



# Reproducing kernel Hilbert space compactification of unitary evolution groups



Suddhasattwa Das<sup>a</sup>, Dimitrios Giannakis<sup>a,\*</sup>, Joanna Slawinska<sup>b</sup>

<sup>a</sup> Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

<sup>b</sup> Finnish Center for Artificial Intelligence, Department of Computer Science, University of Helsinki, Helsinki, FI-00014, Finland

## ARTICLE INFO

### Article history:

Received 1 April 2019

Received in revised form 23

November 2020

Accepted 26 February 2021

Available online 3 March 2021

Communicated by E. Le Pennec

### Keywords:

Koopman operators

Perron-Frobenius operators

Ergodic dynamical systems

Reproducing kernel Hilbert spaces

Spectral theory

## ABSTRACT

A framework for coherent pattern extraction and prediction of observables of measure-preserving, ergodic dynamical systems with both atomic and continuous spectral components is developed. This framework is based on an approximation of the generator of the system by a compact operator  $W_\tau$  on a reproducing kernel Hilbert space (RKHS). The operator  $W_\tau$  is skew-adjoint, and thus can be represented by a projection-valued measure, discrete by compactness, with an associated orthonormal basis of eigenfunctions. These eigenfunctions are ordered in terms of a Dirichlet energy, and provide a notion of coherent observables under the dynamics akin to the Koopman eigenfunctions associated with the atomic part of the spectrum. In addition,  $W_\tau$  generates a unitary evolution group  $\{e^{tW_\tau}\}_{t \in \mathbb{R}}$  on the RKHS, which approximates the unitary Koopman group of the system. We establish convergence results for the spectrum and Borel functional calculus of  $W_\tau$  as  $\tau \rightarrow 0^+$ , as well as an associated data-driven formulation utilizing time series data. Numerical applications to ergodic systems with atomic and continuous spectra, namely a torus rotation, the Lorenz 63 system, and the Rössler system, are presented.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Characterizing and predicting the evolution of observables of dynamical systems is an important problem in the mathematical, physical, and engineering sciences, both theoretically and from an applications standpoint. A framework that has been gaining popularity [1–22] is the operator-theoretic approach to ergodic theory [23–25], where instead of directly studying the properties of the dynamical flow on state space, one characterizes the dynamics through its action on linear spaces of observables. The two classes of operators that have been predominantly employed in these approaches are the Koopman and Perron-Frobenius (transfer) operators, which are duals to one another on appropriate spaces of functions and measures, re-

\* Corresponding author.

E-mail address: [dimitris@cims.nyu.edu](mailto:dimitris@cims.nyu.edu) (D. Giannakis).

spectively. It is a remarkable fact, realized in the work of Koopman in the 1930s [26], that the action of a general nonlinear system on such spaces can be characterized through linear evolution operators, acting on observables by composition with the flow. Thus, despite the potentially nonlinear nature of the dynamics, many relevant problems, such as coherent pattern detection, statistical prediction, and control, can be formulated as intrinsically linear problems, making the full machinery of functional analysis available to construct approximation techniques.

The Koopman operator  $U^t$  associated with a continuous-time, continuous flow  $\Phi^t : M \rightarrow M$  on a manifold  $M$  acts on functions by composition,  $U^t f = f \circ \Phi^t$ . It is a contractive operator on the Banach space  $C^0(M)$  of bounded continuous functions on  $M$ , and a unitary operator on the Hilbert space  $L^2(\mu)$  associated with any invariant Borel probability measure  $\mu$ . Our main focus will be the latter Hilbert space setting, in which  $U = \{U^t\}_{t \in \mathbb{R}}$  becomes a unitary evolution group. In this setting, it is merely a matter of convention to consider Koopman operators instead of transfer operators, for the action of the transfer operator at time  $t$  on densities of measures in  $L^2(\mu)$  is given by the adjoint  $U^{t*} = U^{-t}$  of  $U^t$ . We will also assume that the invariant measure  $\mu$  is ergodic.

In this work, we seek to address the following two broad classes of problems:

- (i) *Coherent pattern extraction*; that is, identification of a collection of observables in  $L^2(\mu)$  having high regularity and an approximately periodic evolution under  $U^t$ . A precise notion of coherent observables stated in terms of Koopman eigenfunctions, or approximate Koopman eigenfunctions, will be given in (5) and (12), respectively.
- (ii) *Prediction*; that is, approximation of  $U^t f$  at arbitrary  $t \in \mathbb{R}$  for a fixed observable  $f \in L^2(\mu)$ .

Throughout, we require that the methods to address these problems are data-driven; i.e., they only utilize information from the values of a function  $F : M \rightarrow Y$  taking values in a data space  $Y$ , sampled finitely many times along an orbit of the dynamics.

### 1.1. Spectral characterization of unitary evolution groups

By Stone's theorem on one-parameter unitary groups [27,28], the Koopman group  $U$  is completely characterized by its generator—a densely defined, skew-adjoint, unbounded operator  $V : D(V) \rightarrow L^2(\mu)$  with domain  $D(V) \subset L^2(\mu)$ , where

$$Vf = \lim_{t \rightarrow 0} \frac{U^t f - f}{t}, \quad f \in D(V).$$

In particular, associated with  $V$  is a unique projection-valued measure (PVM)  $E : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(L^2(\mu))$  acting on the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$  on the real line and taking values in the space  $\mathcal{L}(L^2(\mu))$  of bounded operators on  $L^2(\mu)$ , such that

$$V = \int_{\mathbb{R}} i\omega \, dE(\omega), \quad U^t = \int_{\mathbb{R}} e^{i\omega t} \, dE(\omega). \quad (1)$$

The latter relationship expresses the Koopman operator at time  $t$  as an exponentiation of the generator,  $U^t = e^{tV}$ , which can be thought of as operator-theoretic analog of the exponentiation of a skew-symmetric matrix yielding a unitary matrix. In fact, the map  $V \mapsto e^{tV}$  is an application of the Borel functional calculus, whereby one lifts a Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$  on the imaginary line  $i\mathbb{R} \subset \mathbb{C}$ , to an operator-valued function

$$Z(V) = \int_{\mathbb{R}} Z(i\omega) dE(\omega), \quad (2)$$

acting on the skew-adjoint operator  $V$  via an integral against its corresponding PVM  $E$ .

The spectral representation of the unitary Koopman group can be further refined by virtue of the fact that  $L^2(\mu)$  admits the  $U^t$ -invariant orthogonal splitting

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

where  $H_p$  and  $H_c$  are closed orthogonal subspaces of  $L^2(\mu)$  associated with the atomic (point) and continuous components of  $E$ , respectively. On these subspaces, there exist unique PVMs  $E_p : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(H_p)$  and  $E_c : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(H_c)$ , respectively, where  $E_p$  is atomic and  $E_c$  is continuous, yielding the decomposition

$$E = E_p \oplus E_c. \quad (4)$$

We will refer to  $E_p$  and  $E_c$  as the point and continuous spectral components of  $E$ , respectively.

The subspace  $H_p$  is the closed linear span of the eigenspaces of  $V$  (and thus of  $U^t$ ). Correspondingly, the atoms of  $E_p$ , i.e., the singleton sets  $\{\omega_j\} \subset \mathbb{R}$  for which  $E_p(\{\omega_j\}) \neq 0$ , contain the eigenfrequencies of the generator. In particular, for every such  $\omega_j$ ,  $E_p(\{\omega_j\})$  is equal to the orthogonal projector to the eigenspace of  $V$  at eigenvalue  $i\omega_j$ , and all such eigenvalues are simple by ergodicity of the invariant measure  $\mu$ . As a result,  $H_p$  admits an orthonormal basis  $\{z_j\}$  satisfying

$$Vz_j = i\omega_j z_j, \quad U^t z_j = e^{i\omega_j t} z_j, \quad U^t f = \sum_j e^{i\omega_j t} \langle z_j, f \rangle_\mu z_j, \quad \forall f \in H_p, \quad (5)$$

where  $\langle \cdot, \cdot \rangle_\mu$  is the inner product on  $L^2(\mu)$ . It follows from the above that the Koopman eigenfunctions form a distinguished orthonormal basis of  $H_p$ , whose elements  $z_j$  evolve under the dynamics by multiplication by periodic phase factors at the frequencies  $\omega_j$ , even if the underlying dynamical flow is nonlinear and aperiodic. In contrast, observables  $f \in H_c$  do not exhibit an analogous quasiperiodic evolution, and are characterized instead by a weak-mixing property (decay of correlations), typical of chaotic dynamics,

$$\frac{1}{t} \int_0^t |\langle g, U^s f \rangle_\mu| ds \xrightarrow[t \rightarrow \infty]{} 0, \quad \forall g \in L^2(\mu).$$

## 1.2. Pointwise and spectral approximation techniques

While the two classes of pattern extraction and prediction problems listed above are obviously related by the fact that they involve the same evolution operators, in some aspects they are fairly distinct, as for the latter it is sufficient to perform pointwise (or even weak) approximations of the operators, whereas the former are fundamentally of a spectral nature. In particular, observe that a convergent approximation technique for the prediction problem can be constructed by taking advantage of the fact that  $U^t$  is a bounded (and therefore continuous) linear operator, without explicit consideration of its spectral properties. That is, given an arbitrary orthonormal basis  $\{\phi_0, \phi_1, \dots\}$  of  $L^2(\mu)$  with associated orthogonal projection operators  $\Pi_L : L^2(\mu) \rightarrow \text{span}\{\phi_0, \dots, \phi_{L-1}\}$ , the finite-rank operator  $U_L^t = \Pi_L U^t \Pi_L$  is fully characterized by the matrix elements  $U_{ij}^t = \langle \phi_i, U^t \phi_j \rangle_\mu$  with  $0 \leq i, j \leq L-1$ , and by continuity of  $U^t$ , the sequence of operators  $U_L^t$  converges pointwise to  $U^t$ . Thus, if one has access to data-driven approximations  $U_{N,ij}^t$  of  $U_{ij}^t$  determined from  $N$  measurements of  $F$  taken along an orbit of the dynamics, and these approximations converge as

$N \rightarrow \infty$ , then, as  $L \rightarrow \infty$  and  $N \gg L$ , the corresponding finite-rank operators  $U_{N,L}^t$  converge pointwise to  $U^t$ .

This property was employed in [12] in a technique called diffusion forecasting, whereby the approximate matrix elements  $U_{N,ij}^t$  are evaluated in a data-driven basis constructed from samples of  $F$  using the diffusion maps algorithm (a kernel algorithm for manifold learning) [29]. By spectral convergence results for kernel integral operators [30] and ergodicity, as  $N \rightarrow \infty$ , the data-driven basis functions converge to an orthonormal basis of  $L^2(\mu)$  in an appropriate sense, and thus the corresponding approximate Koopman operators  $U_{N,L}^t$  converge pointwise to  $U^t$  as described above. In [12], it was demonstrated that diffusion forecasts of observables of the Lorenz 63 (L63) system [31] have skill approaching that of ensemble forecasts using the true model, despite the fact that the Koopman group in this case has a purely continuous spectrum (except from the trivial eigenfrequency at 0). Pointwise-convergent approximation techniques for Koopman operators were also studied in [20,32] in the context of extended dynamic mode decomposition (EDMD) algorithms [14]. However, these methods require the availability of an orthonormal basis of  $L^2(\mu)$  of sufficient regularity, which, apart from special cases, is difficult to have in practice (particularly when the support of  $\mu$  is an unknown, measure-zero subset of the ambient state space  $M$ ).

Of course, this is not to say that the spectral decomposition in (4) is irrelevant in a prediction setting, for it reveals that an orthonormal basis of  $L^2(\mu)$  that splits between the invariant subspaces  $H_p$  and  $H_c$  would yield a more efficient representation of  $U^t$  than an arbitrary basis. This representation could be made even more efficient by choosing the basis of  $H_p$  to be a Koopman eigenfunction basis (e.g., [17]). Still, so long as a method for approximating a basis of  $L^2(\mu)$  is available, arranging for compatibility of the basis with the spectral decomposition of  $U^t$  is a matter of optimizing performance rather than ensuring convergence.

In contrast, as has been recognized since the earliest techniques in this area [1–4], in coherent pattern extraction problems the spectral properties of the evolution operators play a crucial role from the outset. In the case of measure-preserving ergodic dynamics studied here, the Koopman eigenfunctions in (5) provide a natural notion of temporally coherent observables that capture intrinsic frequencies of the dynamics. Unlike the eigenfunctions of other operators commonly used in data analysis (e.g., the covariance operators employed in the proper orthogonal decomposition [33]), Koopman eigenfunctions have the property of being independent of the observation map  $F$ , thus leading to a definition of coherence that is independent of the observation modality used to probe the system. In applications in fluid dynamics [6,34], climate dynamics [35], and many other domains, it has been found that the patterns recovered by Koopman eigenfunction analysis have high physical interpretability and ability to recover dynamically significant timescales from multiscale input data.

### 1.3. Review of existing methodologies

Despite the attractive theoretical properties of evolution operators, the design of data-driven spectral approximation techniques that can naturally handle both point and continuous spectra, with rigorous convergence guarantees, is challenging, and several open problems remain. As an illustration of these challenges, and to place our work in context, it is worthwhile noting that besides approximating the continuous spectrum (which is obviously challenging), rigorous approximation of the atomic spectral component  $E_p$  is also non-trivial, since, apart from the case of circle rotations, it is concentrated on a dense, countable subset of the real line. In applications, the density of the atomic part of the spectrum and the possibility of the presence of a continuous spectral component necessitate the use of some form of regularization to ensure well-posedness of spectral approximation schemes. In the transfer operator literature, the use of regularization techniques such as domain restriction to function spaces where the operators are quasicompact [2], or compactification by smoothing by kernel integral operators [8], has been prevalent, though these methods

may require more information than the single observable time series assumed to be available here. On the other hand, many of the popular techniques in the Koopman operator literature, including the dynamic mode decomposition (DMD) [6,7] and EDMD [14] do not explicitly consider regularization, and instead implicitly regularize the operators by projection onto finite-dimensional subspaces (e.g., Krylov subspaces and subspaces spanned by general dictionaries of observables). Despite the practical simplicity of this approach, controlling its asymptotic behavior as the dimension of the approximation space increases is difficult; see, e.g., Fig. 5 in Sect. 9.

To our knowledge, the first spectral convergence results for EDMD [16] were obtained for a variant of the framework called Hankel-matrix DMD [15], which employs dictionaries constructed by application of delay-coordinate maps [36] to the observation function. However, these results are based on an assumption that the observation map lies in a finite-dimensional Koopman invariant subspace (which must be necessarily a subspace of  $H_p$ ); an assumption unlikely to hold in practice. This assumption is relaxed in [20], who establish weak spectral convergence results implied by strongly convergent approximations of the Koopman operator derived through EDMD. This approach makes use of an a priori known orthonormal basis of  $L^2(\mu)$ , the availability of which is not required in Hankel-matrix DMD.

A fairly distinct class of approaches to (E)DMD perform spectral estimation for Koopman operators using harmonic analysis techniques [3,4,21,22]. Among these, [3,4] consider a spectral decomposition of the Koopman operator closely related to (4), though expressed in terms of spectral measures on  $S^1$  as appropriate for unitary operators, and utilize harmonic averaging (discrete Fourier transform) techniques to estimate eigenfrequencies and the projections of the data onto Koopman eigenspaces. While this approach can theoretically recover the correct eigenfrequencies corresponding to eigenfunctions with nonzero projections onto the observation map, its asymptotic behavior in the limit of large data exhibits a highly singular dependence on the frequency employed for harmonic averaging—this hinders the construction of practical algorithms that converge to the true eigenfrequencies by examining candidate eigenfrequencies in finite sets. The method also does not address the problem of approximating the continuous spectrum, or the computation of Koopman eigenfunctions on the whole state space (as opposed to eigenfunctions computed on orbits).

The latter problem was addressed in [22], who employed the theory of reproducing kernel Hilbert spaces (RKHSs) [37,38] to identify conditions for a candidate frequency  $\omega \in \mathbb{R}$  to be a Koopman eigenfrequency based on the RKHS norm of the corresponding Fourier function  $e^{i\omega t}$  sampled on an orbit. For the frequencies meeting these criteria, they constructed pointwise-defined Koopman eigenfunctions in RKHS using out-of-sample extension techniques [39]. While this method also suffers from a singular behavior in  $\omega$ , it was found to perform better than conventional harmonic averaging, particularly in mixed-spectrum systems with non-trivial atomic and continuous spectral components simultaneously present. However, the question of approximating the continuous spectrum remains moot. RKHS-based approaches for spectral analysis of Koopman operators have also been proposed in [40,41], though these methods rely on the strong assumption that the Koopman operator maps the RKHS into itself. The latter is known to be satisfied only in special cases, such as RKHSs with flow-invariant reproducing kernels [22, Corollary 9].

In [21], a promising approach for estimating both the atomic and continuous parts of the spectrum was introduced, based on spectral moment estimation techniques. This approach consistently approximates the spectral measure of the Koopman operator on the cyclic subspace associated with a given scalar-valued observable, and is also capable of identifying its atomic, absolutely continuous, and singular continuous components. However, since it operates on cyclic subspaces associated with individual observables, it is potentially challenging to extend to applications involving a high-dimensional data space  $Y$ , including spatiotemporal systems where the dimension of  $Y$  is formally infinite.

In [13,17,18] a different approach was taken, focusing on approximations of the eigenvalue problem for the skew-adjoint generator  $V$ , as opposed to the unitary Koopman operators  $U^t$ , in an orthonormal basis of an invariant subspace of  $H_p$  (of possibly infinite dimension) learned from observed data

via kernel algorithms [29,30,42–44] as in diffusion forecasting. A key ingredient of these techniques is a family  $K_1, K_2, \dots$  of kernel integral operators on  $L^2(\mu)$  constructed from delay-coordinate-mapped data with  $Q$  delays, such that, in the infinite-delay limit,  $K_Q$  converges in norm to a compact integral operator  $K_\infty : L^2(\mu) \rightarrow L^2(\mu)$  commuting with  $U^t$  for all  $t \in \mathbb{R}$ . Because commuting operators have common eigenspaces, and the eigenspaces of compact operators at nonzero corresponding eigenvalues are finite-dimensional, the eigenfunctions of  $K_\infty$  (approximated by eigenfunctions of  $K_Q$  at large  $Q$ ) provide a highly efficient basis to perform Galerkin approximation of the Koopman eigenvalue problem. In [13,17,18], a well-posed variational eigenvalue problem was formulated by regularizing the raw generator  $V$  by the addition of a small amount of diffusion, represented by a positive-semidefinite self-adjoint operator  $\Delta : D(\Delta) \rightarrow L^2(\mu)$  on a suitable domain  $D(\Delta) \subset D(V)$ . This leads to an advection-diffusion operator

$$L = V - \theta\Delta, \quad \theta > 0, \quad (6)$$

whose eigenvalues and eigenfunctions can be computed through Galerkin schemes based on classical approximation theory for variational eigenvalue problems [45]. The diffusion operator in (6) is constructed so as to commute with  $V$ , so that every eigenfunction of  $L$  is a Koopman eigenfunction, with eigenfrequency equal to the imaginary part of the corresponding eigenvalue. Moreover, it was shown that the variational eigenvalue problem for  $L$  can be consistently approximated from time series data acquired via a generic observation map.

Advection-diffusion operators as in (6) can, in some cases, also provide a notion of coherent observables in the continuous spectrum subspace  $H_c$ , although from this standpoint the results are arguably not very satisfactory. In particular, it follows from results in [46] that if the support  $X \subseteq M$  of the invariant measure  $\mu$  has manifold structure, and  $\Delta$  is chosen to be a Laplacian or weighted Laplacian for a suitable Riemannian metric, then the spectrum of  $L$  contains only isolated eigenvalues, irrespective of the presence of continuous spectrum [17]. However, if  $V$  has a non-empty continuous spectrum, then there exists no smooth Riemannian metric whose corresponding Laplacian commutes with  $V$ , meaning that  $L$  is necessarily non-normal. The spectra of non-normal operators can have several undesirable, or difficult to control, properties, including extreme sensitivity to perturbations and failure to have a complete basis of eigenvectors. The behavior of  $L$  is even more difficult to characterize if  $X$  is not a smooth manifold, and  $V$  possesses continuous spectrum. In [13,17,18], these difficulties are avoided by effectively restricting  $V$  to an invariant subspace of  $H_p$  through a careful choice of data-driven basis, but this approach provides no information about the ability of the method to identify coherent observables in  $H_c$ .

Operators analogous to  $L$  in (6), acting on suitable spaces of distributions, have also been shown to consistently approximate the spectrum of the generator of Anosov flows [47,48], allowing, in particular, to recover Pollicott-Ruelle resonances [49,50] in such systems through zero viscosity ( $\theta \rightarrow 0$ ) limits. However, these approaches make extensive use of the hyperbolic structure of Anosov flows, which is not exhibited by the more general class of ergodic flows studied here. Put together, these facts motivate a different regularization approach to (6) that can seamlessly handle both the point and continuous spectra of  $V$ , while being amenable to data-driven approximation.

#### 1.4. Contributions of this work

In this paper, we propose a data-driven framework for pattern extraction and prediction in measure-preserving, ergodic dynamical systems, which retains the advantageous aspects of [12,13,17,18] through the use of kernel integral operators to provide orthonormal bases of appropriate regularity, while being naturally adapted to dynamical systems with arbitrary (pure point, mixed, or continuous) spectral characteristics. The key element of our approach is to replace the diffusion regularization in (6) by a *compactification* of

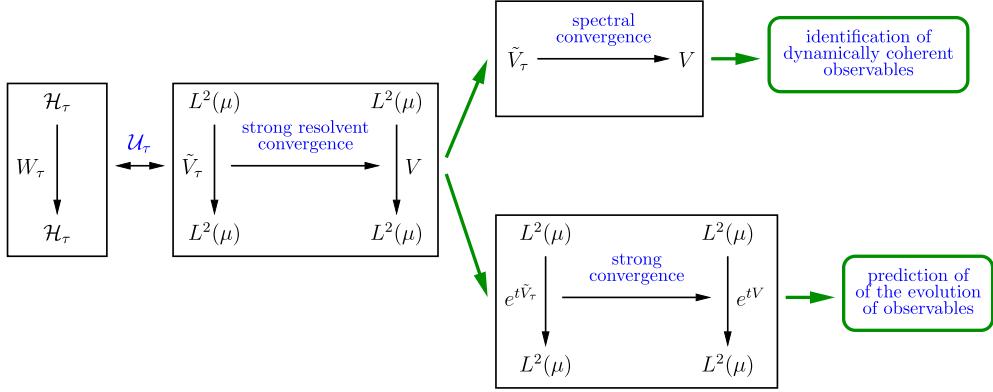
the skew-adjoint generator  $V$  of such systems (which is unbounded, and has complicated spectral behavior), mapping it to a family of compact, skew-adjoint operators  $W_\tau : \mathcal{H}_\tau \rightarrow \mathcal{H}_\tau$ ,  $\tau > 0$ , each acting on an RKHS  $\mathcal{H}_\tau$  of functions on the state space manifold  $M$ . In fact, the operators  $W_\tau$  are not only compact, they are trace-class integral operators with continuous kernels. Moreover, the spaces  $\mathcal{H}_\tau$  employed in this framework are dense in  $L^2(\mu)$ , and have Markovian reproducing kernels. We use the unitary operator group  $\{e^{tW_\tau}\}_{t \in \mathbb{R}}$  generated by  $W_\tau$  as an approximation of the Koopman group  $U$ , and establish spectral and pointwise convergence as  $\tau \rightarrow 0$  in an appropriate sense. This RKHS approach has the following advantages.

- (i) The fact that  $W_\tau$  is skew-adjoint avoids non-normality issues, and allows decomposition of these operators in terms of unique PVMs  $\mathcal{E}_\tau : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(\mathcal{H}_\tau)$ . The existence of  $\mathcal{E}_\tau$  allows in turn the construction of a Borel functional calculus for  $W_\tau$ , meaning in particular that operator exponentiation,  $e^{tW_\tau}$ , is well defined. Moreover, by compactness of  $W_\tau$ , the measures  $\mathcal{E}_\tau$  are purely atomic, have bounded support, and are thus characterized by a countable set of bounded, real-valued eigenfrequencies with a corresponding orthonormal eigenbasis of  $\mathcal{H}_\tau$ . The skew-adjointness of  $W_\tau$  and the generator  $V$  also enables the use of spectral approximation techniques based on strong convergence in a core of  $V$  [51], which are special to skew- or self-adjoint operators and would not be available in a direct approximation of the unitary Koopman group.
- (ii) For systems that do possess nontrivial Koopman eigenfunctions, there exists a subset of the eigenfunctions of  $W_\tau$  converging to them as  $\tau \rightarrow 0$ . These eigenfunctions can be identified a posteriori by monitoring the growth of a Dirichlet energy functional as a function of  $\tau$ . Crucially, however, the eigenfunctions of  $W_\tau$  provide a basis for the whole of  $L^2(\mu)$ , including the continuous spectrum subspace  $H_c$ , that evolves under the dynamics as an approximate Koopman eigenfunction basis.
- (iii) The evaluation of  $e^{tW_\tau}$  in the eigenbasis of  $W_\tau$  leads to a stable and efficient scheme for forecasting observables, which can be initialized with pointwise initial data in  $M$ . This improves upon diffusion forecasting [12], as well as comparable prediction techniques operating directly on  $L^2(\mu)$ , which produce “weak” forecasts (i.e., expectation values of observables with respect to probability densities in  $L^2(\mu)$ ). In addition, being based on an approximation of the generator, the evolution of observables under  $e^{tW_\tau}$  is of a fundamentally generative nature, in contrast with direct approximations of the action of the Koopman group on fixed target observables [52] which would typically be of an interpolatory nature.
- (iv) Our framework is well-suited for data-driven approximation using techniques from statistics and machine learning [30,39,53]. In particular, the theory of interpolation and out-of-sample extension in RKHS allows for consistent and stable approximation of quantities of interest (e.g., the eigenfunctions of  $W_\tau$  and the action of  $e^{tW_\tau}$  on a prediction observable), based on data acquired on a finite trajectory in the state space  $M$ .

In our main results, Theorems 1, 2 and Corollaries 3, 4, we prove the spectral convergence of  $W_\tau$  to  $V$  in an appropriate sense by defining auxiliary compact operators acting on  $L^2(\mu)$ . The relationships between these operators are depicted in Fig. 1 in the form of a flow chart. In Theorem 21, we give a data-driven analog of our main results, indicating how to construct finite-rank operators from finite datasets without prior knowledge of the underlying system and/or state space, and how spectral convergence still holds in an appropriate sense.

### 1.5. Plan of the paper

In Section 2, we make our assumptions on the underlying system precise, and state our main results. This is followed by results on compactification of operators in RKHS, Theorems 5–10, in Section 3, which will be useful for the proofs of the main results. Before proving our main results, we also review some concepts from ergodic theory and functional analysis in Section 4. Then, in Sections 5 and 6, we prove Theorems 5–8 and



**Fig. 1.** Outline of the main results and relationships between operators employed in the paper. Our focus is on approximating, in a spectral sense, the group of unitary Koopman operators  $U^t$  on the  $L^2(\mu)$  space of observables of an ergodic dynamical system with an invariant measure  $\mu$ , and establishing associated data-driven schemes. We approach this problem by approximating the Koopman generator  $V$  and its associated spectral measure given in (1). In particular, we construct a family of compact, skew-adjoint operators  $\tilde{V}_\tau$  which have *strong resolvent convergence* to  $V$  as  $\tau \rightarrow 0$ , resulting in a spectral approximation of  $V$  in the sense of convergence of the corresponding spectral measures. Simultaneously, the exponentiation of  $e^{t\tilde{V}_\tau}$  converges strongly to the Koopman operator  $e^{tV}$ . These results provide means of (i) extraction of coherent observables associated with the approximate point spectrum of  $V$  (Corollary 3); and (ii) prediction of observables under the dynamics (Corollary 4). A key element of the approach is a family of compact, skew-adjoint operators  $W_\tau$  on RKHSs  $\mathcal{H}_\tau$  of appropriate regularity (Theorem 1). These operators are constructed in Theorem 6, and are related to  $\tilde{V}_\tau$  by a canonical unitary transformation  $\mathcal{U}_\tau$  between  $L^2(\mu)$  and  $\mathcal{H}_\tau$ . The main results on the spectral convergence of  $\tilde{V}_\tau$  to  $V$  and strong convergence of the associated spectral measures are stated in Theorem 2. To connect  $W_\tau$  to  $V$ , we rely on two regularizations,  $A_\tau$  (Theorem 5) and  $B_\tau$  (Theorem 7), of  $V$ . Theorems 8 and 9 establish commutation and spectral relationships between  $\tilde{V}_\tau$ ,  $W_\tau$ ,  $A_\tau$ , and  $B_\tau$ , followed by spectral convergence results for these operators in Theorem 10. Theorem 21 and Corollary 22 describe data-driven analogs of the main approximation results utilizing time series data. These approaches are illustrated with numerical examples in Section 8.

