W. Huisinga, P. Deuffhard, Transfer operator approach to conformational dynamics in biomolecular systems, in: B. Fiedler (Ed.), *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, Springer-Verlag, Berlin Heidelberg, 2001, pp. 191–223.
- [34] G. Froyland, G. A. Gottwald, A. Hammerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, *SIAM J. Appl. Dyn. Sys.* 13 (4) (2014) 1816–1846. doi:10.1137/130943637.
- [35] B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, *Proc. Natl. Acad. Sci.* 17 (5) (1931) 315–318.
- [36] D. Ruelle, Statistical mechanics of a one-dimensional lattice gas, *Comm. Math. Phys.* 9 (4) (1968) 267–278.
- [37] R. R. Coifman, S. Lafon, Diffusion maps, *Appl. Comput. Harmon. Anal.* 21 (2006) 5–30. doi:10.1016/j.acha.2006.04.006.
- [38] S. M. Ulam, *Problems in Modern Mathematics*, Dover Publications, Mineola, 1964.
- [39] P. Holmes, J. L. Lumley, G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, 1996.
- [40] J. A. Lee, M. Verleysen, *Nonlinear Dimensionality Reduction, Information Science and Statistics*, Springer, New York, 2007.
- [41] I. P. Kornfeld, Y. G. Sinai, Basic notions of ergodic theory and examples, in: Y. G. Sinai (Ed.), *Dynamical Systems, Ergodic Theory and Applications*, 2nd Edition, Vol. 100 of *Encyclopedia of Mathematical Sciences*, Springer, Berlin, 2000, pp. 2–30.
- [42] B. Hasselblatt, A. Katok, Principal structures, in: B. Hasselblatt, A. Katok (Eds.), *Handbook of Dynamical Systems*, Vol. 1A, North-Holland, Amsterdam, 2002, Ch. 1, pp. 1–203.
- [43] T. Berry, J. Harlim, Variable bandwidth diffusion kernels, *J. Appl. Comput. Harmon. Anal.*, in press (2015). doi:10.1016/j.acha.2015.01.001.
- [44] T. Berry, D. Giannakis, J. Harlim, Nonparametric forecasting of low-dimensional dynamical systems, *Phys. Rev. E.* 91 (2015) 032915. doi:10.1103/PhysRevE.91.032915.
- [45] N. H. Packard, et al., Geometry from a time series, *Phys. Rev. Lett.* 45 (1980) 712–716. doi:10.1103/physrevlett.45.712.
- [46] F. Takens, Detecting strange attractors in turbulence, in: *Dynamical Systems and Turbulence*, Vol. 898 of *Lecture Notes in Mathematics*, Springer, Berlin, 1981, pp. 366–381. doi:10.1007/bfb0091924.
- [47] T. Sauer, J. A. Yorke, M. Casdagli, Embedology, *J. Stat. Phys.* 65 (3–4) (1991) 579–616. doi:10.1007/bf01053745.
- [48] J. C. Robinson, A topological delay embedding theorem for infinite-dimensional dynamical systems, *Nonlinearity* 18 (5) (2005) 2135–2143. doi:10.1088/0951-7715/18/5/013.
- [49] E. R. Deyle, G. Sugihara, Generalized theorems for nonlinear state space reconstruction, *PLoS ONE* 6 (3) (2011) e18295. doi:10.1371/journal.pone.0018295.
- [50] B. Fayad, Analytic mixing reparametrizations of irrational flows, *Ergod. Th. & Dynam. Sys.* 22 (2002) 437–468. doi:10.1017/s0143385702000214.
- [51] J. C. Oxtoby, Stepanoff flows on the torus, *Proc. Amer. Math. Soc.* 4 (1953) 982–987.
- [52] D. Giannakis, J. Slawinska, Z. Zhao, Spatiotemporal feature extraction with data-driven Koopman operators, *J. Mach. Learn. Res. Proceedings* 44 (2015) 103–115.
- [53] D. Giannakis, A. J. Majda, Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability, *Proc. Natl. Acad. Sci.* 109 (7) (2012) 2222–2227. doi:10.1073/pnas.1118984109.
- [54] T. Berry, R. Cressman, Z. Gregurić-Ferenček, T. Sauer, Time-scale separation from diffusion-mapped delay coordinates, *SIAM J. Appl. Dyn. Sys.* 12 (2013) 618–649. doi:10.1137/12088183x.
- [55] A. Katok, J.-P. Thouvenot, Spectral properties and combinatorial constructions in ergodic theory, in: B. Hasselblatt, A. Katok (Eds.), *Handbook of Dynamical Systems*, Vol. 1B, North-Holland, Amsterdam, 2006, Ch. 11, pp. 649–743.
- [56] P. Walters, *An Introduction to Ergodic Theory*, Vol. 79 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, 1981.
- [57] Y. G. Sinai (Ed.), *Dynamical Systems, Ergodic Theory and Applications*, 2nd Edition, Vol. 100 of *Encyclopedia of Mathematical Sciences*, Springer, Berlin, 2000.

- [58] D. Giannakis, Dynamics-adapted cone kernels, *SIAM J. Appl. Dyn. Sys.* 14 (2) (2015) 556–608. doi:10.1137/140954544.
- [59] D. V. Anosov, A. B. Katok, New examples in smooth ergodic theory. *Ergodic diffeomorphisms*, *Trans. Moscow Math. Soc.* 23 (1970) 1–35.
- [60] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (2003) 1373–1396. doi:10.1162/089976603321780317.