9, 10, respectively, while Section 7 contains the proof of our main results. In Section 8, we describe a data-driven method to approximate the compactified generator  $W_\tau$ , and establish its convergence (Theorem 21). In Section 9, we present illustrative numerical examples of our framework applied to dynamical systems with both purely atomic and continuous Koopman spectra, namely a quasiperiodic rotation on a 2-torus, and the Rössler and L63 systems. We state our primary conclusions in Section 9. The paper also includes an appendix on variable-bandwidth Gaussian kernels [43] (Appendix A). Pseudocode is included in Appendix B.

## 2. Main results

All of our main results will use the following standing assumptions and notations.

**Assumption 1.**  $\Phi^t : \mathcal{M} \rightarrow \mathcal{M}$ ,  $t \in \mathbb{R}$ , is a continuous-time, continuous flow on a metric space  $\mathcal{M}$ . There exists a forward-invariant,  $m$ -dimensional,  $C^r$ , compact, connected manifold  $M \subseteq \mathcal{M}$ , such that the restricted flow map  $\Phi^t|_M$  is also  $C^r$ .  $X \subseteq M$  is a compact invariant set, supporting an ergodic, invariant Borel probability measure  $\mu$ .

This assumption is met by many dynamical systems encountered in applications, including ergodic flows on compact manifolds with regular invariant measures (in which case  $\mathcal{M} = M = X$ ), certain dissipative ordinary differential equations on noncompact manifolds (e.g., the L63 system [31], where  $\mathcal{M} = \mathbb{R}^3$ ,  $M$  is an appropriate absorbing ball [54], and  $X$  a fractal attractor [55]), and certain dissipative partial equations with inertial manifolds [56] (where  $\mathcal{M}$  is an infinite-dimensional function space).

In what follows, we seek to compactify the generator  $V$ , whose action is similar to that of a differentiation operator along the trajectories of the flow. Intuitively, one way of achieving this is to compose  $V$  with appropriate smoothing operators. To that end, we will employ kernel integral operators associated with RKHSs.

*Kernels and their associated integral operators.* In the context of interest here, a kernel will be a continuous function  $k : M \times M \rightarrow \mathbb{C}$ , which can be thought of as a measure of similarity or correlation between pairs of points in  $M$ . Associated with every kernel  $k$  and every finite, compactly supported Borel measure  $\nu$  (e.g., the invariant measure  $\mu$ ) is an integral operator  $K : L^2(\nu) \rightarrow C^0(M)$ , acting on  $f \in L^2(\nu)$  as

$$Kf := \int_M k(\cdot, y) f(y) d\nu(y). \quad (7)$$

If, in addition,  $k$  lies in  $C^r(M \times M)$ , then  $K$  imparts this smoothness to  $Kf$ , i.e.,  $Kf \in C^r(M)$ . Note that the compactness of  $\text{supp}(\nu)$  is important for this conclusion to hold. The kernel  $k$  is said to be Hermitian if  $k(x, y) = k^*(y, x)$  for all  $x, y \in M$ . It is called positive-definite if for every sequence of distinct points  $x_1, \dots, x_n \in M$  the  $n \times n$  kernel matrix  $\mathbf{K} = [k(x_i, x_j)]$  is non-negative, and *strictly* positive-definite if  $\mathbf{K}$  is strictly positive-definite. Clearly, every real, Hermitian kernel is symmetric, i.e.,  $k(x, y) = k(y, x)$  for all  $x, y \in M$ .

Aside from inducing an operator mapping into  $C^r(M)$ , a kernel  $k$  also induces an operator  $G = \iota K$  on  $L^2(\nu)$ , where  $\iota : C^0(M) \rightarrow L^2(\nu)$  is the canonical  $L^2$  inclusion map on continuous functions. The operator  $G$  is Hilbert-Schmidt, and thus compact and of finite trace. In particular, its Hilbert-Schmidt norm and trace are given by

$$\|G\|_{\text{HS}} := \sqrt{\text{tr}(G^*G)} = \|k\|_{L^2(\nu \times \nu)}, \quad \text{tr } G = \int_M k(x, x) d\nu(x), \quad (8)$$

respectively. Moreover, if  $k$  is Hermitian,  $G$  is self-adjoint, and there exists an orthonormal basis of  $L^2(\mu)$  consisting of its eigenfunctions. Let  $X_\nu$  denote the support of  $\nu$ . A kernel  $k$  will be called  $L^2(\nu)$ -positive and  $L^2(\nu)$ -strictly-positive if  $G \geq 0$  and  $G > 0$ , respectively; in those cases,  $G$  is also of trace class. Note that if  $k$  is (strictly) positive-definite on  $X_\nu \times X_\nu$ , then it is  $L^2(\nu)$ - (strictly-) positive. Moreover,  $k$  will be called a  $L^2(\nu)$ -Markov kernel if the associated integral operator  $G : L^2(\nu) \rightarrow L^2(\nu)$  is Markov, i.e., (i)  $Gf \geq 0$  if  $f \geq 0$ ; (ii)  $\int_M Gf d\nu = \int_M f d\nu$ , for all  $f \in L^2(\nu)$ ; and (iii)  $Gf = f$  if  $f$  is constant. The Markov kernel  $k$  will be said to be ergodic if  $Gf = f$  iff  $f$  is constant. A sufficient condition for  $k$  to be Markov is that  $k \geq 0$  on  $X_\nu \times X_\nu$ , and  $\int_M k(x, \cdot) d\nu = 1$  for  $\nu$ -a.e.  $x \in M$ . If  $k > 0$  on  $X_\nu \times X_\nu$ , then  $k$  is ergodic.

*Reproducing kernel Hilbert spaces.* An RKHS on  $M$  is a Hilbert space  $\mathcal{H}$  of complex-valued functions on  $M$  with the special property that for every  $x \in M$ , the point-evaluation map  $\delta_x : \mathcal{H} \rightarrow \mathbb{C}$ ,  $\delta_x f = f(x)$ , is a bounded, and thus continuous, linear functional. By the Riesz representation theorem, every RKHS has a unique reproducing kernel, i.e., a kernel  $k : M \times M \rightarrow \mathbb{C}$  such that for every  $x \in M$  the kernel section  $k(x, \cdot)$  lies in  $\mathcal{H}$ , and for every  $f \in \mathcal{H}$ ,

$$f(x) = \delta_x f = \langle k(x, \cdot), f \rangle_{\mathcal{H}},$$

where  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  is the inner product of  $\mathcal{H}$ , assumed conjugate-linear in the first argument. It then follows that  $k$  is Hermitian. Conversely, according to the Moore-Aronszajn theorem [57], given a Hermitian, positive-definite kernel  $k : M \times M \rightarrow \mathbb{C}$ , there exists a unique RKHS  $\mathcal{H}$  for which  $k$  is the reproducing kernel. Moreover, the range of  $K$  from (7) lies in  $\mathcal{H}$ , so we can view  $K$  as an operator  $K : L^2(\nu) \rightarrow \mathcal{H}$  between Hilbert spaces. With this definition,  $K$  is compact, and the adjoint operator  $K^* : \mathcal{H} \rightarrow L^2(\nu)$  maps  $f \in \mathcal{H}$  into its  $L^2(\nu)$  equivalence class, i.e.,  $K^* = \iota|_{\mathcal{H}}$  and  $G = K^*K$ . For any compact subset  $S \subseteq M$ , one can similarly define  $\mathcal{H}(S)$  to be the RKHS induced on  $S$  by the kernel  $k|_{S \times S}$ . In fact, upon restriction to the support  $X_\nu$ , the range of  $K$  is a dense subspace of  $\mathcal{H}(X_\nu)$ . This implies that every function in  $\mathcal{H}(X_\nu)$  has a unique extension to a function in  $\mathcal{H}$  lying in the closed subspace  $\mathcal{K} := \overline{\text{ran } K} \subseteq \mathcal{H}$ .

*Nyström extension.* Let  $\mathcal{H}$  be an RKHS on  $M$  with reproducing kernel  $k$ . Then, the Nyström extension operator  $\mathcal{N} : D(\mathcal{N}) \rightarrow \mathcal{H}$  acts on a subspace  $D(\mathcal{N})$  of  $L^2(\nu)$ , mapping each element  $f$  in its domain to a function  $\mathcal{N}f \in \mathcal{H}$ , such that  $\mathcal{N}f$  lies in the same  $L^2(\nu)$  equivalence class as  $f$ . In other words,  $\mathcal{N}f(x) = f(x)$  for  $\nu$ -a.e.  $x \in M$ , and  $K^*\mathcal{N}$  is the identity on  $D(\mathcal{N})$ . It can also be shown that  $D(\mathcal{N}) = \text{ran } K^*$ ,  $\text{ran } \mathcal{N} = \mathcal{K}$ , and  $\mathcal{N}K^*$  is the identity on  $\mathcal{K}$ . Moreover, if  $k$  is  $L^2(\nu)$ -strictly-positive, then  $D(\mathcal{N})$  is a dense subspace of  $L^2(\nu)$ . In fact,  $D(\mathcal{N})$  can be endowed with the structure of a Hilbert space, equipped with the inner product  $\langle f, g \rangle_{\mathcal{N}} = \langle \mathcal{N}f, \mathcal{N}g \rangle_{\mathcal{H}}$ . If  $k$  is  $L^2(\nu)$ -strictly-positive and Markov ergodic, this space behaves in many ways analogously to a Sobolev space on a compact Riemannian manifold. In particular, equipped with this inner product,  $D(\mathcal{N})$  embeds compactly into  $L^2(\nu)$ , and  $\|f\|_{\mathcal{N}} \geq \|f\|_{L^2(\nu)}$  with equality iff  $f$  is constant. Moreover, the  $\|\cdot\|_{\mathcal{N}}$  norm induces a *Dirichlet energy functional*  $\mathcal{D} : D(\mathcal{N}) \rightarrow \mathbb{R}$ ,

$$\mathcal{D}(f) = \frac{\|f\|_{\mathcal{N}}^2}{\|f\|_{L^2(\nu)}^2} - 1, \quad \forall f \in D(\mathcal{N}) \setminus \{0\}, \quad \text{and} \quad \mathcal{D}(0) = 0, \quad (9)$$

where  $\mathcal{D}(f)$  is non-negative and vanishes iff  $f$  is constant by  $L^2(\nu)$ -Markovianity and ergodicity of  $k$ . Intuitively,  $\mathcal{D}$  can be interpreted as a measure of “roughness” of functions in  $D(\mathcal{N})$ , which vanishes for constant functions, and is large for functions that project strongly to the eigenfunctions of  $G$  with small corresponding eigenvalues. We will give a precise constructive definition of  $\mathcal{N}$ , and discuss its properties, in Section 4.

The following assumption specifies our nominal requirements on kernels pertaining to regularity and existence of an associated RKHS.

**Assumption 2.**  $p : M \times M \rightarrow \mathbb{R}$  is a  $C^r$ , symmetric, positive-definite kernel, and  $\nu$  a Borel probability measure with compact support  $X_{\nu} \subseteq M$ . Moreover,  $p$  is  $L^2(\nu)$ -strictly-positive and Markov ergodic.

We will later describe how kernels satisfying Assumption 2 can easily be constructed from symmetric, positive-definite, positive-valued  $C^r$  kernels using the bistochastic kernel normalization technique proposed in [58]. It should be noted that many of our results will require  $r = 1$  differentiability class in Assumptions 1 and 2, but in some cases that requirement can be relaxed to  $r = 0$ .

*One-parameter kernel families.* Let  $P : L^2(\nu) \rightarrow \mathcal{H}$  be the integral operator associated with a kernel  $p$  satisfying Assumption 2, taking values in the corresponding RKHS  $\mathcal{H}$ . The associated operator  $G = P^*P$  on  $L^2(\nu)$  has positive eigenvalues, which can be ordered as  $1 = \lambda_0 > \lambda_1 \geq \dots$ . Given a real, orthonormal basis  $\{\phi_0, \phi_1, \dots\}$  of  $L^2(\nu)$  consisting of corresponding eigenfunctions, the set  $\{\psi_0, \psi_1, \dots\}$  with  $\psi_j = \lambda_j^{-1/2} P\phi_j$  is an orthonormal basis of  $\overline{\text{ran } P} \subseteq \mathcal{H}$ , and the restrictions of these functions to  $X_{\nu}$  form an orthonormal basis of  $\mathcal{H}(X_{\nu})$ . Defining

$$\lambda_{\tau,j} := \exp(\tau(1 - \lambda_j^{-1})), \quad \psi_{\tau,j} := \sqrt{\lambda_{\tau,j}/\lambda_j} \psi_j, \quad p_{\tau}(x, y) := \sum_{j=0}^{\infty} \psi_{\tau,j}(x) \psi_{\tau,j}(y), \quad (10)$$

where  $\tau > 0$ , and  $x, y$  are arbitrary points in  $M$ , the following theorem establishes the existence of a one-parameter family of RKHSs, indexed by  $\tau$ , and an associated Markov semigroup on  $L^2(\nu)$ .

**Theorem 1 (Markov kernels).** *Let Assumption 2 hold. Then, for every  $\tau > 0$ , the series expansion for  $p_{\tau}(x, y)$  in (10) converges in  $C^r(M \times M)$  norm to a  $C^r$ , symmetric function. Moreover, the following hold:*

- (i) *For every  $\tau > 0$ ,  $p_{\tau}$  is a positive-definite kernel on  $M$ . In addition, it is  $L^2(\nu)$ -strictly-positive and Markov ergodic.*

- (ii) For every  $\tau > 0$ , the RKHS  $\mathcal{H}_\tau$  associated with  $p_\tau$  lies dense in  $L^2(\nu)$ , and for every  $0 < \tau_1 < \tau_2$ , the inclusions  $\mathcal{H}_{\tau_2} \subseteq \mathcal{H}_{\tau_1} \subseteq \mathcal{H}$  hold. Moreover,  $\{\psi_{\tau,0}, \psi_{\tau,1}, \dots\}$  is an orthonormal basis of  $\mathcal{H}_\tau$ .
- (iii) Define  $G_0 := \text{Id}_{L^2(\nu)}$  and  $G_\tau = P_\tau^* P_\tau$ , where  $P_\tau : L^2(\nu) \rightarrow \mathcal{H}_\tau$  is the integral operator associated with  $p_\tau$ . Then, the family  $\{G_\tau\}_{\tau \geq 0}$  forms a strongly continuous, self-adjoint Markov semigroup.

**Remark.** Theorem 1 is independent of the dynamical system in Assumption 1. It is a general RKHS result, allowing one to employ basis functions for the RKHS  $\mathcal{H}(X_\nu)$ , restricted on the support of  $\nu$ , to construct a family of RKHSs  $\mathcal{H}_\tau$  on the entire compact manifold  $M$ . In particular,  $\text{ran } P_\tau$  is a dense subspace of  $\mathcal{H}_\tau$ , while  $\text{ran } P$  is not necessarily dense in  $\mathcal{H}$ .

The semigroup structure of the family  $\{G_\tau\}_{\tau \geq 0}$  in Theorem 1(iii) implies, in particular, that for every  $\tau_1, \tau_2 \geq 0$ ,  $G_{\tau_1+\tau_2} = G_{\tau_1} G_{\tau_2}$ . Moreover, strong continuity is equivalent to pointwise convergence of  $G_\tau$  to the identity operator as  $\tau \rightarrow 0^+$ . These two properties, as well as the Markov ergodic property, will all be important in our compactification schemes for the Koopman generator, presented in Theorem 2 and Section 3 below. The measure  $\nu$  will now be set to the invariant measure  $\mu$ . In what follows,  $\mathcal{N}_\tau : D(\mathcal{N}_\tau) \rightarrow \mathcal{H}_\tau$  will be the Nyström operator associated with  $\mathcal{H}_\tau$ . We also let  $H_\infty = \bigcap_{\tau > 0} D(\mathcal{N}_\tau)$  be the dense subspace of  $L^2(\mu)$  whose elements have  $\mathcal{H}_\tau$  representatives for every  $\tau > 0$ . Note that  $H_\infty$  is dense since it contains all finite linear combinations of the  $\phi_j$ . Similarly, setting  $\mathcal{H}_\infty = \bigcap_{\tau > 0} \mathcal{H}_\tau$ , it follows that  $\mathcal{H}_\infty(X)$  is a dense subspace of  $\mathcal{H}(X)$ . In addition, we will be making use of the polar decomposition of  $P_\tau$ . The latter can be shown to take the form

$$P_\tau = \mathcal{U}_\tau G_\tau^{1/2}, \quad (11)$$

where  $\mathcal{U}_\tau : L^2(\mu) \rightarrow \mathcal{H}_\tau$  is the unitary operator such that  $\mathcal{U}_\tau \phi_j = \psi_{j,\tau}$  for all pairs  $(\phi_j, \psi_{j,\tau})$  from (10). Given a Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$  and a densely-defined skew-adjoint operator  $T$ ,  $Z(T)$  will denote the operator-valued function obtained through the Borel functional calculus as in Section 1. For every set  $\Omega \subset \mathbb{C}$ ,  $\partial\Omega$  will denote its boundary.

**Theorem 2 (Main theorem).** *Under Assumptions 1, 2 with  $r = 1$ , and the definitions in (10), the following hold for every  $\tau > 0$ :*

- (i) *The operator  $W_\tau := P_\tau V P_\tau^* : \mathcal{H}_\tau \rightarrow \mathcal{H}_\tau$  is a well-defined, skew-adjoint, real integral operator of trace class.*
- (ii) *The operator  $G_\tau V : D(V) \rightarrow L^2(\mu)$  extends to a trace class integral operator  $B_\tau : L^2(\mu) \rightarrow L^2(\mu)$ . Moreover, the restriction of  $B_\tau$  to the dense subspace  $D(\mathcal{N}_\tau) \subseteq D(V)$  coincides with the operator  $P_\tau^* W_\tau \mathcal{N}_\tau$ .*
- (iii) *The operators  $B_\tau$  and  $W_\tau$  have the same spectra, including multiplicities of eigenvalues. Moreover, there exists a unique, purely atomic PVM  $\mathcal{E}_\tau : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(\mathcal{H}_\tau)$ , such that  $W_\tau = \int_{\mathbb{R}} i\omega d\mathcal{E}_\tau(\omega)$ .*

In addition, as  $\tau \rightarrow 0^+$ :

- (iv) *For every bounded, Borel-measurable set  $\Omega \subset \mathbb{R}$  such that  $E(\partial\Omega) = 0$ ,  $P_\tau^* \mathcal{E}_\tau(\Omega) \mathcal{N}_\tau$  and  $\mathcal{U}_\tau^* \mathcal{E}_\tau(\Omega) \mathcal{U}_\tau$  converge to  $E(\Omega)$ , in the strong operator topologies of  $H_\infty$  and  $L^2(\mu)$ , respectively.*
- (v) *For every bounded continuous function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ ,  $P_\tau^* Z(W_\tau) \mathcal{N}_\tau$  and  $\mathcal{U}_\tau^* Z(W_\tau) \mathcal{U}_\tau$  converge to  $Z(V)$ , in the strong operator topologies of  $H_\infty$  and  $L^2(\mu)$ , respectively.*
- (vi) *For every holomorphic function  $Z : D(Z) \rightarrow \mathbb{C}$ , with  $i\mathbb{R} \subset D(Z) \subseteq \mathbb{C}$  and  $Z|_{i\mathbb{R}}$  bounded,  $Z(B_\tau)$  converges strongly to  $Z(V)$  on  $L^2(\mu)$ .*
- (vii) *For every element  $i\omega$  of the spectrum of the generator  $V$ , there exists a continuous curve  $\tau \rightarrow \omega_\tau$  such that  $i\omega_\tau$  is an eigenvalue of  $B_\tau$  and  $W_\tau$ , and  $\lim_{\tau \rightarrow 0^+} \omega_\tau = \omega$ .*

The skew-adjoint operator  $W_\tau$  from Theorem 2 can be viewed as a compact approximation to the generator  $V$ . This approximation has a number of advantages for both coherent extraction and prediction. First, although  $V$  is unbounded and could exhibit complex spectral behavior (see Section 1),  $W_\tau$  has a complete orthonormal basis of eigenfunctions, which are  $C^1$  functions lying in  $\mathcal{H}_\tau$ . This suggests that the eigenfunctions of  $W_\tau$  are good candidates for coherent observables of high regularity, which are well defined for systems with general spectral characteristics. Moreover, the discrete spectra of compact, skew-adjoint operators can be used to construct and approximate to any degree of accuracy the Borel functional calculi of these operators, and in particular perform forecasting through exponentiation of  $W_\tau$ . The eigenvalues and eigenfunctions of the smoothing operators  $P_\tau$  employed in the construction of  $W_\tau$  can also be easily derived from those of  $P$  with little computational overhead. In Corollaries 3 and 4 below, we make precise the utility of  $W_\tau$  for the purposes of coherent pattern extraction and forecasting, respectively. See Fig. 3 for an illustration of the dependence of the spectrum of  $W_\tau$  on  $\tau$  for dynamical systems with point and continuous Koopman spectra.

*Approximate point spectrum.* Given  $t \in \mathbb{R}$  and  $\epsilon > 0$ , a complex number  $\gamma$  is said to lie in the  $\epsilon$ -approximate point spectrum of  $U^t$  if there exists a nonzero  $f \in L^2(\mu)$  such that

$$\|U^t f - \gamma f\|_{L^2(\mu)} < \epsilon \|f\|_{L^2(\mu)}. \quad (12)$$

Such observables  $f$  (which include Koopman eigenfunctions as special cases), satisfying (12) for small  $\epsilon$  and  $t$  lying in a given time interval, exhibit a form of dynamical coherence, as they evolve approximately as Koopman eigenfunctions over that time interval. We will refer to  $(\gamma, f)$  satisfying (12) as an  $\epsilon$ -approximate eigenpair of  $U^t$ . A discussion on how the  $\epsilon$ -approximate point spectrum varies with  $\epsilon$ , and its relation to the spectrum, in the context of a general, closed, unbounded operator, can be found in Section 4. The following corollary of Theorem 2 establishes that the eigenvalues of  $W_\tau$  corresponding to eigenfunctions that satisfy certain Dirichlet energy criteria, can be used to identify points in the  $\epsilon$ -approximate point spectrum of the Koopman operator at any  $\epsilon > 0$ . In what follows,  $\mathcal{D} : D(\mathcal{N}) \rightarrow \mathbb{R}$  will denote the Dirichlet energy from (9), induced on  $L^2(\mu)$  by the kernel  $p$  in Assumption 2. We also introduce the function  $R : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ , defined as

$$R(\epsilon, \tau) := \sup\{T > 0 : \|(U^t - e^{tB_\tau})P^*\| < \epsilon, \forall t \in [-T, T]\}.$$

Here, the norm of  $(U^t - e^{tB_\tau})P^*$  is taken as an operator from  $\mathcal{H}$  into  $L^2(\mu)$ . The function  $R(\epsilon, \tau)$  identifies a time interval  $(-R(\tau, \epsilon), R(\tau, \epsilon)) \ni t$  over which the operator families  $U^t$  and  $e^{tB_\tau}$  are  $\epsilon$ -close. We will later show in Proposition 20 that for every  $\epsilon > 0$ ,  $R(\epsilon, \tau)$  diverges as  $\tau \rightarrow 0^+$ .

**Corollary 3 (Coherent observables).** *Let  $(i\omega_\tau, \zeta_\tau)$  be an eigenpair of  $W_\tau$ . Then,  $(e^{i\omega_\tau t}, z_\tau)$ , with  $z_\tau = P_\tau^* \zeta_\tau$ , is an  $\epsilon$ -approximate eigenpair of  $U^t$  for all  $t \in (-T(\epsilon, \tau), T(\epsilon, \tau))$ , where*

$$T(\epsilon, \tau) = R(\epsilon, \tau) / \sqrt{\mathcal{D}(z_\tau) + 1}.$$

*In addition, the following hold:*

- (i) *If  $\lim_{\tau \rightarrow 0^+} \omega_\tau =: \omega$  exists, and  $T(\epsilon, \tau)$  diverges as  $\tau \rightarrow 0^+$  for every  $\epsilon > 0$ , then  $i\omega$  is an element of the spectrum of  $V$ .*
- (ii) *If  $\lim_{\tau \rightarrow 0^+} \omega_\tau =: \omega$  exists, and  $\mathcal{D}(z_\tau)$  is bounded as  $\tau \rightarrow 0^+$ , then  $i\omega$  is an eigenvalue of  $V$ . Moreover, the sequence  $z_\tau$  converges to the eigenspace of  $V$  corresponding to  $i\omega$ .*

**Remark.** An important consideration in spectral approximation techniques is to identify and/or control the occurrence of *spectral pollution* [59], i.e., eigenvalues  $i\omega_\tau$  of the approximating operators  $W_\tau$  converging to points which do not lie in the spectrum of  $V$ . Corollary 3 establishes that the regularity of the corresponding eigenfunctions  $\zeta_\tau$ , as measured through the Dirichlet energy functional associated with the RKHS  $\mathcal{H}$ , provides a useful a posteriori criterion for identifying spectral pollution.

Turning now to forecasting, let  $\{i\omega_{\tau,0}, i\omega_{\tau,1}, \dots\}$  be the set of eigenvalues of  $W_\tau$ . Note that since  $W_\tau$  is a compact, skew-adjoint real operator, the  $i\omega_{j,\tau}$  occur in complex-conjugate pairs, and 0 is the only accumulation point of the sequence  $\omega_{\tau,0}, \omega_{\tau,1}, \dots$ . Let also  $\{\zeta_{\tau,0}, \zeta_{\tau,1}, \dots\}$  be an orthonormal basis of  $\mathcal{H}_\tau$  consisting of corresponding eigenfunctions. The following is a corollary of Theorem 2, which shows that the evolution of an observable in  $L^2(\mu)$  under  $U^t$  can be evaluated to any degree of accuracy by evolution of an approximating observable in  $\mathcal{H}_\infty$  under  $e^{tW_\tau}$ .

**Corollary 4 (Prediction).** *For every  $\tau > 0$ ,  $W_\tau$  generates a norm-continuous group of unitary operators  $e^{tW_\tau} : \mathcal{H}_\tau \rightarrow \mathcal{H}_\tau$ ,  $t \in \mathbb{R}$ . Moreover, for any observable  $f \in L^2(\mu)$ , error bound  $\epsilon > 0$ , and compact set  $\mathcal{T} \subset \mathbb{R}$ , there exists  $f_\epsilon \in \mathcal{H}_\infty$  (independent of  $\mathcal{T}$ ) and  $\tau_0 > 0$ , such that for every  $\tau \in (0, \tau_0)$  and  $t \in \mathcal{T}$ ,*

$$\|U^t f - P_\tau^* e^{tW_\tau} f_\epsilon\|_{L^2(\mu)} < \epsilon, \quad e^{tW_\tau} f_\epsilon = \sum_{j=0}^{\infty} e^{i\omega_{\tau,j} t} \langle \zeta_{\tau,j}, f_\epsilon \rangle_{\mathcal{H}_\tau} \zeta_{\tau,j}.$$

**Remark.** The function  $e^{tW_\tau} f_\epsilon$  lies in  $\mathcal{H}_\tau$ , and is therefore a continuous function which we employ to predict the evolution of the observable  $f$  under  $U^t$ . Corollary 4 suggests that to obtain this function, we first regularize  $f$  by approximating it by a function  $f_\epsilon \in \mathcal{H}_\infty$ , and then invoke the functional calculus for the compact operator  $W_\tau$  to evolve  $f_\epsilon$  as an approximation of  $U^t f$ . Note that analogous error bounds to that in Corollary 4 can be obtained for operator-valued functions  $Z(V)$  of the generator other than the exponential functions,  $Z(V) = e^{tV} = U^t$ . A constructive procedure for obtaining the forecast function in a data-driven setting will be described in Section 8.

The flow chart in Fig. 1 summarizes the relationships between the operator families in Theorem 2 and Corollaries 3, 4 and their associated modes of convergence.

### 3. Compactification schemes for the generator

In this section, we lay out various schemes for obtaining compact operators by composing the generator  $V$  with operators derived from kernels. These schemes are of independent interest, as they are applicable, with appropriate modifications, to more general classes of unbounded, skew- or self-adjoint operators obtained by extension of differentiation operators. In some cases, the following weaker analog of Assumption 2 will be sufficient.

**Assumption 3.**  *$k : M \times M \rightarrow \mathbb{R}$  is a  $C^1$ , symmetric positive-definite kernel.*

Given the RKHS  $\mathcal{H} \subset C^1(M)$  associated with  $k$  from Assumption 3, and the corresponding integral operators  $K : L^2(\mu) \rightarrow \mathcal{H}$ ,  $G = K^* K : L^2(\mu) \rightarrow L^2(\mu)$ , and closed subspace  $\mathcal{K} = \overline{\text{ran } K} \subseteq \mathcal{H}$ , we begin by formally introducing the operators  $A : L^2(\mu) \rightarrow L^2(\mu)$  and  $W : \mathcal{H} \rightarrow \mathcal{H}$ , where

$$A := VG = VK^*K, \quad W := KVK^*. \quad (13)$$

Note that it is not necessarily the case that these operators are well defined, for the ranges of  $G$  and  $K^*$  may lie outside of the domain of  $V$ . Nevertheless, as the following two theorems establish,  $A$  and  $W$  are well-defined, and in fact compact, operators.

**Theorem 5** (*Pre-smoothing*). *Let Assumptions 1 and 3 hold, and define  $k' : M \times M \rightarrow \mathbb{R}$  as the  $C^0$  kernel with  $k'(x, y) := \lim_{t \rightarrow 0} (k(\Phi^t(x), y) - k(x, y))/t$ . Then:*

- (i) *The range of  $G$  lies in the domain of  $V$ .*
- (ii) *The operator  $A$  from (13) is a well-defined, Hilbert-Schmidt integral operator on  $L^2(\mu)$  with kernel  $k'$ , and thus bounded in operator norm by*

$$\|A\| \leq \|A\|_{HS} = \|k'\|_{L^2(\mu \times \mu)} \leq \|k'\|_{C^0(X \times X)}.$$

- (iii)  *$A$  is equal to the negative adjoint,  $-(GV)^*$ , of the densely defined operator  $GV : D(V) \rightarrow L^2(\mu)$ .*

**Remark.** As stated in Section 1,  $V$  is an unbounded operator, whose domain is a strict subspace of  $L^2(\mu)$ . Theorem 5 thus shows that if we regularize this operator by first applying the smoothing operator  $G$ , then not only is  $A$  bounded, it is also Hilbert-Schmidt, and thus compact. In essence, this property follows from the  $C^1$  regularity of the kernel.

Arguably, the regularization scheme leading to  $A$ , which involves first smoothing by application of  $G$ , followed by application of  $V$ , is among the simplest and most intuitive ways of regularizing  $V$ . However, the resulting operator  $A$  will generally not be skew-symmetric; in fact, apart from special cases,  $A$  will be non-normal. Theorem 6 below provides an alternative regularization approach for  $V$ , leading to a Hilbert-Schmidt operator on  $\mathcal{H}$  which is additionally skew-adjoint. Working with this operator also takes advantage of the RKHS structure, allowing pointwise function evaluation by bounded linear functionals.

**Theorem 6** (*Compactification in RKHS*). *Let Assumptions 1 and 3 hold, and define  $\tilde{k}' : M \times M \rightarrow \mathbb{R}$  as the  $C^0$  kernel with  $\tilde{k}'(x, y) = -k'(y, x)$ . Then:*

- (i) *The range of  $K^*$  lies in the domain of  $V$ , and  $VK^* : \mathcal{H} \rightarrow L^2(\mu)$  is a bounded operator.*
- (ii) *The operator  $W$  from (13) is a well-defined, Hilbert-Schmidt, skew-adjoint, real operator on  $\mathcal{H}$ , with  $\text{ran } W \subseteq \mathcal{K}$ , satisfying*

$$Wf = \int_M \tilde{k}'(\cdot, y)f(y) d\mu(y).$$

**Remark.** Because  $W$  is skew-adjoint, real, and compact, it has the following properties, which we will later use.

- (i) Its nonzero eigenvalues are purely imaginary, occur in complex-conjugate pairs, and accumulate only at zero. Moreover, there exists an orthonormal basis of  $\mathcal{H}$  consisting of corresponding eigenfunctions.
- (ii) It generates a norm-continuous, one-parameter group of unitary operators  $e^{tW} : \mathcal{H} \rightarrow \mathcal{H}$ ,  $t \in \mathbb{R}$ .

In the next theorem, we connect the operators  $A$  and  $W$  through the adjoint of  $A$ .

**Theorem 7** (*Post-smoothing*). *Let Assumptions 1 and 3 hold. Then, the adjoint of  $-A$  from (13) is a Hilbert-Schmidt integral operator  $B : L^2(\mu) \rightarrow L^2(\mu)$  with kernel  $\tilde{k}'$ . In addition:*

- (i) *The densely-defined operator  $GV : D(V) \rightarrow L^2(\mu)$  is bounded, and  $B$  is equal to its closure,  $\overline{GV} := (GV)^{**}$ . Moreover,  $B$  is a closed extension of  $KWN : D(\mathcal{N}) \rightarrow L^2(\mu)$ , and if the kernel  $k$  is  $L^2(\mu)$ -strictly-positive, i.e.,  $D(\mathcal{N})$  is a dense subspace of  $L^2(\mu)$ , that extension is unique.*

(ii)  $B$  generates a norm-continuous, 1-parameter group of operators  $e^{tB} : L^2(\mu) \rightarrow L^2(\mu)$ ,  $t \in \mathbb{R}$ , satisfying

$$K^* e^{tW} = e^{tB} K^*, \quad K^* e^{tW} \mathcal{N} = e^{tB}|_{D(\mathcal{N})}, \quad \forall t \in \mathbb{R}.$$

**Remark.** Because  $V$  is an unbounded operator, defined on a dense subset  $D(V) \subset L^2(\mu)$ , the domain of  $GV$  is also restricted to  $D(V)$ . It is therefore a non-intuitive result that a regularization of  $V$  after an application of  $G$  could still result in a bounded operator that can be extended to the entire space  $L^2(\mu)$ .

Theorem 7(i) shows that, on the subspace  $D(\mathcal{N}) \subset L^2(\mu)$ ,  $B$  acts by first performing Nyström extension, then acting by  $W$ , then mapping back to  $L^2(\mu)$  by inclusion via  $K^*$ . In other words,  $B$  is a natural analog of  $W$  acting on  $L^2(\mu)$ , though note that, unlike  $W$ ,  $B$  is generally not skew-adjoint. To summarize, on the basis of Theorems 5–7, we have obtained the following sequence of operator extensions:

$$KWN \subseteq GV \subset B = \overline{GV} = (GV)^{**}.$$

As our final compactification of  $V$ , we will construct a skew-adjoint operator  $\tilde{V}$  on  $L^2(\mu)$  by conjugation by a compact operator. In particular, since  $G$  is positive, it has a square root  $G^{1/2} : L^2(\mu) \rightarrow L^2(\mu)$ , which is the unique positive operator satisfying  $G^{1/2}G^{1/2} = G$ . Note that by compactness of  $G$ ,  $G^{1/2}$  is compact, and its action on functions can be conveniently evaluated in an eigenbasis of  $G$ . Moreover, it can be verified that  $\text{ran } G^{1/2} = \text{ran } K^*$ . In fact, the operators  $K$  and  $G^{1/2}$  are related to  $K$  via the polar decomposition,  $K = \mathcal{U}G^{1/2}$ , where  $\mathcal{U} : L^2(\mu) \rightarrow \mathcal{H}$  is a (uniquely defined) partial isometry with  $\text{ran } \mathcal{U} = \mathcal{K}$ , analogous to  $\mathcal{U}_\tau$  in (11). Note that  $\mathcal{K}$  is an invariant subspace of  $W$ . Moreover, if  $K$  has dense range, then  $\mathcal{U}$  becomes unitary. Using these definitions, we will show in Theorem 8 below that the operator  $G^{1/2}VG^{1/2}$ , defined on the subspace  $\{f \in L^2(\mu) : G^{1/2}f \in D(V)\}$ , actually extends to a well-defined compact operator.

**Theorem 8 (Skew-adjoint compactification).** *Let Assumptions 1 and 3 hold with  $r = 1$ . Then,  $G^{1/2}VG^{1/2}$  is a densely defined, bounded operator with a unique skew-adjoint extension to a Hilbert-Schmidt, real operator  $\tilde{V} : L^2(\mu) \rightarrow L^2(\mu)$ . Moreover,  $\tilde{V}$  is related to the operator  $W$  from Theorem 6 via conjugation by the partial isometry  $\mathcal{U}$ , i.e.,  $\tilde{V} = \mathcal{U}^*W\mathcal{U}$ . In particular, if the kernel  $k$  is  $L^2(\mu)$ -strictly-positive,  $\tilde{V}$  and  $W|_{\mathcal{K}}$  are unitarily equivalent.*

This completes the statement of our compactification schemes for  $V$ . Since these schemes are all carried out using the same kernel  $k$ , one might expect that the spectral properties of the compact operators  $A$ ,  $B$ ,  $\tilde{V}$ , and  $W$ , exhibit non-trivial relationships. These relationships will be made precise in Theorems 9 and 10 below. Hereafter,  $\sigma(T)$  and  $\sigma_p(T)$  will denote the spectrum and point spectrum (set of eigenvalues) of a linear operator  $T$ , respectively.

**Theorem 9 (Spectra of the compactified generators).** *Let Assumptions 1 and 3 hold with  $r = 1$ , and assume further that the kernel  $k$  is  $L^2(\mu)$ -strictly-positive. Let also  $\{\tilde{z}_0, \tilde{z}_1, \dots\}$  be an orthonormal basis of  $L^2(\mu)$ , consisting of eigenfunctions  $\tilde{z}_j$  of  $\tilde{V}$  corresponding to purely imaginary eigenvalues  $i\omega_j$ . Then:*

(i)  *$A$  and  $B$  have the same eigenvalues as  $\tilde{V}$ , including multiplicities. Moreover,  $\sigma_p(W) = \sigma_p(\tilde{V})$ , including multiplicities if  $K$  has dense range.*

*In addition, if the kernel  $k$  is  $L^2(\mu)$ -Markov ergodic:*

(ii) *0 is a simple eigenvalue of each of the operators  $A$ ,  $B$ ,  $\tilde{V}$ , and  $W|_{\mathcal{K}}$ , corresponding to constant eigenfunctions.*

- (iii) Every  $\tilde{z}_j$  lies in the domain of  $G^{-1/2}$ . Moreover, the set  $\{z'_0, z'_1, \dots\}$  with  $z'_j = G^{-1/2}\tilde{z}_j$  consists of eigenfunctions of  $A$ , corresponding to the eigenvalues  $\{i\omega_0, i\omega_1, \dots\}$ , and forms an unconditional Schauder basis of  $L^2(\mu)$ .
- (iv) The set  $\{z_0, z_1, \dots\}$  with  $z_j = G^{1/2}\tilde{z}_j$  is an unconditional Schauder basis of  $L^2(\mu)$ , consisting of eigenfunctions of  $B$  corresponding to the same eigenvalues,  $\{i\omega_0, i\omega_1, \dots\}$ . Moreover, it is the unique dual sequence to the  $\{z_j\}$ , satisfying  $\langle z'_j, z_l \rangle_\mu = \delta_{jl}$ .
- (v) The set  $\{\zeta_0, \zeta_1, \dots\}$  with  $\zeta_j = Kz'_j$  is an orthonormal basis of  $\mathcal{K}$  consisting of eigenfunctions  $\zeta_j$  of  $W$  corresponding to the eigenvalues  $i\omega_j$ .
- (vi) The operators  $A$ ,  $B$ ,  $\tilde{V}$ , and  $W$ , admit the representations

$$A = \sum_{j=0}^{\infty} i\omega_j \langle z_j, \cdot \rangle_{L^2(\mu)} z'_j, \quad B = \sum_{j=0}^{\infty} i\omega_j \langle z'_j, \cdot \rangle_{L^2(\mu)} z_j, \quad \tilde{V} = \sum_{j=0}^{\infty} i\omega_j \langle \tilde{z}_j, \cdot \rangle_{L^2(\mu)} \tilde{z}_j,$$

$$W = \sum_{j=0}^{\infty} i\omega_j \langle \zeta_j, \cdot \rangle_{\mathcal{H}} \zeta_j,$$

where the infinite sums for  $A$  and  $B$  converge strongly, and those for  $\tilde{V}$ , and  $W$  converge in Hilbert-Schmidt norm.

**Remark.** The Markovianity assumption on the kernel was important to conclude that  $A$ ,  $B$ ,  $\tilde{V}$ , and  $W|_{\mathcal{K}}$  have finite-dimensional nullspaces (which may not be the case for a general compact operator), allowing us to establish a one-to-one correspondence of the spectra of these operators, including eigenvalue multiplicities.

An immediate consequence of Theorem 9, in conjunction with Theorems 7 and 8, is that  $\tilde{V}$  and  $W$  are decomposable in terms of unique PVMs  $\tilde{E} : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(L^2(\mu))$  and  $\mathcal{E} : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(\mathcal{H})$ , such that  $\tilde{V} = \int_{\mathbb{R}} i\omega d\tilde{E}(\omega)$ ,  $W = \int_{\mathbb{R}} i\omega d\mathcal{E}(\omega)$ , and

$$\tilde{E}(\Omega) = \sum_{j: \omega_j \in \Omega} \langle \tilde{z}_j, \cdot \rangle_{L^2(\mu)} \tilde{z}_j, \quad \mathcal{E}(\Omega) = \sum_{j: \omega_j \in \Omega} \langle \zeta_j, \cdot \rangle_{\mathcal{H}} \zeta_j + 1_{\Omega}(0) \text{proj}_{\mathcal{K}^\perp}, \quad (14)$$

where  $1_{\Omega}$  is the indicator function on  $\Omega$ , and  $\text{proj}_{\mathcal{K}^\perp} : \mathcal{H} \rightarrow \mathcal{H}$  the orthogonal projection onto  $\mathcal{K}^\perp$ . Moreover,  $\tilde{E}$  and  $\mathcal{E}$  are related by conjugation by the partial isometry  $\mathcal{U} : L^2(\mu) \rightarrow \mathcal{H}$  from Theorem 8,

$$\tilde{E}(\Omega) = \mathcal{U}^* \mathcal{E}(\Omega) \mathcal{U}, \quad \forall \Omega \in \mathcal{B}(\mathbb{R}), \quad (15)$$

and if  $k$  is  $L^2(\mu)$ -strictly positive,  $\tilde{E}(\Omega)$  and  $\mathcal{E}(\Omega)|_{\mathcal{K}}$  are unitarily equivalent. The compactness of  $\tilde{V}$  and  $W$ , which is reflected in the fact that  $\tilde{E}$  and  $\mathcal{E}$  are purely atomic PVMs, allows for simple expressions for the Borel functional calculi of these operators. In particular, for every Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ , we have

$$Z(\tilde{V}) = \int_{\mathbb{R}} Z(i\omega) d\tilde{E}(\omega) = \sum_{j=0}^{\infty} Z(i\omega_j) \langle \tilde{z}_j, \cdot \rangle_{L^2(\mu)} \tilde{z}_j,$$

$$Z(W) = \int_{\mathbb{R}} Z(i\omega) d\mathcal{E}(\omega) = \sum_{j=0}^{\infty} Z(i\omega_j) \langle \zeta_j, \cdot \rangle_{\mathcal{H}} \zeta_j + Z(0) \text{proj}_{\mathcal{K}^\perp},$$

with all limits taken in the strong operator topology. Note that if  $K$  has dense range (as in Theorem 2),  $\mathcal{K}^\perp$  reduces to the zero subspace, and  $\text{proj}_{\mathcal{K}^\perp}$  vanishes in the above expressions.

In the case of  $A$  and  $B$ , the fact that these are, in general, non-normal operators precludes the construction of associated Borel functional calculi. Nevertheless, the compactness of these operators allows one to construct their holomorphic functional calculi in a straightforward manner. Specifically, given any holomorphic function  $Z : D(Z) \rightarrow \mathbb{C}$  on an open set  $D(Z) \subseteq \mathbb{C}$  containing  $\sigma(A) = \sigma(B)$ , we define

$$Z(A) = \oint_{\gamma} Z(z)(z - A)^{-1} dz, \quad Z(B) = \oint_{\gamma} Z(z)(z - B)^{-1} dz,$$

where  $\gamma$  is a Cauchy contour in  $D(Z)$  containing  $\sigma(A)$  in its interior. Now, because  $\tilde{V}G^{1/2} = G^{1/2}VG = G^{1/2}A$ , we have  $A = G^{-1/2}\tilde{V}G^{1/2}$ , and it follows from Taylor series that for any such holomorphic function  $Z$ ,

$$Z(A) = G^{-1/2}Z(\tilde{V})G^{1/2}, \quad Z(B) = Z(-A)^* \supseteq G^{1/2}Z(\tilde{V})G^{-1/2}. \quad (16)$$

The results in Theorems 5–9 are for compactifications based on general kernels satisfying Assumptions 1 and 3 and their associated integral operators. Next, we establish spectral convergence results for one-parameter families of kernels that include the kernels  $p_\tau$  associated with the Markov semigroups in our main result, Theorem 2. Specifically, we assume:

**Assumption 4.**  $\{k_\tau : M \times M \rightarrow \mathbb{R}\}$  with  $\tau > 0$  is a one-parameter family of  $C^1$ , symmetric,  $L^2(\mu)$ -strictly-positive kernels, such that, as  $\tau \rightarrow 0^+$ , the sequence of the corresponding compact operators  $G_\tau = K_\tau^*K_\tau$  on  $L^2(\mu)$  converges strongly to the identity, and the sequence of the skew-adjoint compactified generators  $\tilde{V}_\tau \supseteq G_\tau^{1/2}VG_\tau^{1/2}$  converges strongly to  $V$  on the subspace  $D(V^2) \subset D(V)$ .

Let  $\mathcal{H}_\tau$  be the RKHS on  $M$  with reproducing kernel  $k_\tau$ ,  $\mathcal{N}_\tau : D(\mathcal{N}_\tau) \rightarrow \mathcal{H}_\tau$  the corresponding Nyström extension operator, and  $H_\infty$  the  $L^2(\mu)$  subspace equal to  $\cap_{\tau > 0} D(\mathcal{N}_\tau)$ . Define the partial isometries  $\mathcal{U}_\tau : L^2(\mu) \rightarrow \mathcal{H}_\tau$  through the polar decomposition  $K_\tau = \mathcal{U}_\tau G_\tau^{1/2}$ , as in Theorem 8. Note that, in general,  $H_\infty$  could be the zero subspace, but contains at least constant functions if the  $k_\tau$  are  $L^2(\mu)$ -Markov kernels. As stated in Section 2, if  $H_\infty$  is the space associated with the kernels  $p_\tau$  from (10), whose corresponding integral operators form a Markov semigroup and thus have common eigenspaces, then it is even dense in  $L^2(\mu)$ . With these definitions, we establish the following notion of spectral convergence for approximations of the generator  $V$  by compact operators.

**Theorem 10 (Spectral convergence).** Suppose that Assumptions 1 and 4 hold with  $r = 1$ , and let  $A_\tau, B_\tau, \tilde{V}_\tau : L^2(\mu) \rightarrow L^2(\mu)$  and  $W_\tau : \mathcal{H}_\tau \rightarrow \mathcal{H}_\tau$  with  $\tau > 0$ , be the Hilbert-Schmidt operators from Theorems 5–8, associated with the kernels  $k_\tau$  from Assumption 4. Let also  $\tilde{E}_\tau$  and  $\mathcal{E}_\tau$  be the PVMs associated with  $\tilde{V}_\tau$  and  $W_\tau$ , respectively, constructed as in (14). Then, as  $\tau \rightarrow 0^+$ , the following hold:

- (i) The operator  $B_\tau$  converges strongly to  $V$  on  $D(V)$ .
- (ii) For every bounded continuous function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ ,  $Z(\tilde{V}_\tau)$  and  $\mathcal{U}_\tau^*Z(W_\tau)\mathcal{U}_\tau$  converge strongly to  $Z(V)$  on  $L^2(\mu)$ .
- (iii) For every holomorphic function  $Z : D(Z) \rightarrow \mathbb{C}$ , with  $i\mathbb{R} \subset D(Z) \subseteq \mathbb{C}$  and  $Z|_{i\mathbb{R}}$  bounded,  $Z(A_\tau)$  and  $Z(B_\tau)$  converge strongly to  $Z(V)$  on  $L^2(\mu)$ . Moreover,  $K_\tau^*Z(W_\tau)\mathcal{N}_\tau$  converges strongly to  $Z(V)$  on  $H_\infty$ .
- (iv) For every bounded Borel-measurable set  $\Omega \subset \mathbb{R}$  such that  $\tilde{E}(\partial\Omega) = 0$ ,  $\tilde{E}_\tau(\Omega)$  and  $\mathcal{U}_\tau^*\mathcal{E}_\tau(\Omega)\mathcal{U}_\tau$  converge strongly to  $E(\Omega)$  on  $L^2(\mu)$ .
- (v) For every element  $i\omega$  of the spectrum of  $V$ , there exists a sequence of eigenvalues  $i\omega_\tau$  of  $A_\tau$ ,  $B_\tau$ ,  $\tilde{V}_\tau$ , and  $W_\tau$  converging to  $i\omega$ .

Theorem 10 makes several of the statements of our main result, Theorem 2. In Section 5, we will prove the latter by invoking Theorems 5–10 for the family of Markov kernels  $p_\tau$ . There, the semigroup structure of  $p_\tau$  will allow us to extend the convergence result for  $K_\tau^* Z(W_\tau) \mathcal{N}_\tau$  from holomorphic functions to bounded continuous functions  $Z$ , and further deduce that  $A_\tau$ ,  $B_\tau$ ,  $\tilde{V}_\tau$ , and  $W_\tau$  are of trace class.

#### 4. Results from functional analysis and analysis on manifolds

In this section, we review some basic concepts from RKHS theory, spectral approximation of operators, and analysis on manifolds that will be useful in our proofs of the theorems stated in Sections 2 and 3.

##### 4.1. Results from RKHS theory

*Nyström extension.* We begin by describing the Nyström extension in RKHS. In what follows,  $\mathcal{H}$  will be an RKHS on  $M$  with reproducing kernel  $k$ ,  $\nu$  an arbitrary finite Borel measure with compact support  $X_\nu \subseteq M$ , and  $K : L^2(\nu) \rightarrow \mathcal{H}$  the corresponding integral operator defined via (7). The Nyström extension operator  $\mathcal{N} : D(\mathcal{N}) \rightarrow \mathcal{H}$ , with  $D(\mathcal{N}) \subset L^2(\nu)$ , extends elements of its domain, which are equivalence classes of functions defined up to sets of  $\nu$  measure zero, to functions in  $\mathcal{H}$ , which are defined at every point in  $M$  and can be pointwise evaluated by continuous linear functionals. Specifically, introducing the functions

$$\psi_j = \lambda_j^{-1/2} K \phi_j, \quad j \in J, \quad (17)$$

where  $\{\phi_0, \phi_1, \dots\}$  is an orthonormal set in  $L^2(\nu)$  consisting of eigenfunctions of  $G = K^* K$ , corresponding to strictly positive eigenvalues  $\lambda_0 \geq \lambda_1 \geq \dots$ , and  $J = \{j \in \mathbb{N}_0 : \lambda_j > 0\}$ , we define

$$D(\mathcal{N}) = \left\{ \sum_{j \in J} a_j \phi_j : \sum_{j \in J} |a_j|^2 / \lambda_j < \infty \right\}, \quad \mathcal{N} \left( \sum_{j \in J} a_j \phi_j \right) := \sum_{j \in J} a_j \lambda_j^{-1/2} \psi_j. \quad (18)$$

It follows directly from these definitions that  $\{\psi_j\}_{j \in J}$  is an orthonormal set in  $\mathcal{H}$  satisfying  $K^* \psi_j = \lambda_j^{1/2} \phi_j$ , and  $\mathcal{N}$  is a closed-range, closed operator with  $D(\mathcal{N}) = \text{ran } K^*$  and  $\text{ran } \mathcal{N} = \mathcal{K} := \overline{\text{ran } K} = \overline{\text{span}\{\psi_j\}_{j \in J}}$ . Moreover,  $K^* \mathcal{N}$  and  $\mathcal{N} K^*$  reduce to the identity operators on  $D(\mathcal{N})$  and  $\text{ran } \mathcal{N}$ , respectively. In fact, upon restriction to  $X_\nu$ ,  $\text{ran } \mathcal{N}$  coincides with the RKHS  $\mathcal{H}(X_\nu)$ , and  $\{\psi_j|_X\}_{j \in J}$  forms an orthonormal basis of the latter space. If, in addition, the kernel  $k$  is  $L^2(\nu)$ -strictly-positive, as we frequently require in this paper, then  $D(\mathcal{N})$  is a dense subspace of  $L^2(\nu)$ , and  $K^*$  coincides with the pseudoinverse of  $\mathcal{N}$ . The latter is defined as the unique bounded operator  $\mathcal{N}^\dagger : \mathcal{H} \rightarrow L^2(\mu)$  satisfying (i)  $\ker \mathcal{N}^\dagger = \text{ran } \mathcal{N}^\perp$ ; (ii)  $\overline{\text{ran } \mathcal{N}^\dagger} = \ker \mathcal{N}^\perp$ ; and (iii)  $\mathcal{N} \mathcal{N}^\dagger f = f$ , for all  $f \in \text{ran } \mathcal{N}$ . Note that we have described the Nyström extension for the  $L^2$  space associated with an arbitrary compactly supported Borel measure  $\nu$  since later on we will be interested in applying this procedure not only for the invariant measure  $\mu$  of the system, but also for discrete sampling measures encountered in data-driven approximation schemes.

*Polar decomposition.* A number of the results stated in Sections 2 and 3 make use of the polar decomposition of kernel integral operators associated with RKHSs. We now review this construction. First, recall that the polar decomposition of a bounded linear map  $T : H_1 \rightarrow H_2$  between two Hilbert spaces  $H_1$  and  $H_2$  is the unique factorization  $T = \mathcal{U}|T|$ , where  $|T| = (T^* T)^{1/2}$  is a non-negative, self-adjoint operator on  $H_1$ , and  $\mathcal{U} : H_1 \rightarrow H_2$  is a partial isometry with  $\ker \mathcal{U}^\perp = \overline{\text{ran } |T|}$ . The spaces  $\ker \mathcal{U}^\perp$  and  $\text{ran } \mathcal{U}$  are known as the initial and final spaces of the partial isometry  $\mathcal{U}$ . In the case of the integral operator  $K : L^2(\nu) \rightarrow \mathcal{H}$ , we have  $K = \mathcal{U}|K|$ , where  $|K| = G^{1/2}$  by definition of  $G = K^* K$ . Moreover, it follows from the relationships  $K \phi_j = \lambda_j^{1/2} \psi_j$  and  $G^{1/2} \phi_j = \lambda_j^{1/2} \phi_j$ , which hold for every  $j \in J$ , that  $\mathcal{U} \phi_j = \psi_j$  for  $j \in J$ . Thus, the

initial and final spaces of  $\mathcal{U}$  are given by  $\ker \mathcal{U}^\perp = \overline{\text{ran } K^*} = \overline{D(\mathcal{N})}$  and  $\text{ran } \mathcal{U} = \overline{\text{ran } K} = \mathcal{K}$ , respectively. In addition, since  $K^* \psi_j = \lambda_j^{1/2} \phi_j = G^{1/2} \phi_j$ , we can conclude that  $\text{ran } G^{1/2} = \text{ran } K^*$ , and

$$D(G^{-1/2}) = D(\mathcal{N}), \quad \mathcal{N} = \mathcal{U}G^{-1/2}, \quad K^*\mathcal{U} = G^{1/2}. \quad (19)$$

*Mercer representation.* A classical result in the theory of RKHSs with continuous kernels is the Mercer theorem [60], allowing one to represent the kernel through eigenfunctions. In the following lemma, we will state this result together with a useful integral formula for computing the trace of integral operators associated with continuous kernels.