- [61] M. Hein, J.-Y. Audibert, U. von Luxburg, From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians, in: *Learning Theory*, Vol. 3559 of *Lecture Notes in Computer Science*, Springer, Berlin, 2005, pp. 470–485. doi:10.1007/11503415\_32.
- [62] A. Singer, From graph to manifold Laplacian: The convergence rate, *J. Appl. Comput. Harmon. Anal.* 21 (2006) 128–134. doi:10.1016/j.acha.2006.03.004.
- [63] P. W. Jones, M. Maggioni, R. Schul, Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels, *Proc. Natl. Acad. Sci.* 105 (2008) 1803.
- [64] J. Portegies, Embeddings of Riemannian manifolds with heat kernels and eigenfunctions (2014). arXiv:1311.7568.
- [65] T. Berry, T. Sauer, Local kernels and the geometric structure of data, *J. Appl. Comput. Harmon. Anal.*, in press (2015). doi:10.1016/j.acha.2015.03.002.
- [66] A. Singer, R. R. Coifman, Non-linear independent component analysis with diffusion maps, *J. Appl. Comput. Harmon. Anal.* 25 (2008) 226–239. doi:10.1016/j.acha.2007.11.001.
- [67] R. R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, *Appl. Comput. Harmon. Anal.* 21 (2006) 31–52. doi:j.acha.2005.07.005.
- [68] N. Rabin, R. R. Coifman, Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids, in: *12th SIAM International Conference on Data Mining*, 2012, pp. 189–199. doi:10.1137/1.9781611972825.17.
- [69] Z. Zhao, D. Giannakis, Analog forecasting with dynamics-adapted kernels, in press (2016). arXiv:1412.3831.
- [70] U. von Luxburg, O. Bousquet, M. Belkin, On the convergence of spectral clustering on random samples: The normalized case, in: *17th Annual Conference on Learning Theory*, *COLT 2004*, Springer, 2004, pp. 457–471. doi:10.1007/978-3-540-27819-1\_32.
- [71] M. Belkin, P. Niyogi, Convergence of Laplacian eigenmaps, in: *Advances in Neural Information Processing Systems*, MIT Press, 2007, pp. 129–136.
- [72] U. von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering, *Ann. Stat.* 26 (2) (2008) 555–586. doi:10.1214/009053607000000640.
- [73] T. Berry, T. Sauer, Consistent manifold representation for topological data analysis (2016). arXiv:1606.02353.
- [74] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, A. Singer, Graph laplacian tomography from unknown random projections, *IEEE Trans. Image Process.* 17 (10) (2008) 1891–1899. doi:10.1109/tip.2008.2002305.
- [75] M. Hein, J.-Y. Audibert, Intrinsic dimensionality estimation of submanifolds in Euclidean space, in: *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 2005, pp. 289–296.
- [76] A. V. Little, J. Lee, Y.-M. Jung, M. Maggioni, Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD, in: *Proceedings of the 15th IEEE/SPWorkshop on Statistical Signal Processing*, 2009, pp. 85–88. doi:10.1109/SSP.2009.5278634.
- [77] I. Babuška, J. Osborn, Eigenvalue problems, in: P. G. Ciarlet, J. L. Lions (Eds.), *Finite Element Methods (Part 1)*, Vol. II of *Handbook of Numerical Analysis*, North-Holland, Amsterdam, 1991, pp. 641–787.
- [78] J. M. Melenk, L. P. Kirchner, C. Schwab, Spectral Galerkin discretization for hydrodynamic stability problems, *Computing* 65 (2000) 97–118. doi:10.1007/s006070070014.
- [79] D. Giannakis, P. F. Fischer, R. Rosner, A spectral Galerkin method for the coupled Orr-Sommerfeld and induction equations for free-surface MHD, *J. Comput. Phys.* 228 (4) (2009) 1188–1233. doi:10.1016/j.jcp.2008.10.016.
- [80] B. Franke, Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence-free drift, *Math. Z.* 246 (2004) 373–403. doi:10.1007/s00209-003-0604-1.
- [81] P. Constantin, A. Kiselev, L. Ryzhik, A. Zlatoš, Diffusion and mixing in fluid flow, *Ann. Math.* 168 (2008) 643–674. URL <http://www.jstor.org/stable/40345422>
- [82] B. Franke, C.-R. Hwang, H.-M. Pai, S. Sheu, The behavior of the spectral gap under growing drift, *Trans. Amer. Math. Soc.* 362 (3) (2010) 1325–1350. doi:10.1090/s0002-9947-09-04939-3.
- [83] V. I. Oseledets, A multiplicative ergodic theorem, *Trans. Moscow Math. Soc.* 19 (1968) 197–231.
- [84] L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998.
- [85] R. Fung, A. M. Hanna, O. Vendrell, S. Ramakrishna, T. Seideman, R. Santra, A. Ourmazd, Dynamics from noisy data with extreme timing uncertainty, *Nature* 532 (2016) 471–475. doi:10.1038/nature17627.
- [86] A. V. Kočergin, Time changes in flows and mixing, *Math. USSR Izv.* 7 (6) (1973) 1273–1294. doi:10.1070/IM1973v007n06ABEH002087.
- [87] R. R. Coifman, M. Maggioni, Diffusion wavelets, *Appl. Comput. Harmon. Anal.* 21 (1) (2006) 53–94. doi:10.1016/j.acha.2006.04.004.
- [88] W. K. Allard, G. Chen, M. Maggioni, Multi-scale geometric methods for data sets II: Geometric multi-resolution analysis, *Appl. Comput. Harmon. Anal.* 32 (3) (2012) 435–462. doi:10.1016/j.acha.2011.08.001.