**Lemma 11.** *Let  $\mathcal{H}$  be an RKHS on  $M$  associated with a continuous reproducing kernel  $k$ , and  $\nu$  a finite Borel measure with compact support  $X_\nu \subseteq M$ . Assume, further, the notations in (17). Then, the following hold:*

- (i) *(Mercer theorem) For every  $x, y \in X_\nu$ ,  $k(x, y) = \sum_{j \in J} \psi_j^*(x) \psi_j(y)$ , where the sum converges absolutely and uniformly with respect to  $(x, y) \in X_\nu \times X_\nu$ .*
- (ii) *The trace of the integral operator  $G = K^*K$  is equal to  $\int_M k(x, x) d\nu(x)$ .*

**Proof.** We will only prove Claim (ii). For that, we use Claim (i) to compute explicitly

$$\begin{aligned} \int_M k(x, x) d\nu(x) &= \int_{X_\nu} k(x, x) d\nu(x) = \int_{X_\nu} \sum_{j \in J} \psi_j^*(x) \psi_j(x) d\nu(x) = \sum_{j \in J} \int_{X_\nu} |\psi_j(x)|^2 d\nu(x) \\ &= \sum_{j \in J} \int_{X_\nu} |K^* \psi_j|^2 d\nu = \sum_{j \in J} \int_{X_\nu} \lambda_j |\phi_j|^2 d\nu = \sum_{j \in J} \lambda_j \int_{X_\nu} |\phi_j|^2 d\nu = \sum_{j \in J} \lambda_j = \text{tr } G. \end{aligned}$$

The last equality on the first line follows from the absolute convergence of  $\sum_{j \in J} |\psi_j(x)|^2$  to  $k(x, x)$ . The first equality in the second line follows from the fact that  $K^*$  is the  $L^2(\nu)$ -inclusion operator on  $\mathcal{H}$ .  $\square$

*Bistochastic kernel normalization.* Our main result, Theorem 2, as well as a number of the auxiliary results in Theorem 9, require that the reproducing kernel under consideration be Markovian. However, the notion of Markovianity depends on a choice of measure (e.g., in the case of Theorems 2 and 9, the invariant measure  $\mu$ ), which is usually either unknown, or integrals with respect to it cannot be evaluated in closed form. As a result, a common approach to building Markov kernels is to start from a positive-valued unnormalized kernel, which can be evaluated in closed form, and then perform a normalization procedure to render it Markovian. Such kernel normalizations are widely used in manifold learning [29,43,44], spectral clustering [30], and other applications. However, many of these approaches produce non-symmetric kernels which are not suitable for defining RKHSs. Here, we construct symmetric Markov kernels with associated RKHSs using the bistochastic normalization procedure introduced in [58], which yields symmetric, positive-definite Markov kernels with corresponding RKHSs. The starting point for this construction is a kernel  $k$  on  $M$  satisfying Assumption 3, and in addition, being strictly positive-valued everywhere, i.e.,  $k > 0$ . Given a Borel probability measure  $\nu$  with compact support  $X_\nu \subseteq M$ , the kernel  $k$  induces the functions  $d : M \rightarrow \mathbb{R}$  and  $q : M \rightarrow \mathbb{R}$  such that

$$d(x) = \int_M k(x, y) d\nu(y), \quad q(x) = \int_M \frac{k(x, y)}{d(y)} d\nu(y).$$

By strict positivity and  $C^r$  regularity of  $k$  and compactness of  $X_\nu$ , the functions  $d$ ,  $q$ ,  $1/d$ , and  $1/q$  are strictly positive and  $C^r$ . As a result,  $p : M \times M \rightarrow \mathbb{R}$ , with

$$p(x, y) = \int_M \frac{k(x, z)k(z, y)}{d(x)q(z)d(y)} d\nu(z) \quad (20)$$

is also a  $C^r$ , positive-definite kernel with  $p > 0$ . It then follows by construction that  $p$  is symmetric and satisfies  $\int_M p(x, \cdot) d\nu = 1$  for all  $x \in M$ . That is,  $p$  is a positive-definite, symmetric, and  $L^2(\nu)$ -Markov ergodic kernel. In fact, if the kernel  $k$  is strictly positive-definite on  $X_\nu \times X_\nu$ , then  $p$  is also strictly positive-definite on that set, and thus is  $L^2(\nu)$ -strictly-positive. To verify this, note that

$$p(x, y) = \int_M \frac{\tilde{k}(x, z)\tilde{k}(z, y)}{\tilde{d}(x)\tilde{d}(y)} d\nu(z), \quad \tilde{k}(x, y) = \frac{k(x, y)}{q^{1/2}(x)q^{1/2}(y)}, \quad \tilde{d}(x) = \frac{d(x)}{q^{1/2}(x)},$$

and because  $x \mapsto \tilde{d}(x)$  is a strictly positive continuous function, it suffices to show that the kernel  $\tilde{k}_2(x, y) = \int_M \tilde{k}(x, z)\tilde{k}(z, y) d\nu(z)$  is strictly positive-definite on  $L^2(\nu)$ . Now note that  $\tilde{k}$  is a strictly positive-definite kernel on  $X_\nu \times X_\nu$  by strict positive-definiteness of  $k$  and strict positivity of the continuous function  $x \mapsto \tilde{q}(x)$ . Thus, in order to verify that  $\tilde{k}_2$ , and thus  $p$ , is strictly positive-definite on  $X_\nu \times X_\nu$ , it suffices to show:

**Lemma 12.** *Let  $\nu$  be a finite Borel measure with compact support  $X_\nu \subseteq M$ , and  $k : X_\nu \times X_\nu \rightarrow \mathbb{R}$  a symmetric, strictly positive-definite kernel. Then, the kernel  $k_2 : X_\nu \times X_\nu \rightarrow \mathbb{R}$ , with  $k_2(x, y) = \int_M k(x, z)k(z, y) d\nu(z)$  is strictly positive-definite.*

**Proof.** We must show that for any collection of distinct points  $x_0, \dots, x_{m-1} \in X_\nu$  the  $m \times m$  kernel matrix  $\mathbf{G}_2 := [k_2(x_i, x_j)]$  is positive definite. Defining  $\nu_m = \sum_{j=0}^{m-1} \delta_{x_j}/m$ , this is equivalent to showing that the operator  $G_2 : L^2(\nu_m) \rightarrow L^2(\nu_m)$  with matrix representation  $\mathbf{G}_2$  in the standard orthonormal basis of  $L^2(\nu_m)$  is positive. To that end, observe that  $G_2 = (K^* K_m)^* K^* K_m$ , where  $K_m : L^2(\nu_m) \rightarrow \mathcal{H}(X_\nu)$  and  $K : L^2(\nu) \rightarrow \mathcal{H}(X_\nu)$  are the integral operators associated with  $k$  and the measures  $\nu_m$  and  $\nu$ , respectively, mapping into the RKHS  $\mathcal{H}(X_\nu)$  associated with  $k$ . Because  $K_m$  is an injective operator by strict positive-definiteness of  $k$ , and  $K^*$  is injective by definition,  $K^* K_m$  is injective, and for every nonzero  $f \in L^2(\nu_m)$ ,  $\langle f, G_2 f \rangle_{L^2(\nu_m)} = \langle K^* K_m f, K^* K_m f \rangle_\nu > 0$ . This shows that  $G_2$  is positive, and thus  $k_2$  is a strictly positive-definite kernel, proving the lemma.  $\square$

In summary, we have established that if the kernel  $k$  satisfies Assumption 3, and is also positive-valued and strictly positive-definite on the support of  $\nu$ , then the bistochastic normalization procedure in (20) yields a  $C^r$ , strictly positive definite, and thus  $L^2(\nu)$ -strictly-positive, Markov ergodic kernel. In particular, if it happens that  $k$  is strictly positive-definite on  $M \times M$ , the kernel from (20) is  $L^2(\nu)$ -strictly-positive and Markov ergodic for every compactly supported Borel probability measure  $\nu$ . This approach therefore provides a convenient way of constructing Markov kernels meeting the conditions of Theorem 1. In Section 8, we will employ bistochastic normalization of strictly positive-definite, positive-valued kernels to construct data-driven approximations to the Markov kernels in Theorem 1 that converge in the limit of large data.

#### 4.2. Spectral approximation of operators

*Strong resolvent convergence.* In order to prove the various spectral convergence claims made in Sections 2 and 3, we need appropriate notions of convergence of operators approximating the generator  $V$  that imply spectral convergence. Clearly, because  $V$  is unbounded, it is not possible to employ convergence in operator norm for that purpose. In fact, for the approximations studied here, even strong convergence on the domain of  $V$  may not necessarily hold. For example, in an approximation of  $V$  by  $VT_\tau$ , where  $T_\tau$ ,  $\tau \geq 0$ , is a family of smoothing operators on  $L^2(\mu)$  with  $\text{ran } T_\tau \subseteq D(V)$ , the convergence of  $T_\tau f$  to  $f$  as  $\tau \rightarrow 0^+$  does

not necessarily imply that  $VT_\tau f$  converges to  $Vf$ , as  $V$  is unbounded. In the setting of unbounded, skew-adjoint operators, a weaker form of convergence, which is nevertheless sufficient to establish our spectral convergence claims, is *strong resolvent convergence* [51].

To wit, let  $T : D(T) \rightarrow H$  be a skew-adjoint operator on a Hilbert space  $H$ , and consider a sequence of operators  $T_\tau : D(T_\tau) \rightarrow H$  indexed by a parameter  $\tau > 0$ . The sequence  $T_\tau$  is said to converge to  $T$  as  $\tau \rightarrow 0^+$  in strong resolvent sense if for every complex number  $\rho$  in the resolvent set of  $T$ , not lying on the imaginary line, the resolvents  $(\rho - T_\tau)^{-1}$  converge to  $(\rho - T)^{-1}$  strongly. If  $T_\tau$  is bounded, for every quadratic polynomial  $\mathbf{p}$ ,  $\mathbf{p}(iT_\tau)$  is also bounded. Following [61], we say that the sequence  $T_\tau$  is *p2-continuous* if every  $T_\tau$  is bounded, and the function  $\tau \mapsto \|\mathbf{p}(iT_\tau)\|$  is continuous for every such  $\mathbf{p}$ . Henceforth, when convenient, we will use the notation  $\xrightarrow{s}$  and  $\xrightarrow{sr}$  to indicate strong convergence and strong resolvent convergence, respectively.

As we will see in Lemma 14 below,  $T_\tau \xrightarrow{s} T$  implies  $T_\tau \xrightarrow{sr} T$ . Further, if  $T$  is bounded and the sequence  $T_\tau$  is uniformly bounded in operator norm, then  $T_\tau \xrightarrow{sr} T$  implies  $T_\tau \xrightarrow{s} T$  [51, Proposition 10.1.13]. These facts indicate that strong resolvent convergence can be viewed as a generalization of strong convergence of bounded operators. For our purposes, the usefulness of strong resolvent convergence is that it implies the following convergence results for spectra and Borel functional calculi of skew-adjoint operators.

**Proposition 13.** *Suppose that  $T_\tau : D(T_\tau) \rightarrow H$  is a sequence of skew-adjoint operators converging in strong resolvent sense as  $\tau \rightarrow 0^+$  to a skew-adjoint operator  $T : D(T) \rightarrow H$ . Let also  $\Theta_\tau : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(H)$  and  $\Theta : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{L}(H)$  be the PVMs associated with  $T_\tau$  and  $T$ , respectively. Then:*

- (i) *For every bounded, continuous function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ ,  $Z(T_\tau)$  converges strongly to  $Z(T)$ .*
- (ii) *Let  $J \subset J' \subset i\mathbb{R}$  be two bounded intervals. Then, for every  $f \in L^2(\mu)$ ,  $\limsup_{\tau \rightarrow 0^+} \|1_J(T_\tau)f\|_{L^2(\mu)} \leq \|1_{J'}(T)f\|_{L^2(\mu)}$ .*
- (iii) *For every bounded, Borel-measurable set  $\Omega \subset \mathbb{R}$  such that  $\Theta(\partial\Omega) = 0$ ,  $\Theta_\tau(\Omega)$  converges strongly to  $\Theta(\Omega)$ .*
- (iv) *For every bounded, Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$  of bounded support,  $Z(T_\tau)$  converges strongly to  $Z(T)$ , provided that  $\Theta(S) = 0$ , where  $S \subset \mathbb{R}$  is a closed set such that  $iS$  contains the discontinuities of  $Z$ .*
- (v) *If  $T$  is bounded, (iii) holds for every Borel-measurable set  $\Omega \subseteq \mathbb{R}$ , and (iv) for every bounded Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ .*
- (vi) *If the operators  $T_\tau$  are compact, then for every element  $\theta \in i\mathbb{R}$  of the spectrum of  $T$ , there exists a one-parameter family  $\theta_\tau \in i\mathbb{R}$  of eigenvalues of  $T_\tau$  such that  $\lim_{\tau \rightarrow 0^+} \theta_\tau = \theta$ . Moreover, if the sequence  $T_\tau$  is p2-continuous, the curve  $\tau \rightarrow \theta_\tau$  is continuous.*

**Proof.** Claim (i) is actually an equivalent characterization of strong resolvent convergence [51, Proposition 10.1.9]. Claim (v) is a classical result from spectral approximation theory for normal, bounded operators, e.g., [62, Chapter 8, Theorem 2]. In Claim (vi), the existence of the family  $\theta_\tau$  follows from [51, Corollary 10.2.2], in conjunction with compactness of  $T_\tau$ . The continuity of  $\tau \mapsto \theta_\tau$  follows from [61, Theorem 1].

It now remains to prove Claims (ii)–(iv). Starting from Claim (ii), note that a property of the Borel functional calculus for a skew-adjoint operator  $T : D(T) \rightarrow H$  (more commonly stated for self-adjoint operators, e.g., [63]) is that for any Borel-measurable function  $Z : i\mathbb{R} \rightarrow \mathbb{R}$  lying in  $L^\infty(i\mathbb{R})$ ,  $Z(T)$  is a bounded self-adjoint operator. Moreover, this functional calculus preserves positivity, in the sense that if  $Z$  is positive, then  $Z(T)$  is positive operator, and as a result  $Z(T) \leq Z'(T)$  whenever  $Z \leq Z'$ . With these properties, let  $Z : i\mathbb{R} \rightarrow \mathbb{R}$  be a piecewise-linear continuous function equal to 1 on  $J$ , and with support contained in  $J'$ . Let also  $1_\Omega$  be the indicator function of any set  $\Omega$ . Then, the inequalities  $1_J^2 \leq Z^2 \leq 1_{J'}^2$  hold everywhere in  $i\mathbb{R}$ , so for each  $\tau > 0$ ,  $1_J^2(T_\tau) \leq Z^2(T_\tau) \leq 1_{J'}^2(T)$ . In addition, since  $Z$  is continuous and

bounded,  $Z(T_\tau)$  converges strongly to  $Z(T)$  by Claim (i). The proof of Claim (ii) can now be completed using the following inequality:

$$\begin{aligned}
\limsup_{\tau \rightarrow 0^+} \|1_J(T_\tau)f\|_{L^2(\mu)}^2 &= \limsup_{\tau \rightarrow 0^+} \langle 1_J(T_\tau)f, 1_J(T_\tau)f \rangle_\mu = \limsup_{\tau \rightarrow 0^+} \langle 1_J^2(T_\tau)f, f \rangle_\mu \\
&\leq \limsup_{\tau \rightarrow 0^+} \langle Z^2(T_\tau)f, f \rangle_\mu = \limsup_{\tau \rightarrow 0^+} \langle Z(T_\tau)f, Z(T_\tau)f \rangle_\mu \\
&= \limsup_{\tau \rightarrow 0^+} \|Z(T_\tau)f\|_{L^2(\mu)}^2 = \|Z(T)f\|_{L^2(\mu)}^2 \\
&= \langle Z(T)f, Z(T)f \rangle_\mu = \langle Z^2(T)f, f \rangle_\mu \\
&\leq \langle 1_{J'}^2(T)f, f \rangle_\mu = \|1_{J'}(T)f\|_{L^2(\mu)}^2.
\end{aligned}$$

Next, we will prove Claim (iii) in the case that  $\Omega$  is an interval  $[a, b] \subset \mathbb{R}$  with  $\Theta(\partial\Omega) = \Theta(\{a, b\}) = 0$  (i.e., neither of  $ia$  and  $ib$  is an eigenvalue of  $T$ ). Given any  $w > 0$ , let  $f_w : i\mathbb{R} \rightarrow i\mathbb{R}$  be a continuous function such that  $f_w(i\omega)$  equals  $i\omega$  for  $\omega \in i[a, b]$ , equals 0 outside  $i[a - w, b + w]$ , and is linear on the intervals  $i[a - w, a]$  and  $i[b, b + w]$ . By Claim (i),  $\lim_{\tau \rightarrow 0^+} f_w(T_\tau) = f_w(T)$ . Moreover, the operators  $f_w(T_\tau), f_w(T)$  are bounded and skew-adjoint, and therefore, by Claim (v), for every bounded, measurable  $g : i\mathbb{R} \rightarrow \mathbb{R}$ ,

$$\lim_{\tau \rightarrow 0^+} (g \circ f_w)(T_\tau) = \lim_{\tau \rightarrow 0^+} g(f_w(T_\tau)) = g(f_w(T)) = (g \circ f_w)(T). \quad (21)$$

Setting  $g = 1_{i\Omega}$  then leads to

$$g \circ f_w = 1_{i\Omega} + 1_{J_w}, \quad J_w := [b, b + w] \cap f_w^{-1}(\Omega).$$

Thus, substituting for  $g \circ f_w$  in (21) using the latter identity, and rearranging, we obtain

$$\lim_{\tau \rightarrow 0^+} [\Theta_\tau(\Omega) + \Theta_\tau(J_w) - \Theta(J_w)] = \Theta(\Omega), \quad \forall w > 0. \quad (22)$$

Note that here we have used the fact that for any Borel set  $S \subset \mathbb{R}$ ,  $\Theta(S) = 1_{iS}(T)$ , and a similar fact for  $T_\tau$ . The operator  $\Theta(J_w)$  is the spectral projection onto the subspace  $H_w = \text{ran } \Theta(J_w) \subseteq H$ . Since  $\Theta(\partial\Omega) = 0$  and  $\cap_{w>0} J_w = \{b\}$ , we have  $\cap_{w>0} H_w = \{0\}$ . As a result, as  $w \rightarrow 0^+$ , the  $H_w^\perp$  form an increasing sequence of subspaces with  $\cup_{w>0} H_w^\perp = H$ . Thus, to prove that  $\Theta_\tau(\Omega)$  converges strongly to  $\Theta(\Omega)$ , it is enough to verify the same claim on  $H_{w_0}^\perp$  for every fixed  $w_0 > 0$ . To that end, let  $w_0 > 0$  be fixed, and consider an arbitrary  $f \in H_{w_0}^\perp$ . By construction,  $\Theta(J_w)f$  vanishes for every  $0 < w < w_0$ . Moreover, by Claim (ii),

$$\limsup_{\tau \rightarrow 0^+} \|\Theta_\tau(J_w)f\|_{L^2(\mu)} = \limsup_{\tau \rightarrow 0^+} \|1_{iJ_w}(T_\tau)f\|_{L^2(\mu)} \leq \|1_{iJ_{w'}}(T_\tau)f\|_{L^2(\mu)} = 0, \quad \forall w' \in (w, w_0),$$

from which it follows that  $\lim_{\tau \rightarrow 0^+} \Theta_\tau(J_w)f = 0$ . Thus, substituting the identities  $\Theta(J_w)f = 0$  and  $\lim_{\tau \rightarrow 0^+} \Theta_\tau(J_w)f = 0$  into (22) yields

$$\Theta(\Omega)f = \lim_{\tau \rightarrow 0^+} [\Theta_\tau(\Omega) + \Theta_\tau(J_w) - \Theta(J_w)]f = \lim_{\tau \rightarrow 0^+} \Theta_\tau(\Omega)f + \lim_{\tau \rightarrow 0^+} \Theta_\tau(J_w)f + \Theta(J_w)f = \lim_{\tau \rightarrow 0^+} \Theta_\tau(\Omega)f,$$

proving that Claim (iii) is true for  $\Omega$  equal to an interval.

We now extend this result to the case that  $\Omega$  is an arbitrary bounded Borel subset of  $\mathbb{R}$  with  $\Theta(\partial\Omega) = 0$ . For that, it is sufficient to fix an arbitrary  $b > 0$ , and prove the result for the elements of the set  $\Sigma' = \{\Omega \in \mathcal{B}([-b, b]) : \Theta(\partial\Omega) = 0\}$ , where  $\mathcal{B}([-b, b])$  is the Borel  $\sigma$ -algebra on  $[-b, b]$ . Let then  $\Sigma$  be the collection of subsets  $\Omega \subseteq \mathcal{B}([-b, b])$ , such that  $\Theta_\tau(\Omega)$  converges strongly to  $\Theta(\Omega)$ . It can be shown that  $\Sigma$  is a  $\sigma$ -algebra. Moreover,  $\Sigma$  contains all intervals having zero  $\Theta$  measure on their boundary, and thus must also contain the

$\sigma$ -algebra generated by such intervals. But this latter  $\sigma$ -algebra contains  $\Sigma'$ , and therefore  $\Theta_\tau(\Omega) \xrightarrow{s} \Theta(\Omega)$  for all  $\Omega \in \Sigma'$ , proving Claim (iii).

Finally, we prove Claim (iv). Let  $Z$  be as claimed, with support contained in a bounded open interval  $I \subset i\mathbb{R}$ . Then, the set  $I \setminus S$  is a countable union of bounded open intervals  $I_1, I_2, \dots$ . Note that  $H$  is the direct sum of the mutually orthogonal spaces  $\text{ran } \Theta(I_1), \text{ran } \Theta(I_2), \dots, \text{ran } \Theta(S)$ , and  $\text{ran } \Theta(I^c)$ . Among these,  $\text{ran } \Theta(I^c)$  is contained in the kernel of  $Z(T)$ . Moreover, in a manner similar to the proof of Claim (iii), it can be shown that  $Z|_S(T_\tau)$  converges pointwise to 0. Thus, for every  $f \in \text{ran } \Theta(S)$ ,  $Z(T_\tau)f = Z|_S(T_\tau)f$  converges to  $Tf = 0$ . In light of these facts, Claim (iv) can be simplified to the case that  $Z$  is a continuous function supported on an interval  $(ia, ib) \subset i\mathbb{R}$  with  $\Theta(\{a, b\}) = 0$ . In this case, constructing a function  $f_w$  as in Claim (iii), and using the same line of reasoning, it can be shown that  $Z(T_\tau) \rightarrow Z(T)$  as  $\tau \rightarrow 0^+$ . This proves Claim (iv) and the Proposition.  $\square$

Proposition 13 lays the foundation for many of the spectral convergence results in Theorem 10, and thus Theorem 2. It also highlights, through Claim (iii), the convergence properties for the functional calculus and spectrum lost from the fact that  $V$  is unbounded. Yet, despite the usefulness of the results stated in Proposition 13, the basic assumption made, namely that  $T_\tau$  converges to  $T$  in strong resolvent sense, is oftentimes difficult to explicitly verify. Fortunately, in the case of skew-adjoint operators of interest here, there exist sufficient conditions for strong resolvent convergence, which are easier to verify. Before stating these conditions, we recall that a *core* for a closed operator  $T : D(T) \rightarrow H$  on a Hilbert space  $H$  is any subspace  $C \subseteq D(T)$  such that  $T$  is the closure of the restricted operator  $T|_C$ . In other words,  $C$  is a core if the closure of the graph of  $T|_C$ , as a subset of  $H \times H$ , is the graph of  $T$ . Note that  $T$  may not have a unique core. We also introduce the notion of convergence in the *strong dynamical sense* [51]. Specifically, a sequence  $T_\tau : D(T_\tau) \rightarrow H$ ,  $\tau > 0$ , of skew-adjoint operators is said to converge to  $T : D(T) \rightarrow H$  as  $\tau \rightarrow 0$  in the strong dynamical sense if  $e^{tT_\tau}$  converges strongly to  $e^{tT}$  for every  $t \in \mathbb{R}$ . Note that in the case of the operators  $\tilde{V}_\tau$  from Assumption 4 approximating the generator  $V$ , strong dynamical convergence means that the unitary operators  $e^{t\tilde{V}_\tau}$  converge strongly to the Koopman operator  $U^t = e^{tV}$  for every time  $t \in \mathbb{R}$ .

**Lemma 14.** *Let  $T_\tau : D(T_\tau) \rightarrow H$  and  $T : D(T) \rightarrow H$  be the skew-adjoint operators from Proposition 13. Then, the following hold:*

- (i) *The domain  $D(T^2)$  of the operator  $T^2$  is a core for  $T$ .*
- (ii) *If  $T_\tau$  converges pointwise to  $T$  on a core for  $T$ , then it also converges in strong resolvent sense.*
- (iii) *Strong resolvent convergence of  $T_\tau$  to  $T$  is equivalent to strong dynamical convergence.*

**Proof.** Claim (i) follows from [64, Theorem 5]. Claims (ii) and (iii) follow from Propositions 10.1.18 and 10.1.8, respectively, of [51]. There, the statements are for self-adjoint operators, but they apply to skew-adjoint operators as well.  $\square$

**Remark.** Lemma 14(ii) indicates that a sufficient condition for strong resolvent convergence of a sequence skew-adjoint operators is pointwise convergence in a smaller domain (a core) than the full domain of the limit operator; that is, strong resolvent convergence is weaker than strong convergence for this class of operators. In Proposition 19 ahead, we will see that the operator family  $\tilde{V}_\tau$  employed in Theorem 2 actually converges pointwise to  $V$  on the whole of  $D(V)$ .

*Approximate point spectrum and pseudospectrum.* Generalizing the definition in (12), we say that a complex number  $\gamma$  lies in the  $\epsilon$ -approximate point spectrum  $\sigma_{ap,\epsilon}(T)$  of a closed operator  $T : D(T) \rightarrow H$  on a Hilbert space  $H$  for  $\epsilon > 0$ , if there exists  $f \in H$ , with  $\|f\|_H = 1$ , such that [65,66]

$$\|Tf - \gamma f\|_H < \epsilon. \quad (23)$$

As  $\epsilon$  decreases towards 0,  $\sigma_{ap,\epsilon}(T)$  forms an increasing family of open subsets of the complex plane, such that  $\cup_{\epsilon>0}\sigma_{ap,\epsilon}(T) = \mathbb{C}$ . Moreover, if  $T$  is a normal operator,  $\sigma_{ap,\epsilon}(T)$  is the union of all open  $\epsilon$ -balls in the complex plane with centers lying in its spectrum,  $\sigma(T)$ . If, in addition,  $T$  is bounded,  $\cap_{\epsilon>0}\sigma_{ap,\epsilon}(T) = \sigma(T)$ . The  $\epsilon$ -approximate point spectrum is also a subset of the  $\epsilon$ -pseudospectrum  $\sigma_\epsilon(T)$  of  $T$ , defined as the set of complex numbers  $\gamma$  such that  $\|(T - \gamma)^{-1}\| > 1/\epsilon$ , with the convention that  $\|(T - \gamma)^{-1}\| = \infty$  if  $\gamma \in \sigma(T)$  [67]. Specifically,  $\sigma_\epsilon(T) = \sigma_{ap,\epsilon}(T) \cup \sigma(T)$ , and if  $T$  is normal and bounded,  $\sigma_\epsilon(T) = \sigma_{ap,\epsilon}(T)$ . For our purposes, a distinguished property of each element  $\gamma \in \sigma_{ap,\epsilon}(T)$  is that there exists an associated unit-norm vector  $f \in H$  which behaves approximately as an eigenfunction of  $T$ , in the sense of (23).

#### 4.3. Results from analysis on manifolds

We will state a number of standard results from analysis on manifolds that will be used in the proofs presented in Sections 5 and 7. In what follows, we consider that  $M$  is a  $C^r$  compact manifold, equipped with an arbitrary  $C^{r-1}$  Riemannian metric (e.g., a metric induced from the ambient space  $\mathcal{M}$ , or the embedding  $F : M \rightarrow Y$  into the data space  $Y$  from Section 8), and an associated covariant derivative operator  $\nabla$ . We let  $C^0(M; TM)$  denote the vector space of continuous vector fields on  $M$  (continuous sections of the tangent bundle  $TM$ ), and  $C^q(M; T^{*n}M)$  with  $0 \leq q \leq r$  the vector space of tensor fields  $\alpha$  of type  $(0, n)$  having continuous covariant derivatives  $\nabla^j \alpha \in C^{q-j}(M; T^{*n+j}M)$  up to order  $j = q$ . The Riemannian metric induces norms on these spaces defined by  $\|\Xi\|_{C^0(M; TM)} = \max_{x \in M} \|\Xi\|_x$ ,  $\|\alpha\|_{C^0(M; T^{*n}M)} = \max_{x \in M} \|\alpha\|_x$ , and  $\|\alpha\|_{C^q(M; T^{*n}M)} = \sum_{j=0}^q \|\nabla^j \alpha\|_{C^0(M; T^{*(q+j)}M)}$ , where  $\|\cdot\|_x$  denotes pointwise Riemannian norms on tensors. The case  $C^q(M; T^{*n}M)$  with  $n = 0$  corresponds to the  $C^q(M)$  spaces of functions. All of the  $C^0(M; TM)$  and  $C^q(M; T^{*n}M)$  spaces become Banach spaces with the norms defined above, and by compactness of  $M$ , the topology of these spaces is independent of the choice of Riemannian metric. Hereafter, we will use  $\iota^{(q)}$  to denote the canonical inclusion map of  $C^q(M)$  into  $L^2(\mu)$  ( $\mu$  being the invariant measure), and abbreviate  $\iota^{(0)} = \iota$  as in Section 2. We will also use  $\iota_{\mathcal{H}}$  to denote the inclusion map of an RKHS  $\mathcal{H}$  with a  $C^q$  reproducing kernel into  $C^q(M)$ . It follows from [37, Propositions 6.1 and 6.2] that the latter map is bounded.

The following result expresses how vector fields can be viewed as bounded operators on functions.

**Lemma 15.** *Let  $M$  be a compact,  $C^1$  manifold, equipped with a  $C^0$  Riemannian metric. Then, as an operator from  $C^1(M)$  to  $C^0(M)$ , every vector field  $\Xi \in C^0(M; TM)$  is bounded, with operator norm  $\|\Xi\|$  bounded above by  $\|\Xi\|_{C^0(M; TM)}$ .*

**Proof.** Denoting the gradient operator associated with the Riemannian metric on  $M$  by  $\text{grad}$ , the claim follows by an application of the Cauchy-Schwartz inequality for the Riemannian inner product, viz.

$$\begin{aligned} \|\Xi f\|_{C^0(M)} &= \|\Xi \cdot \text{grad} f\|_{C^0(M)} \leq \|\Xi\|_{C^0(M; TM)} \|\text{grad} f\|_{C^0(M; TM)} = \|\Xi\|_{C^0(M; TM)} \|\nabla f\|_{C^0(M; T^*M)} \\ &\leq \|\Xi\|_{C^0(M; TM)} \|f\|_{C^1(M)}. \quad \square \end{aligned}$$

In particular, under Assumption 1, the dynamical flow  $\Phi^t$  on  $M$  is generated by a vector field  $\vec{V} \in C^0(M; TM)$ , for which Lemma 15 applies. This vector field is related to the generator  $V$  by a conjugacy with  $\iota$  and  $\iota^{(1)}$ , namely,  $\iota \vec{V} = V \iota^{(1)}$ .

The following is a well known result from analysis [68].

**Lemma 16 (C<sup>1</sup> convergence theorem).** *Let  $M$  be a compact, connected,  $C^1$  manifold equipped with a  $C^0$  Riemannian metric. Let also  $f_j : M \rightarrow \mathbb{R}$  be a sequence of tensor fields in  $C^1(M; T^{*(n+1)}M)$ , such that the sequence  $\{\|\nabla f_j\|_{C^0(M; T^{*(n+1)}M)}\}_{j \in \mathbb{N}}$  is summable. Then, if there exists  $x \in M$  such that the series  $F_x := \sum_{j \in \mathbb{N}} f_j(x)$  converges in Riemannian norm, the series  $\sum_{j \in \mathbb{N}} f_j$  converges uniformly to a tensor field  $F \in C^1(M; T^{*(n+1)}M)$  such that  $F(x) = F_x$ .*

This lemma leads to the following  $C^r$  convergence result for functions, which will be useful for establishing the smoothness of kernels constructed as infinite sums of  $C^r$  eigenfunctions.

**Lemma 17.** *Let  $M$  be a compact, connected,  $C^r$  manifold with  $r \geq 1$ , equipped with a  $C^{r-1}$  Riemannian metric. Suppose that  $f_j : M \rightarrow \mathbb{R}$  is a sequence of real-valued  $C^r(M)$  functions such that (i) the sequence  $\{\|f_j\|_{C^r(M)}\}_{j \in \mathbb{N}}$  is summable; and (ii) there exists  $x \in M$  such that the series  $F_x = \sum_{j=0}^{\infty} f_j(x)$  converges. Then, the series  $\sum_{j=0}^{\infty} f_j$  converges absolutely and in  $C^r(M)$  norm to a  $C^r$  function  $F$ , such that  $F(x) = F_x$ .*

**Proof.** We will prove this lemma by induction over  $q \in \{1, \dots, r\}$ , invoking Lemma 16 as needed. First, note that summability of  $\{\|f_j\|_{C^r(M)}\}_{j \in \mathbb{N}}$  implies summability of  $\{\|\nabla^q f_j\|_{C^0(M; T^{*q} M)}\}_{j \in \mathbb{N}}$  for all  $q \in \{1, \dots, r\}$ . Because of this, and the fact that  $\sum_{j \in \mathbb{N}} f_j(x)$  converges, it follows from Lemma 16 that  $\sum_{j \in \mathbb{N}} f_j$  converges in  $C^1$  norm to some  $C^1$  function  $F$ . This establishes the base case for the induction ( $q = 1$ ). Now suppose that it has been shown that  $\sum_{j \in \mathbb{N}} f_j$  converges to  $F$  in  $C^q(M)$  norm for  $1 < q < r$ . In that case,  $\sum_{j \in \mathbb{N}} \nabla^q f_j(x)$  converges, and by summability of  $\{\|\nabla^{q+1} f_j\|_{C^0(M; T^{*(q+1)} M)}\}_{j \in \mathbb{N}}$ , it follows from Lemma 16 that  $\nabla^q F = \sum_{j \in \mathbb{N}} \nabla^q f_j$  converges in  $C^1(M; T^{*q} M)$  norm. Thus,  $\nabla^{q+1} F = \sum_{j \in \mathbb{N}} \nabla^{q+1} f_j$  converges in  $C^0(M; T^{*(q+1)} M)$  norm, which in turn implies that  $\sum_{j \in \mathbb{N}} f_j$  converges to  $F$  in  $C^{q+1}(M)$  norm, and the lemma is proved by induction.  $\square$

## 5. Proof of Theorems 5–8

**Proof of Theorem 5.** By Assumption 3,  $\mathcal{H}$  is a subspace of  $C^1(M)$ , and therefore for every  $f \in \mathcal{H}$ ,  $K^* f = \iota^{(1)} f$ . Claim (i) then follows from the facts that  $\text{ran } \iota^{(1)} \subset D(V)$ , and  $K$  is bounded. To prove Claim (ii), let  $K' : L^2(\mu) \rightarrow C^0(M)$  be the kernel integral operator associated with the continuous kernel  $k'$ , and  $\iota$  the  $C^0(M) \rightarrow L^2(\mu)$  inclusion map. Because  $\iota K'$  is a Hilbert-Schmidt integral operator on  $L^2(\mu)$ , with operator norm bounded above by its Hilbert-Schmidt norm,  $\|\iota K'\| \leq \|k'\|_{L^2(\mu \times \mu)} \leq \|k'\|_{C^0(X \times X)}$ , the claim will follow if it can be shown that  $\iota K' = VG$ . To that end, note that for every  $f \in L^2(\mu)$  and  $x \in M$  we have  $K' f(x) = \langle k'(x, \cdot), f \rangle_{\mu}$ . Thus, using the  $C^0(M)$  limit  $k'(x, \cdot) = \lim_{t \rightarrow 0} g_t$ , where  $g_t = (k(\Phi^t(x), \cdot) - k(x, \cdot))/t$ , and continuity of inner products, we obtain

$$K' f(x) = \langle k'(x, \cdot), f \rangle_{\mu} = \langle \lim_{t \rightarrow 0} g_t, f \rangle_{\mu} = \lim_{t \rightarrow 0} \langle g_t, f \rangle_{\mu} = \lim_{t \rightarrow 0} \frac{1}{t} [\langle k(\Phi^t(x), \cdot), f \rangle_{\mu} - \langle k(x, \cdot), f \rangle_{\mu}] = \vec{V} K f(x).$$

As a result, because  $\text{ran } K \subset C^1(M)$ , for any  $f \in L^2(\mu)$  it follows that

$$\iota K' f = \iota \vec{V} K f = V \iota^{(1)} K f = V K^* K f = V G f,$$

proving Claim (ii). Finally, to prove Claim (iii), we have by definition of the adjoint,

$$D((GV)^*) := \{f \in L^2(\mu) : \exists h \in L^2(\mu) \text{ such that } \forall g \in D(V), \langle f, GVg \rangle_{\mu} = \langle h, g \rangle_{\mu}\}, \quad (GV)^* f := h,$$

where  $h$  is unique by the Riesz representation theorem and density of  $D(V)$  in  $L^2(\mu)$ . We will now use this definition to show that  $(GV)^* = -VG = -A$ . Indeed, for every  $f \in D(A) = L^2(\mu)$  and every  $g \in D(V)$ , setting  $h = -Af$ , we obtain

$$\langle h, g \rangle_{\mu} = \langle -Af, g \rangle_{\mu} = -\langle VGf, g \rangle_{\mu} = \langle Gf, Vg \rangle_{\mu} = \langle f, GVg \rangle_{\mu}.$$

This satisfies the definition of  $(GV)^*$ , proving the claim and the theorem.  $\square$

**Proof of Theorem 6.** We begin with the proof of Claim (i). The inclusion  $\text{ran } K^* \subset D(V)$  holds because  $\mathcal{H}$  is a subspace of  $C^1$ . To prove that  $V K^*$  is bounded, note that by Lemma 15, and the fact that the inclusion map  $\iota_{\mathcal{H}} : \mathcal{H} \rightarrow C^1(M)$  is bounded,

$$\|VK^*f\|_{L^2(\mu)} = \|\iota\vec{V}f\|_{L^2(\mu)} \leq \|\vec{V}f\|_{C^0(M)} \leq \|\vec{V}\|\|f\|_{C^1(M)} \leq \|\iota_{\mathcal{H}}\|\|\vec{V}\|\|f\|_{\mathcal{H}},$$

proving that  $VK^*$  is bounded and completing the proof of Claim (i). Turning to Claim (ii), that  $W$  is compact follows from the fact that it is a composition of a compact operator,  $K$ , by a bounded operator,  $VK^*$ . Moreover,  $W$  is skew-symmetric by skew-adjointness of  $V$ , and thus skew-adjoint because it is bounded.  $W$  is also real because  $K$  and  $V$  are real operators. It thus remains to verify the integral formula for  $Wf$  stated in the theorem. For that, it follows from the Leibniz rule for vector fields and the fact that  $k$  lies in  $C^1(M \times M)$  that for every  $f \in C^1(M)$  and  $x \in X$ ,

$$k(x, \cdot)\vec{V}f = \vec{V}(k(x, \cdot)f) - (\vec{V}k(x, \cdot))f = \vec{V}(k(x, \cdot)f) + \tilde{k}'(x, \cdot)f.$$

Using this result, and the fact that  $\int_M \vec{V}(k(x, \cdot)f) d\mu = \langle 1_M, V(k(x, \cdot)f) \rangle_{\mu}$  vanishes by skew-adjointness of  $V$ , we obtain

$$KVK^*f(x) = KV\iota^{(1)}f(x) = K\vec{V}f(x) = \int_M k(x, \cdot)\vec{V}f d\mu = \int_M \tilde{k}'(x, \cdot)f d\mu. \quad \square$$

**Proof of Theorem 7.** That  $B = -A^*$  is a Hilbert-Schmidt integral operator with kernel  $\tilde{k}'$  follows from standard properties of integral operators. Next, to prove Claim (i), note that  $GV$  is bounded as it has a bounded adjoint,  $(GV)^* = -A$ , by Theorem 5, and therefore has a unique closed extension  $\overline{GV} : L^2(\mu) \rightarrow L^2(\mu)$  equal to  $(GV)^{**}$ . In order to verify that  $\overline{GV} = B$ , it suffices to show that  $GVf = Bf$  for all  $f$  in any dense subspace of  $D(V)$ ; in particular, we can choose the subspace  $\iota^{(1)}C^1(M)$ . For any observable  $\iota^{(1)}f$  in this subspace, we have  $Bf = \iota\tilde{K}'f$  and  $GVf = \iota^{(1)}K\vec{V}f$ , where  $\tilde{K}' : L^2(\mu) \rightarrow C^0(M)$  is the integral operator with kernel  $\tilde{K}'$ , defined analogously to the operator  $K'$  in the proof of Theorem 5. Employing the Leibniz rule as in the proof of Theorem 6, it is straightforward to verify that  $Bf$  is indeed equal to  $GVf$ , proving that  $B$  is the unique closed extension of  $GV$ . Next, to show that  $B$  is also an extension of  $K^*W\mathcal{N}$ , it suffices to show that  $GV \supseteq K^*W\mathcal{N}$ . For that, note that  $K^*W\mathcal{N}$  is a well-defined operator by Theorem 6, and thus, substituting the definition of  $W$  in (13), and using the fact that  $K^*\mathcal{N}$  is the identity on  $D(\mathcal{N})$ , we obtain

$$K^*W\mathcal{N} = GV\iota^{(1)}\mathcal{N} = GV|_{D(\mathcal{N})}.$$

This shows that  $K^*W\mathcal{N} \subseteq GV \subset B$ , confirming that  $B$  is a closed extension of  $K^*W\mathcal{N}$ . If  $k$  is strictly positive, then  $D(\mathcal{N})$  is dense, and  $B$  is the unique closed extension of  $K^*W\mathcal{N}$ . This completes the proof of Claim (i).

Next, to prove Claim (ii), note that because  $B$  is bounded, the Taylor series  $e^{tB} = \sum_{n=0}^{\infty} (tB)^n/n!$  converges in operator norm for every  $t \in \mathbb{R}$ , and the set  $\{e^{tB}\}_{t \in \mathbb{R}}$  clearly forms a group under composition of operators. This group is norm-continuous by boundedness of  $B$ . Similarly, we have  $e^{tW} = \sum_{n=0}^{\infty} (tW)^n/n!$  in operator norm, and observing that for every  $n \in \mathbb{N}$ ,  $K^*W^n = B^nK^*$ , we arrive at the claimed identity,

$$K^*e^{tW} = \sum_{n=0}^{\infty} \frac{1}{n!} t^n K^*W^n = \sum_{n=0}^{\infty} \frac{1}{n!} t^n B^n K^* = e^{tB} K^*.$$

The identity  $K^*e^{tW}\mathcal{N} = e^{tB}|_{D(\mathcal{N})}$  then follows from the fact that  $K^*\mathcal{N}$  is the identity on  $D(\mathcal{N})$ .  $\square$

**Proof of Theorem 8.** Let  $\{\phi_j\}_{j=0}^{\infty}$  be an orthonormal basis of  $L^2(\mu)$  consisting of eigenfunctions  $\phi_j$  of  $G$  corresponding to eigenvalues  $\lambda_j$  ordered in decreasing order. Let also  $\{\psi_j\}_{j=0}^{\infty}$  be an orthonormal basis of  $\mathcal{H}$ , whose first  $J$  elements are given by (17) (with some abuse of notation as  $J$  may be infinite). Recall from Section 4.1 that  $\mathcal{U}\phi_j = \psi_j$ . To prove the theorem, it suffices to show that  $G^{1/2}VG^{1/2}$  is well-defined on a

dense subspace of  $L^2(\mu)$ , and on that subspace,  $G^{1/2}VG^{1/2}$  and  $\mathcal{U}^*W\mathcal{U}$  are equal. To verify that  $G^{1/2}VG^{1/2}$  is densely defined, note first that  $G^{1/2}\phi_j$  trivially vanishes for  $j \notin J$ , and therefore  $G^{1/2}VG^{1/2}\phi_j$  is well-defined and vanishes too. Moreover if  $j \in J$ ,  $G^{1/2}\phi_j = K^*\psi_j$ , and  $G^{1/2}VG^{1/2}\phi_j$  is again well defined since  $\text{ran } K^* \subset D(V)$ . As a result, the domain of  $G^{1/2}VG^{1/2}$  contains all linear combinations of  $\phi_j$  with  $j \notin J$ , and all finite combinations with  $j \in J$ , and is therefore a dense subspace of  $L^2(\mu)$ . Next, to show that  $\mathcal{U}^*W\mathcal{U}$  and  $G^{1/2}VG^{1/2}$  are equal on this subspace, it suffices to show that they have the same matrix elements in the  $\{\phi_j\}$  basis of  $L^2(\mu)$ , i.e., that  $\langle \phi_i, G^{1/2}VG^{1/2}\phi_j \rangle_\mu$  is equal to  $\langle \phi_i, \mathcal{U}^*W\mathcal{U}\phi_j \rangle_{\mathcal{H}}$  for all  $i, j \in \mathbb{N}_0$ . Indeed, because  $\ker \mathcal{U} = \ker G^{1/2} = (\text{ran } G^{1/2})^\perp$ , both  $\mathcal{U}\phi_j$  and  $G^{1/2}\phi_j$  vanish when  $j \notin J$ . We therefore deduce that if either of  $i$  and  $j$  does not lie in  $J$ , the matrix elements  $\langle \phi_i, \mathcal{U}^*W\mathcal{U}\phi_j \rangle_\mu$  and  $\langle \phi_j, G^{1/2}VG^{1/2}\phi_i \rangle_\mu$  both vanish. On the other hand, if  $i, j \in J$ , we have

$$\begin{aligned} \langle \phi_i, \mathcal{U}^*W\mathcal{U}\phi_j \rangle_\mu &= \langle \psi_i, W\psi_j \rangle_{\mathcal{H}} = \langle K^*\psi_i, VK^*\psi_j \rangle_\mu = \langle \lambda_i^{-1/2}K^*K\phi_i, \lambda_j^{-1/2}K^*K\phi_j \rangle_\mu \\ &= \langle G^{1/2}\phi_i, VG^{1/2}\phi_j \rangle_\mu = \langle \phi_i, G^{1/2}VG^{1/2}\phi_j \rangle_\mu. \end{aligned}$$

We have thus shown that  $\mathcal{U}^*W\mathcal{U}$  and  $G^{1/2}VG^{1/2}$  are equal on a dense subspace of  $L^2(\mu)$ , and because the former operator is bounded and defined on the whole of  $L^2(\mu)$ , it follows that  $\tilde{V} = \mathcal{U}^*W\mathcal{U}$  is the unique closed extension of  $G^{1/2}VG^{1/2}$ . That  $\tilde{V}$  is skew-adjoint, Hilbert-Schmidt, and real follows immediately.  $\square$

## 6. Proof of Theorems 9 and 10

We will need the following lemma, describing how to convert between eigenfunctions of  $A$ ,  $B$ ,  $\tilde{V}$ , and  $\tilde{W}$ . The proof will be omitted since it follows directly from the definitions of these operators.

**Lemma 18.** *Let Assumptions 1 and 3 hold with  $r = 1$ . Then,*

- (i) *If  $\zeta \in \mathcal{K}$  is an eigenfunction of  $\tilde{W}$  at eigenvalue  $i\omega$ , then  $K^*\zeta$  is an eigenfunction of  $B$  at eigenvalue  $i\omega$ .*
- (ii)  *$z' \in L^2(\mu)$  is an eigenfunction of  $A$  at eigenvalue  $i\omega$  iff  $Kz'$  is an eigenfunction of  $\tilde{W}$  at eigenvalue  $i\omega$ .*
- (iii) *If  $z'$  is an eigenfunction of  $A$  at eigenvalue  $i\omega$ , then  $G^{1/2}z'$  is an eigenfunction of  $\tilde{V}$  at eigenvalue  $i\omega$ .*
- (iv) *If  $\tilde{z} \in L^2(\mu)$  is an eigenfunction of  $\tilde{V}$  at eigenvalue  $i\omega$ , then  $G^{1/2}\tilde{z}$  is an eigenfunction of  $B$  at eigenvalue  $i\omega$ .*

**Proof of Theorem 9.** Starting from Claim (i), let  $\tilde{W}$  be the restriction of  $W$  onto the closed subspace  $\mathcal{K} \subseteq \mathcal{H}$ . Since  $\mathcal{K}$  is invariant under  $W$ , and  $\ker W \supseteq \mathcal{K}^\perp$  by definition, we have  $\sigma_p(W) = \sigma_p(\tilde{W})$  if  $\mathcal{K}^\perp = \{0\}$  (i.e.,  $K$  has dense range), and  $\sigma_p(W) = \sigma_p(\tilde{W}) \cup \{0\} = \sigma_p(\tilde{W})$  otherwise. Thus, to prove the claim, it is enough to show that  $\sigma_p(A) = \sigma_p(B) = \sigma_p(\tilde{V}) = \sigma_p(\tilde{W})$ , including eigenvalue multiplicities. To that end, note first that  $\tilde{W}$  and  $\tilde{V}$  are unitarily equivalent by Theorem 8 and strict  $L^2(\mu)$ -positivity of  $k$ , and thus  $\sigma_p(\tilde{W}) = \sigma_p(\tilde{V})$ , including multiplicities. Moreover, by Lemma 18,  $\sigma_p(A) \subseteq \sigma_p(\tilde{W}) \subset i\mathbb{R}$ , and because  $A$  is a real operator, it follows that  $\sigma_p(A)$  is symmetric about the origin of the imaginary line  $i\mathbb{R}$ , so that

$$\sigma_p(A) = -\sigma_p(A) = -\sigma_p(A)^* = -\sigma_p(A^*) = -\sigma_p(-B) = \sigma_p(B).$$

Thus, the equality of  $\sigma_p(A)$ ,  $\sigma_p(B)$ ,  $\sigma_p(\tilde{V})$ , and  $\sigma_p(\tilde{W})$  will follow if it can be shown that  $\sigma_p(A) = \sigma_p(\tilde{V})$ . Indeed, it follows from Lemmas 18(iii) and 18(iv) that  $\sigma_p(A) \subseteq \sigma_p(\tilde{V})$  and  $\sigma_p(\tilde{V}) \subseteq \sigma_p(B)$ , respectively. These relationships, together with the fact that  $\sigma_p(A) = \sigma_p(B)$ , imply that  $\sigma_p(A) = \sigma_p(\tilde{V})$ , and thus  $\sigma(A) = \sigma(B) = \sigma(\tilde{V}) = \sigma(\tilde{W})$ , as claimed. The equality of the multiplicities of the eigenvalues of  $A$ ,  $B$ , and  $\tilde{V}$  follows from the facts that  $K$  and  $G^{1/2}$  are injective operators. This completes the proof of Claim (i).

To prove Claim (ii), note that because  $k$  is  $L^2(\mu)$ -Markov ergodic,  $Gf = f$  implies that  $f$  is  $\mu$ -a.e. constant. In addition, by ergodicity of the flow  $\Phi^t$ ,  $Vf = 0$  implies again that  $f$  is  $\mu$ -a.e. constant. It then follows that

$$Af = 0 \implies V(Gf) = 0 \implies Gf = \mu\text{-a.e. constant} \implies f = \mu\text{-a.e. constant.}$$

This shows that 0 is a simple eigenvalue of  $A$  with constant corresponding eigenfunctions. Therefore, since  $\sigma_p(A) = \sigma_p(B) = \sigma_p(\tilde{V}) = \sigma_p(\tilde{W})$ , including multiplicities, 0 is also a simple eigenvalue of  $B$ ,  $\tilde{V}$ , and  $\tilde{W}$ , and the constancy of the corresponding eigenfunctions follows directly from the definition of these operators.

Next, to prove Claims (iii) and (iv), fix a nonzero eigenvalue  $i\omega_j$  of  $A$ . By compactness of this operator, the corresponding eigenspace is finite-dimensional, and thus the injective operator  $G^{1/2}$  maps every basis of this eigenspace to a linearly independent set. By Lemma 18(iii) and Claim (i), this set is actually a basis of the eigenspace of  $\tilde{V}$  at eigenvalue  $i\omega_j$ . As result, every eigenfunction of  $\tilde{V}$  at nonzero corresponding eigenvalue lies in the range of  $G^{1/2}$ . Moreover, it follows from Claim (ii) that every eigenfunction of  $\tilde{V}$  at eigenvalue 0 is constant, and thus also lies in the range of  $G^{1/2}$ . We therefore conclude that every eigenfunction of  $\tilde{V}$  lies in the range of  $G^{1/2}$ , and thus in the domain of  $G^{-1/2}$ , as claimed. The fact that  $\tilde{z}_j \in \text{ran } G^{1/2}$  for all  $j \in \mathbb{N}_0$  also implies that  $z'_j$  is an eigenfunction of  $A$  at eigenvalue  $i\omega_j$ , since

$$\tilde{V}\tilde{z}_j = \tilde{V}G^{1/2}G^{-1/2}\tilde{z}_j = G^{1/2}VGG^{-1/2}\tilde{z}_j = G^{1/2}Az'_j.$$

In addition, we can deduce directly from Lemma 18(iv) that each of the  $z_j$  are eigenfunctions of  $B$  at eigenvalue  $i\omega_j$ , as stated in Claim (iv).

To complete the proof of Claims (iii) and (iv), it remains to show that  $\{z'_0, z'_1, \dots\}$  and  $\{z_0, z_1, \dots\}$  form unconditional Schauder bases of  $L^2(\mu)$ . For that, note first that  $z_j$  is a dual sequence to the  $z'_j$ , i.e.,

$$\langle z'_j, z_l \rangle_\mu = \langle G^{-1/2}\tilde{z}_j, G^{1/2}\tilde{z}_l \rangle_\mu = \langle \tilde{z}_j, \tilde{z}_l \rangle_\mu = \delta_{jl}.$$

As a result, since every Schauder basis has a unique dual sequence, which is also a Schauder basis [69], Claims (iii) and (iv) will be proved if it can be shown that  $\{z_0, z_1, \dots\}$  is an unconditional Schauder basis. To verify that this is indeed the case, fix  $\{\phi_0, \phi_1, \dots\}$  from (17) as an orthonormal basis of  $L^2(\mu)$  (corresponding to the eigenvalues  $\lambda_0, \lambda_1, \dots$ ), and  $\{e_0, e_1, \dots\}$  as the standard orthonormal basis of  $\ell^2$ , and define the unbounded operator  $Z' : D(Z') \subset \ell^2 \rightarrow \ell^2$ , the bounded operator  $L : \ell^2 \rightarrow \ell^2$ , the unitary operator  $U : \ell^2 \rightarrow \ell^2$ , and the diagonal operator  $\Lambda : \ell^2 \rightarrow \ell^2$  such that

$$\langle e_i, Z'e_j \rangle_{\ell^2} = \langle \phi_i, z'_j \rangle_\mu, \quad \langle e_i, Le_j \rangle_{\ell^2} := \langle z_i, \phi_j \rangle_\mu, \quad \langle e_i, Ue_j \rangle_{\ell^2} = \langle \phi_i, \tilde{z}_j \rangle_\mu, \quad \langle e_i, \Lambda e_j \rangle_{\ell^2} = \lambda_i \delta_{ij}.$$

Here,  $D(Z')$  is defined as the dense subspace of  $\ell^2$  whose elements  $\sum_{j=0}^{\infty} c_j e_j$  satisfy  $\sum_{i,j=0}^{\infty} |\langle \phi_i, z'_j \rangle_\mu c_j|^2 < \infty$ . Note that  $Z'^*$  and  $L$  are the matrix representations of the mappings  $\phi_j \mapsto z'_j$ , and  $\phi_j \mapsto z_j$ , respectively, in the orthonormal basis  $\{\phi_0, \phi_1, \dots\}$ . With these definitions, the  $\mathbb{N} \times \mathbb{N}$  matrices with elements  $\langle e_i, Z'e_j \rangle_{\ell^2}$  and  $\langle e_i, Le_j \rangle_{\ell^2}$ , which represent  $Z'$  and  $L$ , respectively, have  $\ell^2$  summable columns and rows respectively.

Next, note that  $L$  is a left inverse of  $Z'$ , as can be verified by computing

$$\begin{aligned} \langle e_i, LZ'e_j \rangle_\mu &= \sum_{j=0}^{\infty} \langle z_i, \phi_j \rangle_\mu \langle \phi_j, z'_l \rangle_\mu = \sum_{j=0}^{\infty} \langle z_i, \phi_j \rangle_\mu \langle \phi_j, \phi_l \rangle_\mu \langle \phi_l, z'_l \rangle_\mu = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \langle z_i, \phi_j \rangle_\mu \langle \phi_j, \phi_k \rangle_\mu \langle \phi_k, z'_l \rangle_\mu \\ &= \left\langle \sum_{j=0}^{\infty} \langle \phi_j, z_i \rangle_\mu \phi_j, \sum_{k=0}^{\infty} \langle \phi_j, z'_l \rangle_\mu \phi_k \right\rangle_\mu = \langle z_i, z'_l \rangle_\mu = \delta_{il}. \end{aligned} \tag{24}$$

Similarly, one can verify the identities  $L = U^* \Lambda^{1/2}$  and  $Z' = \Lambda^{-1/2} U$ . Using these results, and defining  $\Pi_l : \ell^2 \rightarrow \ell^2$  as the canonical orthogonal projection onto  $\text{span}\{e_0, \dots, e_{l-1}\}$ , we obtain

$$Z' \Pi_l L = \Lambda^{-1/2} U \Pi_l U^* \Lambda^{1/2} = \Lambda^{-1/2} \Pi_l \Lambda^{1/2} = \Pi_l, \quad Z' \Pi_l L \xrightarrow[l \rightarrow \infty]{s} \text{Id}. \quad (25)$$

By [70, Lemma 2.1], (24) and (25) imply that the columns of the matrix representation of  $Z'$ , i.e., the eigenfunctions  $z'_j$ , form a Schauder basis of  $L^2(\mu)$ . The unconditionality of this basis follows from the fact that if the  $z_j$  are permuted, (25) still holds, but with the rows and columns of the matrix representations of  $U$ ,  $Z'$ ,  $L$ , and  $\Lambda$  correspondingly permuted. This completes the proof of Claims (iii) and (iv).

In Claim (v), the fact that the  $\zeta_j$  are eigenfunctions of  $\tilde{W}$  follows from Lemma 18 (ii). We also have

$$\zeta_j = K G^{-1/2} \tilde{z}_j = \mathcal{U} \tilde{z}_j,$$

and because  $\mathcal{U}$  acts as a unitary operator from  $L^2(\mu)$  to  $\mathcal{K}$ , the fact that  $\{\tilde{z}_0, \tilde{z}_1, \dots\}$  is an orthonormal basis of  $L^2(\mu)$  implies that  $\{\zeta_0, \zeta_1, \dots\}$  is an orthonormal basis of  $\mathcal{K}$ , proving the claim.

Finally, in Claim (vi), note first that all of the summations are well defined and independent of ordering due to the unconditionality of all the bases involved. The results for  $\tilde{V}$  and  $\tilde{W}$  follow from standard properties of Hilbert-Schmidt, skew-adjoint operators. Here, we will only verify the representation of  $B$ , as the case for  $A$ , is analogous. By Claim (iv), every  $f \in L^2(\mu)$  has a unique expansion  $f = \sum_{j=0}^{\infty} a_j z_j$ , with the summation holding in  $L^2(\mu)$  sense. Then, since  $Bz_j = i\omega_j z_j$  and  $B$  is bounded,

$$Bf = B \sum_{j=0}^{\infty} a_j z_j = \sum_{j=0}^{\infty} a_j Bz_j = \sum_{j=0}^{\infty} a_j i\omega_j z_j.$$

The fact that  $Bf = \sum_{j=0}^{\infty} \langle z'_j, f \rangle_{\mu} i\omega_j z_j$  then follows from the identity below for the coefficients  $a_j$ :

$$\langle z'_j, f \rangle_{\mu} = \left\langle z'_j, \sum_{k=0}^{\infty} a_k z_k \right\rangle_{\mu} = \sum_{k=0}^{\infty} a_k \langle z'_j, z_k \rangle_{\mu} = \sum_{k=0}^{\infty} a_k \delta_{jk} = a_j.$$

This completes the proof of the claim and Theorem 9.  $\square$

**Proof of Theorem 10.** It follows from the strong convergence  $G_{\tau} \xrightarrow{s} \text{Id}$  in Assumption 4 that

$$\lim_{\tau \rightarrow 0^+} \|(B_{\tau} - V)f\|_{L^2(\mu)} = \lim_{\tau \rightarrow 0^+} \|(G_{\tau}V - V)f\|_{L^2(\mu)} = \lim_{\tau \rightarrow 0^+} \|(G_{\tau} - \text{Id})Vf\|_{L^2(\mu)} = 0, \quad \forall f \in D(V^2),$$

proving Claim (i). Turning to Claim (ii), it follows from Lemma 14(i) that  $D(V^2)$  is a core for  $V$ , and thus by Assumption 4 and Lemma 14(ii) that, as  $\tau \rightarrow 0^+$ ,  $\tilde{V}_{\tau}$  converges to  $V$  in strong resolvent sense. The strong convergence of  $Z(\tilde{V}_{\tau})$  to  $Z(V)$  then follows by Proposition 13(i). The result for  $\mathcal{U}_{\tau}^* Z(W_{\tau}) \mathcal{U}_{\tau}$  follows from the fact that this operator is equal to  $Z(\tilde{V}_{\tau})$ , by (15).

To prove Claim (iii) note first that, by standard properties of the Borel functional calculus,  $Z(\tilde{V}_{\tau})$  is a uniformly bounded family of operators with  $\|Z(\tilde{V}_{\tau})\| \leq \|Z\|_{C^0(i\mathbb{R})}$ . As a result, it follows from the uniform boundedness principle that  $Z(\tilde{V}_{\tau})G_{\tau}^{1/2} \xrightarrow{s} Z(V)$ , as  $\tau \rightarrow 0^+$ . Similarly,  $G_{\tau}^{1/2}$  is uniformly bounded, so the strong  $\tau \rightarrow 0^+$  limit of  $G_{\tau}^{1/2} Z(A_{\tau})$  is equal to the strong  $\tau \rightarrow 0^+$  limit of  $Z(A_{\tau})$ . However,  $G_{\tau}^{1/2} Z(A_{\tau}) = Z(\tilde{V}_{\tau})G_{\tau}^{1/2}$  by (16), and we conclude that  $Z(A_{\tau}) \xrightarrow{s} Z(V)$ , as claimed. That  $Z(B_{\tau}) \xrightarrow{s} Z(V)$  then follows immediately from the fact that  $B_{\tau} = -A_{\tau}^*$ . The latter result leads in turn to the strong convergence  $K_{\tau}^* Z(W_{\tau}) \mathcal{N}_{\tau} \xrightarrow{s} Z(V)$  on  $H_{\infty}$ , since, by Theorem 7(ii) and complex analyticity of  $Z$ ,  $K_{\tau}^* Z(W_{\tau}) \mathcal{N}_{\tau}$  and  $Z(B_{\tau})$  are equal operators on  $H_{\infty}$ .

Next, the strong convergence of  $\tilde{E}_{\tau}(\Omega)$  to  $E(\Omega)$  in Claim (iv) follows from Proposition 13(iii). Equation (15) then leads to the result for  $\mathcal{U}_{\tau}^* \mathcal{E}_{\tau}(\Omega) \mathcal{U}_{\tau}$ . Finally, Claim (v) follows from Proposition 13(vi).  $\square$

## 7. Proof of Theorems 1, 2 and Corollaries 3, 4

**Proof of Theorem 1.** First, note that by Lemma 11(ii), the sequence  $\{\lambda_j\}_{j=0}^\infty$  is summable. Moreover, since  $\lambda_j \leq 1$ ,  $\{\lambda_j^q\}_{j=0}^\infty$  is summable for every  $q \geq 1$ . Now define  $r_{\tau,j} = (\lambda_{\tau,j}/\lambda_j)^{1/2}$ . Due to the exponential decay of the  $\lambda_{\tau,j}$  in (10), the sequences  $\{r_{\tau,j}^q\}_{j=0}^\infty$  and  $\{r_{\tau,j}^q/\lambda_j\}_{j=0}^\infty$  are summable for every  $q \geq 1$  and  $\tau > 0$ . Observe now that  $\psi_{\tau,j}$  and  $p_\tau$  can be expressed as

$$\psi_{\tau,j} = r_{\tau,j}\psi_j = r_{\tau,j}\lambda_{\tau,j}^{-1/2}P\phi_j, \quad p_\tau(x, y) = \sum_{j=0}^\infty r_{\tau,j}^2\psi_j(x)\psi_j(y). \quad (26)$$

It therefore follows from the summability of  $\{r_{\tau,j}^2\}_{j=0}^\infty$  that for every  $\tau > 0$ , the series for  $p_\tau(x, y)$  also converges absolutely and uniformly on  $X_\nu \times X_\nu$ , and condition (ii) of Lemma 17 is satisfied. Next, observe that for every  $j \in \mathbb{N}_0$  and  $\alpha \in \{1, \dots, r\}$ ,

$$\psi_{\tau,j} = r_{\tau,j}\lambda_j^{-1/2} \int_M p(\cdot, y)\phi_j(y) d\nu(y), \quad \nabla^\alpha \psi_{\tau,j} = r_{\tau,j}\lambda_j^{-1/2} \int_M \nabla^\alpha p(\cdot, y)\phi_j(y) d\nu(y),$$

and thus  $\|\psi_{\tau,j}\|_{C^r(M)} \leq r_{\tau,j}\lambda_j^{-1/2}\|p\|_{C^r(M \times M)}$ . Let now  $\nabla_1 f$  and  $\nabla_2 f$  denote the covariant derivatives of  $f \in C^1(M \times M)$  with respect to the first and second variables, respectively. Defining  $f_j(x, y) = \psi_{\tau,j}(x)\psi_{\tau,j}(y)$ , and noting that  $f_j$  is a  $C^r(M \times M)$  function by  $C^r$  regularity of  $p$  (and thus  $\psi_j$ ), we have

$$\|f_j\|_{C^r(M)} = \sum_{\substack{\alpha, \beta \in \{0, \dots, r\}, \\ \alpha + \beta = m}} \left\| \left( \nabla_1^\alpha \nabla_2^\beta \right) f_j(x, y) \right\|_{C^0(M; T^{*(\alpha+\beta)} M)} \leq C r_{\tau,j}^2 / \lambda_j,$$

where,  $C$  is a constant equal to a multiple of  $\|p\|_{C^r(M \times M)}^2$ . This bound implies that  $\{\|f_j\|_{C^r(M)}\}_{j=0}^\infty \in \ell^1$ , and condition (i) of Lemma 17 is satisfied. We therefore conclude that Lemma 17 applies, and as a result, for every  $x, y \in M$ ,  $\sum_{j=0}^\infty f_j(x, y) = \sum_{j=0}^\infty \psi_{\tau,j}(x)\psi_{\tau,j}(y)$  converges in  $C^r(M \times M)$  norm to a  $C^r(M \times M)$  function,  $p_\tau$ , as claimed.

Next, we begin our proof of Claim (i) by showing that  $p_\tau$  is the reproducing kernel for an RKHS. Fixing  $\tau > 0$ , we start from the pre-Hilbert space  $H_0 = \text{span}\{\psi_{\tau,j}\}$ , equipped with the inner product

$$\left\langle \sum_{i=0}^{m-1} a_i \psi_{\tau,j}, \sum_{j=0}^{n-1} b_j \psi_{\tau,j} \right\rangle_{H_0} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} a_i^* \delta_{ij} b_j.$$

By (26), for every  $f = \sum_{j=0}^{n-1} c_j \psi_{\tau,j} \in H_0$ , we have

$$\|f\|_{\mathcal{H}}^2 = \left\| \sum_{j=0}^{n-1} c_j r_{\tau,j} \psi_j \right\|_{\mathcal{H}}^2 = \sum_{j=0}^{n-1} |r_{\tau,j}|^2 |c_j|^2 \leq C \sum_{j=0}^{n-1} |c_j|^2 = C \|f\|_{H_0}^2,$$

where  $C = \max_{j \in \mathbb{N}_0} |r_{\tau,j}|^2$ . This implies that every Cauchy sequence in  $H_0$  is a Cauchy sequence in  $\mathcal{H}$ , and as a result the Hilbert space completion of  $H_0$ , denoted  $H$ , can be identified with a subspace of  $\mathcal{H}$ . In particular,  $H$  is a Hilbert space of functions on  $M$  with an orthonormal basis  $\{\psi_{\tau,j}\}_{j=0}^\infty$ . We will next show that  $H$  is an RKHS with reproducing kernel  $p_\tau$  by showing that, for every  $x \in M$ , the kernel sections  $p_\tau(x, \cdot)$  lie in  $H$ , and function evaluation at  $x$  is a bounded linear functional on  $H$  equal to an inner product with these sections. Indeed, since  $p$  is the reproducing kernel for  $\mathcal{H}$ , for every  $x \in M$ , the section  $p(x, \cdot)$  lies in  $\mathcal{H}$ , and thus, by the Mercer representation for  $p$ ,  $\sum_{j=0}^\infty |\psi_j(x)|^2 < \infty$ . It therefore follows that

$$\sum_{j=0}^{\infty} |\psi_{\tau,j}(x)|^2 = \sum_{j=0}^{\infty} r_{\tau,j}^2 |\psi_j(x)|^2 < \infty,$$

and because  $\{\psi_{\tau,j}\}_{j=0}^{\infty}$  is an orthonormal basis of  $H$ ,  $p_{\tau}(x, \cdot) = \sum_{j=0}^{\infty} \psi_{\tau,j}(x) \psi_{\tau,j}$  lies in  $H$ . Moreover, for every  $x \in M$  and  $f \in H$ ,

$$f(x) = \sum_{j=0}^{\infty} \langle \psi_{\tau,j}, f \rangle_H \psi_{\tau,j}(x) = \left\langle \sum_{j=0}^{\infty} \psi_{\tau,j}(x) \psi_{\tau,j}, f \right\rangle_H = \langle p_{\tau}(x, \cdot), f \rangle_H,$$

which shows that pointwise evaluation on  $H$  is given by inner products with the kernel sections  $p_{\tau}$ . We therefore conclude that  $H$  is an RKHS, denoted  $\mathcal{H}_{\tau}$ , with  $p_{\tau}$  as its  $C^r$  reproducing kernel. As a result,  $p_{\tau}$  is positive-definite, and it induces integral operators  $P_{\tau} : L^2(\nu) \rightarrow \mathcal{H}_{\tau}$  and  $G_{\tau} = P_{\tau}^* P_{\tau}$ . It also follows from the Mercer representation for  $p_{\tau}$  that  $G_{\tau}$  is a strictly positive, compact operator with the same eigenfunctions  $\phi_j$  as  $G$ , corresponding to the eigenvalues  $0 < \lambda_{\tau,j} \leq 1$ , where  $\lambda_{\tau,0} = 1$  is simple. What remains to prove Claim (i) is to show that  $G_{\tau}$  is  $L^2(\nu)$ -Markov ergodic. We will verify this assertion following the proof of Claim (iii).

Turning to Claim (ii), note that for every  $j \in \mathbb{N}_0$ , the function  $\psi_j/\lambda_j^{1/2}$  equals  $\psi_{\tau,j}/\lambda_{\tau,j}$ , and thus lies in  $\mathcal{H}_{\tau}$ . Moreover, this function lies in the same  $L^2(\nu)$  equivalence class as  $\phi_j$ , and because the  $\phi_j$  form an orthonormal basis of  $L^2(\nu)$ , it follows that  $\mathcal{H}_{\tau}$  is dense in  $L^2(\nu)$ . To verify the claimed inclusion relationships between  $\mathcal{H}$  and  $\mathcal{H}_{\tau}$ , we use (26) to characterize these spaces as

$$\mathcal{H} = \left\{ \sum_{j=0}^{\infty} a_j \psi_j : \sum_{j=0}^{\infty} |a_j|^2 < \infty \right\}, \quad \mathcal{H}_{\tau} = \left\{ \sum_{j=0}^{\infty} a_j r_{\tau,j} \psi_j : \sum_{j=0}^{\infty} |a_j|^2 < \infty \right\}.$$

Now note that since  $\lambda_{\tau,j} = \exp[-\tau(1/\lambda_j - 1)]$ , for every  $\tau_2 > 0$  and  $\tau_1 \in (0, \tau_2)$ , we have

$$\lambda_{\tau_2,j}/\lambda_{\tau_1,j} = \exp[(\tau_1 - \tau_2)(1/\lambda_j - 1)] < 1, \quad (27)$$

which shows that  $\mathcal{H}_{\tau_2} \subseteq \mathcal{H}_{\tau_1}$ . That  $\mathcal{H}_{\tau_1} \subseteq \mathcal{H}$  follows from the fact that the  $r_{\tau,j}$  are bounded. This completes the proof of Claim (ii).

Next, turning to Claim (iii), we have already established in Claim (i) that for every  $\tau > 0$ ,  $G_{\tau}$  is an  $L^2(\nu)$ -strictly positive, compact, contraction on  $L^2(\nu)$  with a simple eigenvalue  $\lambda_{\tau,0} = 1$ . The semigroup property follows directly from the facts the  $\phi_j$  form an orthonormal eigenbasis for all  $G_{\tau}$ ,  $\tau \geq 0$ , with eigenvalues  $\lambda_{\tau,j}$ , and for each  $j \in \mathbb{N}_0$  and  $\tau_1, \tau_2 \geq 0$ ,  $\lambda_{\tau_1+\tau_2,j} = \lambda_{\tau_1,j} \lambda_{\tau_2,j}$ . To establish strong continuity of this semigroup, it is enough to show that for every  $f \in L^2(\nu)$  and  $\epsilon > 0$ ,

$$\lim_{\tau \rightarrow 0^+} \|(G_{\tau} - \text{Id})f\|_{L^2(\nu)} < 2\epsilon. \quad (28)$$

Indeed, expanding  $f = \sum_{j=0}^{\infty} a_j \phi_j$ , the partial sum  $f_L = \sum_{j=0}^{L-1} a_j \phi_j$  with  $L$  large-enough satisfies  $\|f - f_L\|_{L^2(\nu)} < \epsilon$ . Then, because

$$(G_{\tau} - \text{Id})f = (G_{\tau} - \text{Id})f_L + (G_{\tau} - \text{Id})(f - f_L) = \sum_{j=0}^L a_j (\lambda_{\tau,j} - 1) \phi_j + (G_{\tau} - \text{Id})(f - f_L),$$

and  $\|G_{\tau}\| = \lambda_{\tau,0} = 1$ , the last term in the above equation can be bounded as  $\|(G_{\tau} - \text{Id})(f - f_L)\|_{L^2(\nu)} < \epsilon$ . Now note that for each  $j$ ,  $\lambda_{\tau,j} - 1 = \exp(\tau(1 - \lambda_j^{-1})) - 1$  converges to 0 as  $\tau \rightarrow 0^+$ , so that  $\|\sum_{j=0}^L a_j (\lambda_{\tau,j} - 1) \phi_j\|_{L^2(\nu)} < \epsilon$  for  $\tau$  small-enough, and (28) is satisfied. This proves Claim (iii).

We will now show that  $G_\tau$  is  $L^2(\nu)$ -Markov, completing the proof of Claim (i) and the theorem. By Hille-Yosida theory for strongly continuous, contraction semigroups of positive, compact operators [71], there exists a positive, self-adjoint operator  $\mathcal{L} : D(\mathcal{L}) \rightarrow L^2(\nu)$  with compact resolvent such that, for all  $\tau \geq 0$ ,  $G_\tau = e^{-\tau\mathcal{L}}$ .  $\mathcal{L}$  is a diagonal operator with eigenbasis  $\phi_j$  and corresponding eigenvalues

$$-\frac{d}{d\tau}\lambda_{\tau,j}\Big|_{\tau=0} = \frac{1}{\lambda_j} - 1.$$

In particular, since  $\lambda_0 = 1$  is simple,  $\mathcal{L}$  has a simple eigenvalue 0 corresponding to the constant eigenfunction  $\phi_0 \equiv 1_M$ . It then follows from results on Markov semigroups (e.g., [72, Chapter 14, Theorem 2]) that the semigroup generated by  $-\mathcal{L}$  is actually  $L^2(\nu)$ -Markov ergodic. That is, for every  $\tau > 0$ ,  $G_\tau = e^{-\tau\mathcal{L}}$  is a Markov operator with transition probability density  $p_\tau(x, \cdot)$  relative to  $\nu$ . This completes the proof of Claim (i) and Theorem 1.  $\square$

Before proceeding with the proof of Theorem 2, we will state a useful proposition, which is a consequence of the semigroup structure of the operator family  $\{G_\tau\}_{\tau \geq 0}$ . In what follows,  $\Pi_L : L^2(\mu) \rightarrow L^2(\mu)$  will denote the orthogonal projection onto the subspace spanned by  $\{\phi_0, \dots, \phi_{L-1}\}$ .

**Proposition 19.** *Under the assumptions of Theorem 2:*

- (i) *As  $\tau \rightarrow 0^+$ ,  $G_\tau^{-1/2}$  converges pointwise to the identity on  $H_\infty$ .*
- (ii) *For every  $\tau > 0$ , the compactified generator  $\tilde{V}_\tau : L^2(\mu) \rightarrow L^2(\mu)$  from Assumption 4 is equal to  $G_\tau^{1/2} V G_\tau^{1/2}$ .*
- (iii) *For every  $\tau > 0$ ,  $A_\tau$ ,  $B_\tau$ ,  $\tilde{V}_\tau$ , and  $W_\tau$  are trace class operators.*
- (iv) *The operator families  $\{A_\tau\}_{\tau > 0}$ ,  $\{B_\tau\}_{\tau > 0}$ ,  $\{\tilde{V}_\tau\}_{\tau > 0}$ , and  $\{W_\tau\}_{\tau > 0}$  are p2-continuous.*
- (v) *As  $\tau \rightarrow 0^+$ ,  $A_\tau$ ,  $B_\tau$ ,  $\tilde{V}_\tau$ , and  $\mathcal{U}_\tau^* W_\tau \mathcal{U}_\tau$  converge pointwise to  $V$  on  $D(V)$ .*

**Proof.** By (27), for every  $j$ ,  $\lambda_{\tau,j}$  increases strictly monotonically as  $\tau \rightarrow 0^+$ , which means that  $\lambda_{\tau,j}^{-1/2}$  decreases strictly monotonically. Now, since  $D(G_\tau^{-1/2}) = D(\mathcal{N}_\tau)$  (see Section 4.1) and  $G_\tau^{-1/2} : \phi_j \mapsto \lambda_{\tau,j}^{-1/2} \phi_j$ , for every  $f \in H_\infty$  and  $0 < \tau' < \tau$ , we have  $\|G_{\tau'}^{-1/2} f\|_{L^2(\mu)} \leq \|G_\tau^{-1/2} f\|_{L^2(\mu)}$ , and thus

$$H_\infty \subseteq D(G_\tau^{-1/2}) \subseteq D(G_{\tau'}^{-1/2}).$$

Therefore, fixing  $\epsilon > 0$  and  $\tau_0 > 0$ , to prove Claim (i) it is enough to show that

$$\lim_{\tau \rightarrow 0^+} \|G_\tau^{-1/2} f - f\|_{L^2(\mu)} = \lim_{\tau \rightarrow 0^+} \|(G_\tau^{-1/2} - \text{Id})f\|_{L^2(\mu)} < 2\epsilon, \quad \forall f \in D(G_{\tau_0}^{-1/2}). \quad (29)$$

To that end, we begin by using the triangle inequality to write down the bound

$$\|(G_\tau^{-1/2} - \text{Id})f\|_{L^2(\mu)} \leq \|(G_\tau^{-1/2} - \text{Id})\Pi_L f\|_{L^2(\mu)} + \|(G_\tau^{-1/2} - \text{Id})(\text{Id} - \Pi_L)f\|_{L^2(\mu)}. \quad (30)$$

Now, since  $G_\tau$  and  $G_\tau^{-1/2}$  are diagonal operators, they commute with  $\Pi_L$  and  $\text{Id} - \Pi_L$ . As a result, for every  $\tau \in (0, \tau_0)$ , by (27),

$$\|(G_\tau^{-1/2} - \text{Id})(\text{Id} - \Pi_L)f\|_{L^2(\mu)} \leq \|(G_{\tau_0}^{-1/2} - \text{Id})(\text{Id} - \Pi_L)f\|_{L^2(\mu)} = \|(\text{Id} - \Pi_L)(G_{\tau_0}^{-1/2} - \text{Id})f\|_{L^2(\mu)},$$

and the last term vanishes as  $L \rightarrow \infty$ . Therefore, for  $L$  large-enough, the second term on the right-hand side of (30) is less than  $\epsilon$  for every  $\tau \in (0, \tau_0)$ . Similarly, for any fixed  $L$ , for  $\tau$  small-enough, the first term is also less than  $\epsilon$ , proving (29) and Claim (i).

To prove Claim (ii), note that for every  $\tau' > 0$ ,

$$\tilde{V}_{2\tau'} \supseteq G_{2\tau'}^{1/2} V G_{2\tau'}^{1/2} = G_{\tau'} V G_{\tau'},$$

where the last equality follows from the semigroup structure of  $\{G_\tau\}_{\tau \geq 0}$ . However, the range of  $G_\tau$  lies in the domain of  $V$ , so we conclude that  $\tilde{V}_{2\tau'} = G_{\tau'} V G_{\tau'} = G_{\tau'} A_{\tau'}$ . Setting  $\tau' = \tau/2$  and noting that  $G_{\tau/2} = G_\tau^{1/2}$  leads to the claim.

Next, to prove Claim (iii), observe that  $\tilde{V}_\tau = G_{\tau/2} B_{\tau/2}$ , which shows that  $\tilde{V}_\tau$  is trace class since  $G_{\tau/2}$  is trace class and  $B_{\tau/2}$  is bounded. Similarly, we have  $B_\tau = B_{\tau/2} G_{\tau/2}$ , which shows that  $B_\tau$  is trace class. That  $W_\tau$  and  $A_\tau$  are trace class then follows from the fact that the former is unitarily equivalent to  $\tilde{V}_\tau$  and the latter equal to the negative adjoint of  $B_\tau$ .

Turning to Claim (iv), we will only prove p2-continuity for  $\{\tilde{V}_\tau\}_{\tau > 0}$ . The result for  $\{W_\tau\}_{\tau > 0}$  follows immediately by unitary equivalence of  $\tilde{V}_\tau$  and  $W_\tau$ , and the results for  $\{A_\tau\}_{\tau > 0}$  and  $\{B_\tau\}_{\tau > 0}$  can be verified analogously to the proof for  $\{\tilde{V}_\tau\}_{\tau > 0}$  below.

First, by Claim (ii), it is sufficient to establish p2-continuity for the family of operators  $\{G_\tau A_\tau\}_{\tau > 0}$ . That is, fixing a quadratic polynomial  $Q$ , we have to show that the operator norm  $\|Q(G_\tau A_\tau)\|_{L^2(\mu)}$  is a continuous function of  $\tau > 0$ . This is in turn equivalent to showing that  $\tau \mapsto Q(G_\tau A_\tau)$  is a continuous map in the  $L^2(\mu)$  operator norm topology. Note that this continuity is not affected by the addition of a constant term to the polynomial  $Q$ . Thus, without loss of generality, we may assume that  $Q$  is a homogeneous polynomial of the form  $Q(x) = \alpha x^2 + \beta x$ . By Theorems 1 and 5,  $G_\tau$  and  $A_\tau$  are both Hilbert-Schmidt integral operators with kernels  $p_\tau$  and  $p'_\tau$ , respectively. Since the composition of a bounded operator with a Hilbert-Schmidt operator is again a Hilbert-Schmidt operator, it follows that

$$Q(G_\tau A_\tau) = \alpha G_\tau \circ A_\tau \circ G_\tau \circ A_\tau + \beta G_\tau \circ A_\tau$$

is Hilbert-Schmidt. As a result, because the Hilbert-Schmidt norm induces a stronger topology than the  $L^2(\mu)$  operator norm, it is sufficient to prove the stronger claim that  $\tau \mapsto Q(G_\tau A_\tau)$  is a continuous map in the Hilbert-Schmidt norm topology.

By (8), the Hilbert-Schmidt norm of the kernel integral operator  $Q(G_\tau A_\tau)$  is just the  $L^2(\mu \times \mu)$  norm of its kernel. Thus, denoting this kernel by  $q_\tau : M \times M \rightarrow \mathbb{R}$ , the task now is to show that  $\tau \mapsto \|q_\tau\|_{L^2(\mu \times \mu)}$  is a continuous function of  $\tau$ , or, equivalently, that  $\tau \mapsto q_\tau$  is continuous in the  $L^2(\mu \times \mu)$  norm topology. That this is indeed the case follows from the claims below.

(a)  $\tau \mapsto p_\tau$  and  $\tau \mapsto p'_\tau$  are continuous in the  $L^2(\mu \times \mu)$  norm topology. Indeed, by (10) and (26),

$$\|p_\tau - p_{\tau'}\|_{L^2(\mu)}^2 = \sum_{j=0}^{\infty} \left| \lambda_{\tau,j}^{1/2} - \lambda_{\tau',j}^{1/2} \right|^2 \leq \sum_{j=0}^L \left| \lambda_{\tau,j}^{1/2} - \lambda_{\tau',j}^{1/2} \right|^2 + \sum_{j=L+1}^{\infty} \lambda_{\tau,j} + \sum_{j=L+1}^{\infty} \lambda_{\tau',j},$$

so that for  $L$  sufficiently large, the last two terms can be made arbitrarily small, whereas for every fixed  $L$  the term  $\sum_{j=0}^L |\lambda_{\tau,j}^{1/2} - \lambda_{\tau',j}^{1/2}|^2$  converges to 0 as  $\tau \rightarrow \tau'$ . This establishes  $L^2(\mu \times \mu)$  continuity of  $\tau \mapsto p_\tau$ . The claim for  $\tau \mapsto p'_\tau$  follows analogously.

(b) If  $a_\tau, b_\tau : M \times M \rightarrow \mathbb{R}$  are two kernel families depending continuously on  $\tau$  with respect to  $L^2(\mu \times \mu)$  norm, then their composition,  $c_\tau(x, y) = \int_M a_\tau(x, z) b_\tau(z, y) d\mu(z)$ , is also continuous. This claim can be verified via a standard calculation in analysis, which will be omitted here.

The continuity of  $\tau \mapsto q_\tau$  then follows from these results since  $q_\tau$  is equal to a sum of various compositions of  $p_\tau$  and  $p'_\tau$ . This completes the proof of Claim (iv).

Finally, to prove Claim (v), fix  $f = \sum_{j=0}^{\infty} a_j \phi_j \in D(V)$ , and observe the following:

(a)  $G_\tau f$  is a family of functions in  $D(V)$ , converging, as  $\tau \rightarrow 0^+$  to  $f$ . The convergence follows from Theorem 1(iii). Moreover, since  $G_\tau$  has a  $C^1$  kernel  $p_\tau$ , the  $G_\tau f$  have  $C^1$  representatives. Thus,  $G_\tau f$  lies in  $D(V)$ , as claimed.

(b)  $P_\tau f$  is a Cauchy sequence in  $\mathcal{H}$ . To verify this, fix  $\tau_0 > 0$ . Then, for every  $\tau, \tau' \in (0, \tau_0)$ ,

$$\|P_{\tau'} f - P_\tau f\|_{\mathcal{H}}^2 = \sum_{j=0}^{\infty} \left( \lambda_{\tau',j}^{1/2} - \lambda_{\tau,j}^{1/2} \right)^2 |a_j|^2 = \sum_{j=0}^L \left( \lambda_{\tau',j}^{1/2} - \lambda_{\tau,j}^{1/2} \right)^2 |a_j|^2 + \sum_{j=L+1}^{\infty} \left( \lambda_{\tau',j}^{1/2} - \lambda_{\tau,j}^{1/2} \right)^2 |a_j|^2,$$

and therefore, since  $\lambda_{\tau,j} \in [0, 1)$  for every  $\tau > 0$  and  $j \in \mathbb{N}_0$ , we obtain

$$\limsup_{\tau_0 \rightarrow 0^+} \|P_{\tau'} f - P_\tau f\|_{\mathcal{H}}^2 \leq \limsup_{\tau_0 \rightarrow 0^+} \sum_{j=0}^L \left( \lambda_{\tau',j}^{1/2} - \lambda_{\tau,j}^{1/2} \right)^2 |a_j|^2 + 2 \sum_{j=L+1}^{\infty} |a_j|^2 = 2 \sum_{j=L+1}^{\infty} |a_j|^2.$$

The above inequality holds for every  $L \in \mathbb{N}$ , and the last term vanishes as  $L \rightarrow \infty$ , proving the claim.

(c)  $A_\tau f$  is a Cauchy sequence in  $L^2(\mu)$ . To verify this, note that  $VP^* : \mathcal{H} \rightarrow L^2(\mu)$  is a bounded operator by Theorem 6(i), and therefore, since  $P_\tau f$  is a Cauchy sequence in  $\mathcal{H}$ ,  $A_\tau f = VP_\tau^* P_\tau f = VP^*(P_\tau f)$  is a Cauchy sequence in  $L^2(\mu)$ .

We have thus shown that  $G_\tau f$  is a family of functions in  $D(V)$  which converges as  $\tau \rightarrow 0^+$  to  $f$ , and whose image under  $V$ , namely  $V(G_\tau f) = A_\tau f$ , forms a Cauchy sequence. Since  $V$  is a closed operator, the limit of this Cauchy sequence is equal to  $Vf$ . Thus,  $A_\tau f$  converges to  $Vf$ , and since  $f$  was arbitrary, it follows that  $A_\tau$  converges to  $V$  pointwise on  $D(V)$ . In addition, because  $G_\tau A_\tau = \tilde{V}_{2\tau}$ , and  $G_\tau$  is uniformly bounded and converges to the identity, it follows that  $\tilde{V}_\tau$  also converges pointwise to  $V$  on  $D(V)$ . Note that we have used Claim (ii) to deduce equality of  $G_\tau A_\tau$  and  $\tilde{V}_{2\tau}$ . Finally, the pointwise convergence of  $B_\tau = G_\tau V$  to  $V$  follows directly from the pointwise convergence of  $G_\tau$  to the identity, and the result for  $\mathcal{U}_\tau^* W_\tau \mathcal{U}_\tau = \tilde{V}_\tau$  is obvious. This completes the proof of Proposition 19.  $\square$

**Proof of Theorem 2.** First, Proposition 19(iii) established that  $W_\tau$  and  $B_\tau$  are trace class. Moreover, Claim (i) of the theorem follows from Theorem 6(ii), Claim (ii) follows from Theorem 7, and Claim (iii) follows from Theorem 9(i) and (14). Aside from the convergence of  $P_\tau^* Z(W_\tau) \mathcal{N}_\tau$  to  $Z(V)$  for bounded continuous (as opposed to holomorphic) functions, Claims (iv)–(vii) will follow from Theorems 9, 10 and Proposition 19(iv) if we can show that  $p_\tau$  satisfies Assumption 4. Theorem 1(iii) establishes the condition in this assumption that  $G_\tau$  converges pointwise to the identity. In order to verify Assumption 4, it thus remains to be shown that, as  $\tau \rightarrow 0^+$ ,  $\tilde{V}_\tau$  converges pointwise to  $V$  on  $D(V^2)$ . This follows immediately from Proposition 19(v), where we have shown the stronger result that  $\tilde{V}_\tau$  converges to  $V$  pointwise on the whole of  $D(V)$ .

What remains to complete the proof of Theorem 2 is to show that  $P_\tau^* Z(W_\tau) \mathcal{N}_\tau$  converges strongly on  $H_\infty$  to  $Z(V)$  for bounded continuous  $Z$ . By (19), for every  $f \in H_\infty$  we have

$$P_\tau^* Z(W_\tau) \mathcal{N}_\tau f = P_\tau^* \mathcal{U}_\tau \mathcal{U}_\tau^* Z(W_\tau) \mathcal{U}_\tau G_\tau^{-1/2} f = G_\tau^{1/2} Z(\tilde{V}_\tau) G_\tau^{-1/2} f,$$

and therefore

$$\begin{aligned} P_\tau^* Z(W_\tau) \mathcal{N}_\tau f - Z(V) f &= G_\tau^{1/2} Z(\tilde{V}_\tau) G_\tau^{-1/2} f - Z(V) f \\ &= G_\tau^{1/2} Z(\tilde{V}_\tau) (G_\tau^{-1/2} - \text{Id}) f + (G_\tau^{1/2} Z(\tilde{V}_\tau) - Z(V)) f. \end{aligned}$$

By Proposition 19(i) and the fact that  $G_\tau^{1/2} Z(\tilde{V}_\tau)$  is a uniformly bounded family of operators converging pointwise to  $Z(V)$ , as  $\tau \rightarrow 0^+$ , each of the terms in the right-hand side of the last equation converges to 0. This shows that  $P_\tau^* Z(W_\tau) \mathcal{N}_\tau f \xrightarrow{s} Z(V)$  on  $H_\infty$ , completing the proof of Theorem 2.  $\square$

Before proving Corollary 3, we will state and prove a proposition on the  $\epsilon$ -approximate spectrum of  $U^t$ . One of the important claims we make is that, suitably restricted to the space  $P^* \mathcal{H} = D(\mathcal{N}) \subset L^2(\mu)$ ,  $e^{tB_\tau}$  converges in norm to  $U^t$ , as opposed to merely strongly as shown in Theorem 2(vi). In particular, we will consider the quantity  $Q(t, \tau) = \|(U^t - e^{tB_\tau})P^*\|$  for  $t \in \mathbb{R}, \tau > 0$ , where  $\|\cdot\|$  denotes  $\mathcal{H} \rightarrow L^2(\mu)$  operator norm.

**Proposition 20.** *Let Assumptions 1, 2 hold, with  $r = 2$ . Then, the function  $Q$  is continuous, vanishes at  $t = 0$  for every  $\tau \in (0, \infty)$ , and converges to 0 as  $\tau \rightarrow 0^+$  for every  $t \in \mathbb{R}$ . Moreover, for every eigenfunction  $\zeta_\tau$  of  $W_\tau$  with eigenvalue  $i\omega_\tau$  and every  $t \in \mathbb{R}$ ,  $e^{i\omega_\tau t}$  lies in the  $\epsilon$ -approximate point spectrum of  $U^t$ , with*

$$\epsilon = Q(t, \tau) \sqrt{\mathcal{D}(z_\tau) + 1}, \quad z_\tau := P_\tau^* \zeta_\tau / \|P_\tau^* \zeta_\tau\|_{L^2(\mu)}, \quad \|U^t z_\tau - e^{i\omega_\tau t} z_\tau\|_{L^2(\mu)} < \epsilon.$$

Moreover, for every fixed  $\epsilon > 0$ ,  $R(\epsilon, \tau)$  defined in Corollary 3 diverges as  $\tau \rightarrow 0^+$ .

**Proof.** By arguments analogous to those used to prove Proposition 19(iv), the map  $(t, \tau) \mapsto (U^t - e^{tB_\tau})P^*$  is continuous in the Hilbert-Schmidt norm topology of operators from  $\mathcal{H}$  into  $L^2(\mu)$  at every  $(t, \tau) \in \mathbb{R} \times \mathbb{R}_+$ . This implies continuity of  $(t, \tau) \mapsto (U^t - e^{tB_\tau})$  in the operator norm topology, and thus continuity of  $Q$ . That  $Q(t, \tau)$  vanishes as  $\tau \rightarrow 0^+$  at fixed  $t$  follows from the fact that  $U^t - e^{tB_\tau}$  converges pointwise to 0, and  $P^*$  is compact. That  $Q(0, \cdot) = 0$  is obvious. Next, to verify that  $e^{i\omega_\tau t}$  lies in the  $\epsilon$ -approximate point spectrum of  $U^t$  with  $\epsilon = Q(t, \tau) \sqrt{\mathcal{D}(z_\tau) + 1}$ , we use Theorem 7(ii) to compute

$$\begin{aligned} \|U^t z_\tau - e^{i\omega_\tau t} z_\tau\|_{L^2(\mu)} &= \frac{\|U^t P_\tau^* \zeta_\tau - e^{i\omega_\tau t} P_\tau^* \zeta_\tau\|_{L^2(\mu)}}{\|P_\tau^* \zeta_\tau\|_{L^2(\mu)}} = \frac{\|U^t P_\tau^* \zeta_\tau - P_\tau^* e^{tW_\tau} \zeta_\tau\|_{L^2(\mu)}}{\|P_\tau^* \zeta_\tau\|_{L^2(\mu)}} \\ &= \frac{\|(U^t - e^{tB_\tau})P_\tau^* \zeta_\tau\|_{L^2(\mu)}}{\|P_\tau^* \zeta_\tau\|_{L^2(\mu)}} = \frac{\|(U^t - e^{tB_\tau})P^* \zeta_\tau\|_{L^2(\mu)}}{\|P^* \zeta_\tau\|_{L^2(\mu)}} \\ &\leq \|(U^t - e^{tB_\tau})P^*\| \frac{\|\zeta_\tau\|_{\mathcal{H}}}{\|P^* \zeta_\tau\|_{L^2(\mu)}} = Q(t, \tau) \sqrt{\mathcal{D}(z_\tau) + 1}, \end{aligned} \quad (31)$$

Finally, fix an  $\epsilon > 0$ . It follows by continuity of  $Q$ , that for every  $T > 0$ ,  $t \in [-T, T]$ , and small-enough  $\tau > 0$ , we have  $Q(t, \tau) < \epsilon$ . This implies that  $R(\epsilon, \tau) > T$  for small enough  $\tau$ . Since  $T$  was arbitrary, we can conclude that  $R(\epsilon, \tau)$  diverges as  $\tau \rightarrow 0^+$ .  $\square$

**Proof of Corollary 3.** The first inequality follows from the definition of  $Q(t, \tau)$  and  $R(\epsilon, \tau)$ , in conjunction with (31). Next, to prove Claim (i), it is sufficient to show that  $e^{i\omega t}$  lies in the spectrum of  $U^t$  for every  $t \in \mathbb{R}$ . To that end, we use the triangle inequality and the fact that  $\|z_\tau\|_{L^2(\mu)} = 1$  to obtain the bound

$$\|U^t z_\tau - e^{i\omega t} z_\tau\|_{L^2(\mu)} \leq \|U^t z_\tau - e^{i\omega_\tau t} z_\tau\|_{L^2(\mu)} + |e^{i\omega_\tau t} - e^{i\omega t}|, \quad \forall \tau \in \mathbb{R}_+. \quad (32)$$

Now, because  $\lim_{\tau \rightarrow 0^+} \omega_\tau = \omega$ , there exists  $\tau_0 > 0$  such that for all  $\tau \in (0, \tau_0)$ ,  $|e^{i\omega_\tau t} - e^{i\omega t}| < \epsilon/2$ . Moreover, because  $T(\epsilon, \tau)$  is unbounded, there exists  $\tau_1 \in (0, \tau_0]$  such that  $t$  lies in the interval  $(-T(\epsilon/2, \tau), T(\epsilon/2, \tau))$  from Claim (i) for all  $\tau \in (0, \tau_1)$ . As a result, the bound in (31) becomes

$$\|U^t z_\tau - e^{i\omega t} z_\tau\|_{L^2(\mu)} \leq \epsilon/2 + \epsilon/2 = \epsilon, \quad \forall \tau \in (0, \tau_1).$$

Since  $\epsilon$  was arbitrary, we conclude that  $U^t - e^{i\omega t}$  has no bounded inverse, i.e.,  $e^{i\omega t}$  lies in the spectrum of  $U^t$ , as claimed.

Claim (ii) will be proven by contradiction. In particular, assume that there exists a sequence  $\tau_j > 0$  monotonically converging to 0 as  $j \rightarrow \infty$ , and  $\delta > 0$  such that for every  $j \in \mathbb{N}$ ,  $z_{\tau_j}$  is at distance at least  $\delta$

from the 1-dimensional eigenspace  $\mathcal{Z}$  of  $V$  corresponding to  $i\omega$ . Here, as a measure of distance of a vector  $z \in L^2(\mu)$  from  $\mathcal{Z}$  we use  $d(z, \mathcal{Z}) := \inf\{\|z - z'\|_{L^2(\mu)} : z' \in \mathcal{Z}\}$ . Since  $\|z_{\tau_j}\|_{L^2(\mu)} = 1$ , it follows from the boundedness of  $\mathcal{D}(z_{\tau_j})$  that  $\|z_{\tau_j}\|_{\mathcal{N}}$  is bounded. Therefore, by compactness of the embedding of  $D(\mathcal{N})$  into  $L^2(\mu)$ ,  $z_{\tau_j}$  has a subsequence converging to some vector  $z \in L^2(\mu)$ . By assumption on the  $\tau_j$ ,  $d(z, \mathcal{Z})$  is greater than  $\delta$ . We will complete the proof by showing that  $z$  lies, in fact, in  $\mathcal{Z}$ , leading to a contradiction. To that end, note that the condition that  $\mathcal{D}(z_{\tau})$  is bounded, together with the fact that  $R(\epsilon, \tau)$  diverges from Proposition 20, implies that  $T(\epsilon, \tau)$  diverges. Thus, the conclusion of Claim (i) holds, and  $z$  satisfies  $\|U^t z - e^{i\omega t} z\|_{L^2(\mu)} < \epsilon$  for every  $\epsilon > 0$ . We therefore conclude that  $U^t z = e^{i\omega t} z$  for every  $t \in \mathbb{R}$ , i.e., that  $z$  lies in  $\mathcal{Z}$ , in contradiction with the assumption that  $d(z, \mathcal{Z}) > \delta$ . This completes the proof of Claim (ii).  $\square$

**Proof of Corollary 4.** Since  $\{\phi_0, \phi_1, \dots\}$  is an orthonormal basis of  $L^2(\mu)$ , for  $L$  large enough,  $f_L := \Pi_L f$  satisfies  $\|f - f_L\|_{L^2(\mu)} < \epsilon/2$ . Moreover, since  $\{U^t\}_{t \in \mathbb{R}}$  is a unitary group, the inequality  $\|U^t f - U^t f_L\|_{L^2(\mu)} < \epsilon/2$  is preserved for all  $t \in \mathbb{R}$ . Moreover,  $f_L$  lies in  $H_\infty$  as it is a finite linear combination of the  $\phi_j$ . Now define  $\hat{f}_\epsilon = \mathcal{N}f_L$ , so that  $\hat{f}_\epsilon \in \mathcal{H}_\infty$ , and  $P^* \hat{f}_\epsilon = f_L$ . An application of Theorem 2(v) with  $Z(i\omega) = e^{i\omega t}$  then shows that, as  $\tau \rightarrow 0^+$ ,  $\|U^t f_L - P_\tau^* e^{tW_\tau} \hat{f}_\epsilon\|_{L^2(\mu)}$  converges to zero, where the convergence is uniform for  $t \in \mathcal{T}$  by continuity of the map  $t \mapsto U^t f_L - P_\tau^* e^{tW_\tau} \hat{f}_\epsilon$ . Therefore, there exists  $\tau_0 > 0$  such that for all  $\tau \in (0, \tau_0)$  and  $t \in \mathcal{T}$ ,  $\|U^t f_L - P_\tau^* e^{itW_\tau} \hat{f}_\epsilon\|_{L^2(\mu)} < \epsilon/2$ . Corollary 4 is then proved by the bound

$$\left\| U^t f - P_\tau^* e^{itW_\tau} \hat{f}_\epsilon \right\|_{L^2(\mu)} < \|U^t f - U^t f_L\|_{L^2(\mu)} + \left\| U^t f_L - P_\tau^* e^{itW_\tau} \hat{f}_\epsilon \right\|_{L^2(\mu)} < \epsilon/2 + \epsilon/2 = \epsilon. \quad \square$$

## 8. Data-driven approximation

We now take up the problem of approximating the operators in Theorems 1 and 2 from a finite time series of observed data and without prior knowledge of the dynamical flow  $\Phi^t$ . Specifically, we consider that available to us is a time series  $F(x_0), F(x_1), \dots, F(x_{N-1})$ , consisting of the values of an observation function  $F : M \rightarrow Y$  that takes values in a data space  $Y$ , sampled at a fixed time interval  $\Delta t > 0$  along an orbit  $x_0, x_1, \dots, x_{N-1}$  of the dynamics. As already alluded to in Section 1, besides the lack of knowledge of the dynamical flow map  $\Phi^t$ , this task presents a number of obstacles, including:

- (i) In general, one does not have direct access to the ergodic invariant measure  $\mu$  and the associated  $L^2(\mu)$  space, but is limited to working with the sampling measure  $\mu_N = \sum_{n=0}^{N-1} \delta_{x_n}/N$  supported on the finite trajectory  $\{x_0, \dots, x_{N-1}\}$ . In fact, even if  $\mu$  were explicitly known, its support  $X$  would typically be a non-smooth subset of the ambient manifold  $M$ , of zero Lebesgue measure (e.g., a fractal attractor), significantly hindering the construction of orthonormal bases of  $L^2(\mu)$  by restriction of smooth basis functions defined on  $M$ .
- (ii) In many experimental scenarios, the sampled states will not lie exactly on the invariant set  $X$ , as it is not feasible to achieve complete convergence of the trajectory to that set.
- (iii) Measurements are not taken continuously in time, preventing direct evaluation of the action of the dynamical vector field  $\tilde{V}$  on functions.

To address the first two issues, we take advantage of the fact that in many ergodic dynamical systems encountered in applications, the statistical properties of observables with respect to the sampling measures associated with a suitable class of initial points  $x_0$  coincide with those of the invariant measure [73], as discussed below.

*Basin of a measure.* The basin of an invariant measure  $\mu$  is the set of initial points such that the sampling measures  $\mu_N$  on the trajectories starting from them converge weakly to  $\mu$ . More specifically, it is the set of points  $x_0 \in M$  such that for every continuous function  $f \in C^0(M)$ ,

$$\lim_{N \rightarrow \infty} \int_M f \, d\mu_N = \int_M f \, d\mu, \quad \mu_N = \frac{1}{N} \sum_{n=0}^{N-1} \delta_{x_n}, \quad x_n = \Phi^{n \Delta t}(x_0).$$

This set will be denoted  $B_\mu$ . If  $\mu$  is ergodic, as assumed throughout this work, then  $\mu$ -a.e. point in  $M$  lies in its basin. The invariant measure  $\mu$  is said to be *physical* if  $B_\mu$  has nonzero measure with respect to some reference measure in the ambient manifold  $M$ . For instance, in typical experimental scenarios, initial points are drawn from some distribution equivalent to a smooth volume measure on  $M$ . In such cases, physicality of  $\mu$  ensures convergence of the data-driven techniques for a “large” set of initial conditions. While, in what follows, we will not require that  $\mu$  be physical as an explicit assumption, it should be kept in mind that some type of physicality is oftentimes an implied assumption in practical applications.

*Finite-difference approximation.* Following [17,18,74], to address the discrete-time sampling of the data, we approximate the action of the dynamical vector field  $\vec{V}$  on  $C^r$  functions using finite differences. As a concrete example, a scheme appropriate to the  $C^1$  regularity in Theorem 2 and Assumption 5 is a central finite-difference scheme  $\vec{V}_{\Delta t} : C^0(M) \rightarrow C^0(M)$ , given by

$$\vec{V}_{\Delta t} f(x) = \frac{f(\Phi^{\Delta t}(x)) - f(\Phi^{-\Delta t}(x))}{2 \Delta t}. \quad (33)$$

By compactness of  $M$ , for any  $f \in C^1(M)$ , the error  $\|\vec{V}_{\Delta t} f - \vec{V} f\|_{C^0(M)}$  of this scheme vanishes as  $\Delta t \rightarrow 0$ , and is  $o(\Delta t)$  and  $O((\Delta t)^2)$  if  $f$  lies in  $C^2(M)$  or  $C^3(M)$ , respectively.

The assumptions underlying our data-driven approximation schemes are as follows.

**Assumption 5.** *The dataset  $\{y_0, \dots, y_{N-1}\}$  consists of the values  $y_n = F(x_n)$  of an injective,  $C^1$  observation map  $F : M \mapsto Y$  into a manifold  $Y$ , sampled along a trajectory  $x_0, \dots, x_{N-1}$ ,  $x_n = \Phi^{n \Delta t}(x_0)$ , starting from a point  $x_0 \in B_\mu$  which is not a fixed point of the dynamics. Moreover:*

- (i) *The sampling interval  $\Delta t$  is such that  $\mu$  is an ergodic invariant measure of the map  $\Phi^{\Delta t} : M \rightarrow M$ .*
- (ii)  *$\kappa : Y \times Y \mapsto \mathbb{R}$  is a  $C^1$  symmetric, strictly positive-definite kernel with  $\kappa > 0$ .*

Note that Assumption 5(i) is satisfied iff  $\omega \Delta t$  is not a multiple of  $2\pi$  for any Koopman eigenfrequency  $\omega$ . For dynamics on a separable space, there can only be countably many such  $\omega$ , and thus Assumption 5(i) is satisfied for every  $\Delta t$  in a full-measure, co-countable subset of the real line. The manifold  $Y$  will be referred to as the data space. While it usually has the structure of a linear space (e.g.,  $Y = \mathbb{R}^m$ ), in a number of scenarios  $Y$  can be nonlinear (e.g., directional measurements with  $Y = S^2$ ).

The techniques described below will be based on the kernel  $k : M \times M \mapsto \mathbb{R}$ ,

$$k(x, x') := \kappa(F(x), F(x')), \quad (34)$$

induced from the kernel  $\kappa$  on data space. Note that  $k(x, x')$  can be evaluated given the data points  $F(x)$  and  $F(x')$ , without explicit knowledge of the underlying dynamical states  $x$  and  $x'$ . Moreover, the assumptions on  $\kappa$  and  $F$  in Assumption 5 ensure that  $k$  is also a  $C^1$  symmetric, strictly positive-definite kernel. We will discuss how to construct  $\kappa$  when the injectivity condition on  $F$  is not satisfied below.

*Data-driven Hilbert spaces.* Since the starting point  $x_0$  is not a fixed point, and  $\mu$  is an ergodic invariant measure of  $\Phi^{\Delta t}$ , all sampled states  $x_0, \dots, x_{N-1}$  are distinct. Therefore  $L^2(\mu_N)$ , is an  $N$ -dimensional Hilbert space, equipped with the inner product  $\langle f, g \rangle_{\mu_N} := \sum_{n=0}^{N-1} f^*(x_n)g(x_n)/N$ . This space consists of equivalence classes of complex-valued functions on  $M$  having common values at  $x_0, \dots, x_{N-1}$  (i.e., the support of  $\mu_N$ ). It is clear that  $L^2(\mu_N)$  is isomorphic to the space  $\mathbb{C}^N$  equipped with a normalized Euclidean inner product.

Note that one issue with establishing convergence of data-driven approximation techniques in this setting is that there is no obvious way of comparing functions in  $L^2(\mu_N)$  and  $L^2(\mu)$ . Here, we avoid this issue by performing our approximations in suitable RKHSs, whose elements can be projected into both  $L^2(\mu_N)$  and  $L^2(\mu)$ . The main elements of our approach, which closely parallel the theoretical results in Section 2, are (i) construction of a family of  $L^2(\mu_N)$ -Markov kernels with its associated semigroup and RKHSs,  $\mathcal{H}_{\tau,N}$ ; (ii) construction skew-adjoint operators  $W_{\tau,N}$  on  $\mathcal{H}_{\tau,N}$  approximating the compactified generator  $W_\tau$ , and evaluation of the spectral decomposition and functional calculus of these operators; and (iii) prediction of observables by exponentiation of the data-driven generators. We will now describe these procedures, and then, in Theorem 21, establish their convergence in the limit of large data. Pseudocode implementing our approach is included in Algorithms 1–4 in Appendix B.

*Markov kernels.* Using the bistochastic normalization procedure described in Section 4.1 with  $\nu$  set to the sampling measure  $\mu_N$  and  $k$  to the pullback kernel from (34), we construct a  $C^1$ ,  $L^2(\mu_N)$ -strictly-positive, Markov ergodic kernel  $p_N : M \times M \rightarrow \mathbb{R}$ . We then apply the construction in (10) with  $\nu = \mu_N$  to obtain a family of kernels  $p_{\tau,N} : M \times M \rightarrow \mathbb{R}$ ,  $\tau > 0$ , which are also  $L^2(\mu_N)$ -strictly-positive and Markov ergodic. Associated with  $p_N$  and  $p_{\tau,N}$  are RKHSs  $\mathcal{H}_N$  and  $\mathcal{H}_{\tau,N}$ , respectively, as well as the corresponding integral operators  $P_N : L^2(\mu_N) \rightarrow \mathcal{H}_N$ ,  $P_{\tau,N} : L^2(\mu_N) \rightarrow \mathcal{H}_{\tau,N}$ ,  $G_N = P_N^* P_N$ , and  $G_{\tau,N} = P_{\tau,N}^* P_{\tau,N}$ . In accordance with Theorem 1, the latter form  $L^2(\mu_N)$ -ergodic Markov semigroups for each  $N$ , with associated eigenvalues  $1 = \lambda_{\tau,N,0} > \lambda_{\tau,N,1} \geq \dots \geq \lambda_{\tau,N,N-1} > 0$ ,  $L^2(\mu_N)$ -orthonormal eigenfunctions  $\{\phi_{\tau,N,0}, \dots, \phi_{\tau,N,N-1}\}$ , and  $\mathcal{H}_{\tau,N}$ -orthonormal functions  $\{\psi_{\tau,N,0}, \dots, \psi_{\tau,N,N-1}\}$  (the latter, defined analogously to the  $\psi_{\tau,j}$  in (10)). The RKHSs  $\mathcal{H}_{\tau,N}$  also have associated Nyström extension operators,  $\mathcal{N}_{\tau,N} : L^2(\mu_N) \rightarrow \mathcal{H}_{\tau,N}$ . Note that because the  $L^2(\mu_N)$  are finite-dimensional spaces, and the eigenvalues  $\lambda_{\tau,N,j}$  are strictly positive, the Nyström operators  $\mathcal{N}_{\tau,N}$  are everywhere-defined.

*Data-driven generator.* Next, we construct finite-rank approximations of the compactified generator  $W_\tau$ . For that, note that every finite-difference scheme  $\vec{V}_{\Delta t}$  for the dynamical vector field induces a corresponding operator  $\tilde{V}_{N,\Delta t}$  on  $L^2(\mu_N)$ . For instance, the central finite-difference scheme in (33) leads to

$$\tilde{V}_{N,\Delta t} f(x_n) = \frac{f(x_{n+1}) - f(x_{n-1})}{2\Delta t}, \quad n \in \{1, \dots, N-2\}, \quad \tilde{V}_{N,\Delta t} f(x_0) = \tilde{V}_{N,\Delta t}(x_{N-1}) = 0.$$

While this operator is generally not skew-adjoint, it can be employed to construct a skew-adjoint operator  $V_{N,\Delta t} : L^2(\mu_N) \rightarrow L^2(\mu_N)$  by antisymmetrization, namely,

$$V_{N,\Delta t} = \frac{\tilde{V}_{N,\Delta t} - \tilde{V}_{N,\Delta t}^*}{2}. \quad (35)$$

The latter is a data-driven approximation of  $V$ , which adheres to our general scheme of approximating  $V$  using skew-adjoint operators. Note that  $V_{N,\Delta t}$  is fully characterized through its matrix elements  $\langle \phi_{\tau,N,i}, V_{N,\Delta t} \phi_{\tau,N,j} \rangle_{\mu_N}$  in the  $\phi_{\tau,N,j}$  basis of  $L^2(\mu_N)$ , which are in turn computable by applying (35) to the eigenfunction time series  $\phi_{\tau,N,j}(x_n)$ .

Next, using  $V_{N,\Delta t}$ , we construct the skew-adjoint operators  $W_{\tau,N,\Delta t}$  on  $\mathcal{H}_{\tau,N}$ , defined as

$$W_{\tau,N,\Delta t} = P_{\tau,N}^* V_{N,\Delta t} P_{\tau,N}.$$

It follows by definition of the  $\psi_{\tau,N,j}$  basis functions of  $\mathcal{H}_{\tau,N}$  that the matrix elements of  $W_{\tau,N,\Delta t}$  are related to those of  $V_{N,\Delta t}$  by

$$\langle \psi_{\tau,N,i}, W_{\tau,N,\Delta t} \psi_{\tau,N,j} \rangle_{\mathcal{H}_{\tau,N}} = \lambda_{\tau,N,i}^{1/2} \langle \phi_{\tau,N,i}, V_{N,\Delta t} \phi_{\tau,N,j} \rangle_{\mu_N} \lambda_{\tau,N,j}^{1/2}. \quad (36)$$

Note that as  $i$  and  $j$  grow, the matrix elements of  $W_{\tau,N,\Delta t}$  diminish in magnitude compared to those of  $V_{N,\Delta t}$  due to the decay of the eigenvalues. This is a manifestation of the RKHS regularization resulting from conjugation of  $V_{N,\Delta t}$  by  $P_{\tau,N}$ .

**Remark.** As is well known, finite-difference schemes are prone to errors if the sampling interval  $\Delta t$  is not sufficiently small, or high-frequency noise is present in the data. The extent to which the data-driven generator matrix in (36) is susceptible to these issues ultimately depends on the kernel  $k_\tau$ , as it governs the data-driven basis functions  $\psi_{\tau,N,j}$  appearing in the approximation. For instance, in [17, Theorem 22] it was shown that for quasiperiodic systems, incorporating delay-coordinate maps in the construction of the kernel can remove temporally i.i.d. noise of arbitrarily large variance, allowing (36) to be evaluated with “clean” eigenfunctions. While in systems with continuous spectrum that technique may have limitations (as adding delays would eventually suppress the kernel eigenfunctions spanning the continuous spectrum subspace  $H_c$  [18]), delay-coordinate maps should still be a useful tool for enhancing the noise robustness of the approximation in (36). Methods for controlling errors with respect to  $\Delta t$  would include composing the kernel with averaging operators to suppress high frequencies, or performing differentiation in the Fourier domain using spectral tapering methods. While exploring the efficacy of such approaches lies beyond the scope of this work, for the purposes of the methods presented in this paper, any convergent approximation of  $V$  can be employed in place of  $\vec{V}_{\Delta t}$  from (33).

*Spectral truncation.* In what follows, we will perform coherent pattern extraction and prediction using various spectrally truncated observables and operators. For that, we will need the orthogonal projections  $\Pi_{N,L} : L^2(\mu_N) \rightarrow L^2(\mu_N)$  and  $\Pi_{\tau,N,L} : \mathcal{H}_{\tau,N} \mapsto \mathcal{H}_{\tau,N}$  mapping into  $\text{span}\{\phi_{N,0}, \dots, \phi_{N,L-1}\}$  and  $\text{span}\{\psi_{\tau,N,0}, \dots, \psi_{\tau,N,L-1}\}$ , respectively. With some abuse of notation, in what follows  $\iota : B(M) \rightarrow L^2(\mu)$  and  $\iota_N : B(M) \rightarrow L^2(\mu_N)$  will be the canonical inclusion and restriction maps, respectively, on the space  $B(M)$  of bounded, complex-valued, Borel functions on  $M$ , equipped with the supremum norm. With these definitions, given an observable  $f \in B(M)$  that is to be predicted, we will treat it by first mapping it into the spectrally truncated observable

$$f_{N,L} = \Pi_{N,L} \iota_N f \in L^2(\mu_N), \quad 1 \leq L \leq N. \quad (37)$$

Moreover, the operators used for coherent pattern extraction and prediction will be spectrally truncated analogs of  $W_{\tau,N,\Delta t}$ , namely

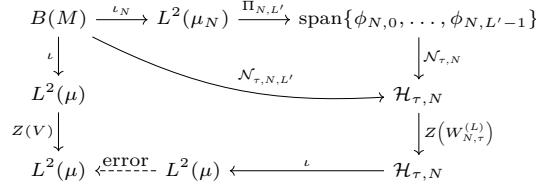
$$W_{\tau,N,\Delta t}^{(L)} := \Pi_{\tau,N,L} W_{\tau,N,\Delta t} \Pi_{\tau,N,L}, \quad 1 \leq L \leq N. \quad (38)$$

The reason for these spectral truncations will become clear below. Note that in applications the parameters  $L$  in (37) and (38) need not be equal. Moreover, since  $\Pi_{\tau,N,N} = \text{Id}$ ,  $W_{\tau,N,\Delta t}^{(N)}$  is equal to  $W_{\tau,N,\Delta t}$ .

*Coherent pattern extraction.* For any given  $L$ ,  $W_{\tau,N,\Delta t}^{(L)}$  is a skew-symmetric operator of rank at most  $L$ . In particular, it is diagonal in an orthonormal basis of eigenfunctions  $\zeta_{\tau,N,\Delta t,j}^{(L)} \in \mathcal{H}_{\tau,N}$ ,  $j \in \{0, \dots, L-1\}$ , corresponding to purely imaginary eigenvalues  $i\omega_{\tau,N,\Delta t,j}^{(L)}$ , i.e.,

$$W_{\tau,N,\Delta t}^{(L)} \zeta_{\tau,N,\Delta t,j}^{(L)} = i\omega_{\tau,N,\Delta t,j}^{(L)} \zeta_{\tau,N,\Delta t,j}^{(L)}. \quad (39)$$

The eigenfunctions  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  will act as data-driven coherent observables, approximating the eigenfunctions  $\zeta_{\tau,j}$  of  $W_\tau$ . It should be noted that  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  is a continuous function, constructed from the training data  $F(x_0), \dots, F(x_{N-1})$ , which can be evaluated at any  $x \in M$  from the corresponding value  $F(x)$  of the observation map in  $Y$ . This procedure is known as out-of-sample evaluation.



**Fig. 2.** Diagram illustrating the construction of the data-driven forecast function  $g_{\tau,N,\Delta t,L,L'}$  from (40) and its relationship to  $Z(V)\iota f$ . Starting with a bounded observable  $f \in B(M)$ , the left loop in the diagram leads to  $Z(V)\iota f$ , and the right loop to  $\iota g_{\tau,N,\Delta t,L,L'}$ . Note that  $Z(W_{\tau,N,\Delta t}^{(L)})$  maps into the  $L$ -dimensional subspace of  $\mathcal{H}_{\tau,N}$  spanned by  $\{\psi_{\tau,N,0}, \dots, \psi_{\tau,N,L-1}\}$ . The dashed arrow indicates discrepancy (error) between the data-driven function  $g_{\tau,N,\Delta t,L,L'}$  and the true observable,  $Z(V)\iota f$ . The composition of maps  $\mathcal{N}_{\tau,N} \circ \Pi_{N,L'} \circ \iota_N$  has been demarcated separately as an operator  $\mathcal{N}_{\tau,N,L'}$ . This operator represents an entire data-driven procedure which takes as input a  $B(M)$  function  $f$ , projects it onto its first  $L'$  components of a basis for  $L^2(\mu_N)$ , and then outputs the Nyström extension in  $\mathcal{H}_{\tau,N}$ . This output can then be the input of any operator  $Z(W_{N,\tau}^{(L)})$ , as above.

The reason for working with  $W_{\tau,N,\Delta t}^{(L)}$ , as opposed to the bare data-driven generator  $W_{\tau,N,\Delta t}$ , is twofold. First, in what follows, we will be interested in establishing a form of spectral convergence for the data-driven generators in the limit of large data—keeping  $L$  fixed while increasing  $N$  will allow us to ensure uniform convergence of the  $\psi_{\tau,N,\Delta t,j}$  with  $j \leq L-1$  to the corresponding  $\psi_{\tau,j}$ . Moreover, working at a fixed  $L \ll N$  allows to control the computational cost of data-driven approximations of  $W_{\tau}$ . In fact, following the computation of the  $L \times L$  matrix representing  $W_{\tau,N,\Delta t}^{(L)}$ , the cost of acting with this operator on observables becomes independent of the much larger data size  $N$ .

*Functional calculus and forecasting.* By skew-adjointness, the functional calculi of  $W_{\tau,N,\Delta t}^{(L)}$  can be conveniently constructed by applying any given function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$  to their eigenvalues, and projecting to the corresponding eigenspaces. That is,

$$Z(W_{\tau,N,\Delta t}^{(L)}) = \sum_{j=0}^{L-1} Z(i\omega_{\tau,N,\Delta t,j}^{(L)}) \langle \zeta_{\tau,N,\Delta t,j}^{(L)}, \cdot \rangle_{\mathcal{H}_{\tau,N}} \zeta_{\tau,N,\Delta t,j}^{(L)}.$$

Given such a bounded continuous function  $Z$  and a continuous observable  $f \in C^0(M)$ , our approximation for  $Z(V)\iota f$  is the  $\mathcal{H}_{\tau,N}$  function

$$g_{\tau,N,\Delta t,L,L'} = Z(W_{\tau,N,\Delta t}^{(L)}) \mathcal{N}_{\tau,N} f_{N,L'}, \quad (40)$$

where  $f_{N,L'}$  is given by (37), and  $L, L'$  are chosen such that  $1 \leq L' \leq L \leq N$ . Here, the role of the constraint  $L' \leq L$  is to control the error in the dynamical evolution of  $f_{N,L'}$  by the operator  $e^{tW_{\tau,N}^{(L)}}$  as  $L$  increases, keeping  $L'$  fixed. As with the eigenfunctions in (39),  $g_{\tau,N,\Delta t,L,L'}$  can be evaluated at an arbitrary state  $x \in M$ , given knowledge of  $F(x) \in Y$ . The relationships between the various maps employed in the construction of this approximation are depicted diagrammatically in Fig. 2.

With these constructions, we have the following convergence result.

**Theorem 21** (Data-driven approximation). *Let Assumptions 1, 2, and 5 hold. Then:*

- (i) *Every eigenfrequency  $\omega_{\tau,j}$  of  $W_{\tau}$ ,  $\tau > 0$ , can be consistently approximated by the eigenfrequencies  $\omega_{\tau,N,\Delta t,j}^{(L)}$  of  $W_{\tau}^{(L)}$ , in the sense that  $\lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0, N \Delta t \rightarrow \infty} \omega_{\tau,N,\Delta t,j}^{(L)} = \omega_{\tau,j}$ .*
- (ii) *For every eigenfunction  $\zeta_{\tau,j}$  of  $W_{\tau}$  corresponding to  $\omega_{\tau,j}$ , there exist eigenfunctions  $\zeta_{\tau,N,\Delta t}^{(L)}$  of  $W_{\tau,N,\Delta t}^{(L)}$  corresponding to  $\omega_{\tau,N,\Delta t,j}^{(L)}$  such that  $\lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0, N \Delta t \rightarrow \infty} \|\zeta_{\tau,N,\Delta t,j}^{(L)} - \zeta_{\tau,j}\|_{C^0(M)} = 0$ .*
- (iii) *For every bounded, continuous function  $Z : i\mathbb{R} \rightarrow \mathbb{C}$  and every bounded observable  $f \in B(M)$ ,*

$$\lim_{L' \rightarrow \infty} \lim_{\tau \rightarrow 0^+} \lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|Z(V)\iota f - \iota g_{\tau,N,\Delta t,L,L'}\|_{L^2(\mu)} = 0, \quad (41)$$

where  $g_{\tau,N,\Delta t,L,L'} \in \mathcal{H}_{\tau,N}$  is the data-driven approximation from (40).

**Remark.** Theorem 21 establishes convergence of the data-driven approximation  $g_{\tau,N,\Delta t,L,L'}$ , constructed for  $L' \leq L$ . Since  $L'$  is the last asymptotic control parameter taken to  $\infty$  in (41), an alternative formulation would be to keep  $L'$  constant, and state

$$\lim_{\tau \rightarrow 0^+} \lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|Z(V)\iota f - \iota g_{\tau,N,\Delta t,L,L'}\|_{L^2(\mu)} = 0, \quad \forall f \in \text{span}\{\phi_0, \dots, \phi_{L'-1}\}.$$

In this formulation,  $L'$  controls the dimension of the space of response observables on which we perform prediction, whereas  $L$  controls the dimension of the hypothesis space [53] in which the forecast function lies. Theoretically, we fix  $L'$  to attain convergence since  $\text{span}\{\phi_0, \dots, \phi_{L'-1}\}$  need not be invariant under  $Z(V)$ . In a numerical application with a fixed training dataset, the parameters  $L$  and  $L'$  can be tuned independently in a cross-validation step aiming to balance bias errors (increasing with decreasing  $L, L'$ ) and generalization errors (increasing with increasing  $L, L'$ ).

An application of Theorem 21(ii) for  $Z(i\omega) = e^{i\omega t}$  leads to the following corollary, establishing the convergence of the data-driven forecast functions for  $U^t f$ .

**Corollary 22.** *For every  $f \in B(M)$ , the function  $f_{\tau,N,\Delta t,L,L'}^{(t)} \in \mathcal{H}_{\tau,N}$  defined as*

$$f_{\tau,N,\Delta t,L,L'}^{(t)} = e^{tW_{\tau,N,\Delta t}^{(L)}} \mathcal{N}_{\tau,N} \Pi_{N,L'} \iota_N f$$

*is an approximation of  $U^t f$ , satisfying*

$$\lim_{L' \rightarrow \infty} \lim_{\tau \rightarrow 0} \lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|U^t f - f_{\tau,N,\Delta t,L,L'}^{(t)}\|_{L^2(\mu)} = 0.$$

*Moreover, the map  $t \mapsto f_{\tau,N,\Delta t,L,L'}^{(t)}$  is continuous, and the convergence is uniform for  $t$  lying in compact intervals.*

**Remark.** The order in which the limits in Theorem 21 and Corollary 22 are taken is important. In particular, the first limits taken are those of  $N$  and  $\Delta t$ . This corresponds to the limit of large data, i.e., infinitely many samples taken at arbitrarily small sampling interval. As stated above, in order to control sampling errors and ensure spectral convergence of the data-driven operators, the limit of large data must be taken at a fixed resolution  $L$ . After this, the limit  $L \rightarrow \infty$  is taken to facilitate a finite-rank approximation of  $W_\tau$  and  $Z(W_\tau)$ . Next, the limit  $\tau \rightarrow 0^+$  is taken as the 0-time limit of the Markov semi-group  $G_\tau$ , leading to convergence of  $W_\tau$  and  $Z(W_\tau)$  to  $V$  and  $Z(V)$ , respectively, in the sense of Theorem 2. Finally, in Theorem 21(iii) and Corollary 22, the limit  $L' \rightarrow \infty$  is taken to facilitate convergence to the spectrally truncated observable  $\Pi_{L'} \iota f \in L^2(\mu)$  to  $\iota f$ . The latter limit is analogous to an  $\epsilon \rightarrow 0^+$  limit of the tolerance  $\epsilon$  in Corollary 4.

Before proving Theorem 21, we will state an auxiliary lemma. In what follows,  $\Pi_{\tau,L}$  will denote the orthogonal projection on  $\mathcal{H}_\tau$  mapping into  $\text{span}\{\psi_{\tau,0}, \dots, \psi_{\tau,L-1}\}$ .

**Lemma 23.** *Under the assumptions of Theorem 21, the following hold:*

- (i) *The eigenvalues  $\lambda_{\tau,N,j}$  of  $G_{\tau,N}$  converge to those of  $G_\tau$ , i.e., for every  $\tau \geq 0$  and  $j \in \mathbb{N}_0$ ,  $\lim_{N \rightarrow \infty} \lambda_{\tau,N,j} = \lambda_{\tau,j}$ . Moreover, for every  $\mathcal{H}_\tau$  basis function  $\psi_{\tau,j}$  there exists a sequence of  $\mathcal{H}_{\tau,N}$  basis functions  $\psi_{\tau,N,j}$  converging to it in  $C^0(M)$  norm in the same limit.*
- (ii) *For every  $\tau > 0$ , the matrix elements of  $W_{\tau,N,\Delta t}$  from (36) converge to the corresponding matrix elements of  $W_\tau$ , i.e.,  $\lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \langle \psi_{\tau,N,i}, W_{\tau,N,\Delta t} \psi_{\tau,N,j} \rangle_{\mathcal{H}_{\tau,N}} = \langle \psi_{\tau,i}, W_\tau \psi_{\tau,j} \rangle_{\mathcal{H}_\tau}$ .*

(iii) As  $L \rightarrow \infty$ , the finite-rank, skew-adjoint operators  $W_\tau^{(L)} := \Pi_{\tau,L} W_\tau \Pi_{\tau,L}$  converge to  $W_\tau$  in Hilbert-Schmidt norm, and thus in  $\mathcal{H}_\tau$  operator norm.

**Proof.** Claim (i) was proved in [18], following the approach of [30], for Markov kernels constructed via the kernel normalization procedure introduced in the diffusion maps algorithm [29]. The result for the bistochastic Markov kernels from (20) follows analogously.

To verify Claim (ii), note first that the matrix elements  $\langle \phi_{N,i}, V_{N,\Delta t} \phi_{N,j} \rangle_{L^2(\mu_N)}$  converge to  $\langle \phi_i, V \phi_j \rangle_{L^2(\mu)}$  by convergence of the finite-difference approximation in (33) for  $C^1$  functions, in conjunction with the fact that the measure  $\mu$  is physical; see [18] for further details. The convergence of the  $\langle \psi_{\tau,N,i}, W_{\tau,N,\Delta t} \psi_{\tau,N,j} \rangle_{\mathcal{H}_{\tau,N}}$  to  $\langle \psi_{\tau,i}, W_\tau \psi_{\tau,j} \rangle_{\mathcal{H}_\tau}$  then follows from this result in conjunction with Claim (i).

Claim (iii) follows from the fact that  $\{\psi_{\tau,ij} := \langle \psi_{\tau,j}, \cdot \rangle_{\mathcal{H}_\tau} \psi_{\tau,i} : i, j \in \mathbb{N}_0\}$  is an orthonormal basis of the Hilbert space of Hilbert-Schmidt operators on  $\mathcal{H}_\tau$ , and in this basis, every Hilbert-Schmidt operator  $T : \mathcal{H}_\tau \rightarrow \mathcal{H}_\tau$  can be decomposed as  $T = \sum_{i,j=0}^{\infty} \langle \psi_{\tau,i}, T \psi_{\tau,j} \rangle_{\mathcal{H}_\tau} \psi_{\tau,ij}$ . In this expansion, the partial sums  $\sum_{i,j=0}^{L-1} \langle \psi_{\tau,i}, T \psi_{\tau,j} \rangle_{\mathcal{H}_\tau} \psi_{\tau,ij}$  are equal to  $\Pi_{\tau,L} T \Pi_{\tau,L}$ , and converge in Hilbert-Schmidt norm. Applying these results for  $T = W_\tau$  leads directly to the claim.  $\square$

**Proof of Theorem 21.** Because, by Lemma 23(iii),  $W_\tau^{(L)}$  is a sequence of compact operators converging in operator norm to the compact operator  $W_\tau$ , it follows that for every  $j \in \mathbb{N}_0$  such that  $\omega_{\tau,j} \neq 0$ , the eigenvalues  $i\omega_{\tau,j}^{(L)}$  of  $W_\tau^{(L)}$  converge to  $i\omega_j$ . Moreover, since, as follows directly from their definition, all  $W_\tau^{(L)}$  have an eigenvalue at zero corresponding to constant eigenfunctions, we conclude that the convergence  $\omega_{\tau,j}^{(L)} \xrightarrow[L \rightarrow \infty]{} \omega_{\tau,j}$  holds for all  $j \in \mathbb{N}_0$ . The convergence of the eigenvalues implies in turn that for every eigenfunction  $\zeta_{\tau,j}$  of  $W_\tau$  there exists a sequence of eigenfunctions  $\zeta_{\tau,j}^{(L)}$  of  $W_\tau^{(L)}$  converging to it in  $\mathcal{H}_\tau$  norm as  $L \rightarrow \infty$ . Claims (i) and (ii) will then follow if it can be shown that the eigenvalues of  $W_{\tau,N,\Delta t}^{(L)}$  converge to those of  $W_\tau^{(L)}$ , and the corresponding eigenfunctions converge in  $C^0(M)$  norm.

The convergence of the eigenvalues  $i\omega_{\tau,N,\Delta t}^{(L)}$  to  $i\omega_\tau^{(L)}$  follows from the convergence of the matrix elements of  $W_{\tau,N,\Delta t}^{(L)}$  to  $W_\tau^{(L)}$ , as established in Lemma 23(ii). The existence of eigenfunctions  $\zeta_{\tau,N,\Delta t}^{(L)}$  of  $W_{\tau,N,\Delta t}^{(L)}$  converging to  $\zeta_{\tau,j}^{(L)}$  in  $C^0(M)$  norm follows from the fact that both  $\zeta_{\tau,N,\Delta t}^{(L)}$  and  $\zeta_{\tau,j}^{(L)}$  are expressible as finite linear combinations of the  $\psi_{\tau,N,j}$  and  $\psi_{\tau,j}$ , namely

$$\zeta_{\tau,N,\Delta t,j}^{(L)} = \sum_{l=0}^{L-1} c_{\tau,N,\Delta t,l,j}^{(L)} \psi_{\tau,N,l}, \quad \zeta_{\tau,j}^{(L)} = \sum_{l=0}^{L-1} c_{\tau,l,j}^{(L)} \psi_{\tau,l},$$

where  $\bar{c}_{\tau,N,j}^{(L)} = (c_{\tau,N,\Delta t,0,j}^{(L)}, \dots, c_{\tau,N,\Delta t,L-1,j}^{(L)})^\top$  and  $\bar{c}_{\tau,j}^{(L)} = (c_{\tau,0,j}^{(L)}, \dots, c_{\tau,L-1,j}^{(L)})^\top$  are eigenvectors of the  $L \times L$  matrices representing  $W_{\tau,N,\Delta t}^{(L)}$  and  $W_\tau^{(L)}$ , respectively. By Lemma 23(i), the  $\psi_{\tau,N,j}$  converge to  $\psi_{\tau,j}$  in  $C^0(M)$  norm, and moreover for every eigenvector  $\bar{c}_{\tau,j}^{(L)}$  there exist  $\bar{c}_{\tau,N,j}^{(L)}$  converging to it in any vector norm. We therefore conclude that  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  converges in  $C^0(M)$  norm to  $\zeta_{\tau,j}^{(L)}$ , proving Claims (i) and (ii).

Turning to Claim (iii), we will verify that the limits in (41) hold in a sequential manner. First, defining  $f_{L'} = \Pi_{L'} \iota f$ , note that because  $Z(V)$  is a bounded operator and the  $\Pi_{L'}$  converge pointwise to the identity,  $\lim_{L' \rightarrow \infty} \|Z(V)\iota f - Z(V)f_{L'}\|_{L^2(\mu)} = 0$ . Thus, to verify (41), it suffices to show that

$$\lim_{\tau \rightarrow 0^+} \lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|Z(V)f_{L'} - \iota g_{\tau,N,\Delta t,L,L'}\|_{L^2(\mu)} = 0. \quad (42)$$

Second, observe that  $f_{L'}$  lies in  $H_\infty$ , and thus, by Theorem 2(v),  $\lim_{\tau \rightarrow 0^+} \|Z(V)f_{L'} - \iota Z(W_\tau)\mathcal{N}_\tau f_{L'}\|_{L^2(\mu)} = 0$ . As a result, to prove (42), it is enough to show that

$$\lim_{L \rightarrow \infty} \lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|\iota Z(W_\tau)\mathcal{N}_\tau f_{L'} - \iota g_{\tau,N,\Delta t,L,L'}\|_{L^2(\mu)} = 0. \quad (43)$$

Next, by Lemma 23(iii),  $\lim_{L \rightarrow \infty} \|Z(W_\tau)\mathcal{N}_\tau f_{L'} - Z(W_\tau^{(L)})\mathcal{N}_\tau f_{L'}\|_{\mathcal{H}_\tau} = 0$ , which implies that (43) holds if it can be shown that

$$\lim_{\Delta t \rightarrow 0^+, N \Delta t \rightarrow \infty} \|\iota Z(W_\tau^{(L)})\mathcal{N}_\tau f_{L'} - \iota g_{\tau, N, \Delta t, L, L'}\|_{L^2(\mu)} = 0. \quad (44)$$

The latter will follow in turn if it can be established that the vectors  $g_{N, \Delta t} = Z(W_{\tau, N, \Delta t}^{(L)})\mathcal{N}_{\tau, N} \Pi_{N, L'} f_{N, L'}$ , with  $f_{N, L'}$  given by (37), converge to  $g = Z(W_\tau^{(L)})\mathcal{N}_\tau \Pi_{L'} f_{L'}$  in  $C^0(M)$  norm. This fact follows from arguments similar to the proof of Claim (i). That is, writing  $g_{N, \Delta t} = \sum_{j=0}^{L-1} c_{N, \Delta t, j} \psi_{\tau, N, j}$  and  $g = \sum_{j=0}^{L-1} c_j \psi_{\tau, j}$ , one can verify the claimed convergence from the facts that (a) the functions  $\psi_{\tau, N, j}$  converge to  $\psi_{\tau, j}$  in  $C^0(M)$  norm; and (b) the expansion coefficients  $c_{N, \Delta t, j}$  and  $c_j$  are determined from the action of  $L \times L'$  matrices representing  $Z(W_{\tau, N, \Delta t}^{(L)})\mathcal{N}_{\tau, N} \Pi_{N, L'}$  and  $Z(W_\tau^{(L)})\mathcal{N}_\tau \Pi_{L'}$  on the  $L'$ -dimensional vectors representing  $f_{N, L'}$  and  $f_{L'}$ , respectively, all of which converge in the appropriate limit by Lemma 23. The sequence of limits in (42)–(44) then leads to (41), proving Claim (iii). This completes the proof of the theorem.  $\square$

*Approximation errors.* As stated above, the convergence results in Theorem 21 and Corollary 22 require that the limit of large data has to be taken before the limits involving the spectral truncation ( $L$  and  $L'$ ) and RKHS regularization ( $\tau$ ) parameters. Yet, in practical applications, one typically works with a fixed number of samples  $N$  and sampling interval  $\Delta t$ , and is faced with the question of tuning  $L$ ,  $L'$ , and  $\tau$  so as to achieve optimal performance. In particular, even though from a theoretical standpoint one would like to employ arbitrarily large  $L, L'$  and arbitrarily small  $\tau$ , such a choice would invariably lead to overfits of the training data and/or numerical instability. In the context of prediction (i.e., Theorem 21(iii) and Corollary 22), appropriate parameter values can be determined using cross-validation, i.e., by setting aside a portion of the available training data as verification data, and choosing  $L$ ,  $L'$ , and  $\tau$ , as well as other parameters (e.g., bandwidth parameters of Gaussian kernels as in (46) ahead), so as to maximize prediction skill in the verification dataset. In the context of spectral estimation (i.e., in the present work, coherent pattern extraction), parameter selection is more challenging, as typically there is no a priori known ground truth that can be employed for cross-validation. Instead, one way to proceed is through a posteriori analysis of the results, seeking to identify eigenvalues and eigenfunctions of  $W_{\tau, N, \Delta t}^{(L)}$  with minimal risk of being affected by sampling errors.

One such a posteriori metric is the Dirichlet energy of the eigenfunctions,  $\mathcal{D}_N(P_N^* \zeta_{\tau, N, \Delta t, j}^{(L)})$ , induced on  $L^2(\mu_N)$  by the RKHS  $\mathcal{H}_N$  according to (9). On the basis of well known results from statistical learning theory [53] (generally established for i.i.d. data, though analogous results hold for equidistributed data in an ergodic sense, as in the present work), the functional is a useful proxy for the sensitivity of  $\zeta_{\tau, N, \Delta t, j}^{(L)}$  to sampling errors. In Corollary 3, we established that the Dirichlet energy is also useful for identifying dynamical coherence. As a result,  $\mathcal{D}_N(P_N^* \zeta_{\tau, N, \Delta t, j}^{(L)})$  is a natural quantity to monitor for the purpose of identifying robust, data-driven coherent observables. Still, the raw Dirichlet energy does not take into account another source of error in our data-driven approximations, namely that we are approximating the unbounded generator  $V$  by a finite-difference operator  $V_{N, \Delta t}$  of the form in (35). Such operators, and as a result  $W_{\tau, N, \Delta t}^{(L)}$ , have  $L^2(\mu_N)$  operator norm of at most  $1/\Delta t$ , placing an effective Nyquist limit on the eigenfrequencies  $\omega_{\tau, N, \Delta t}^{(L)}$ , that can be recovered at a given sampling interval  $\Delta t$ . In particular, eigenfrequencies close to that limit are expected to have high sensitivity to  $\Delta t$ . The above suggests assessing the robustness of the data-driven eigenfunctions  $\zeta_{\tau, N, \Delta t, j}^{(L)}$  using a functional that depends on both the Dirichlet energy and eigenfrequency. In the experiments presented in Section 9 ahead, we will employ the frequency-adjusted Dirichlet energy on  $L^2(\mu_N)$  given by

$$\mathcal{D}_{N, \Delta t}(f) = \mathcal{D}_N(f) \left( 1 - \frac{(\Delta t \|V_{N, \Delta t} f\|_{L^2(\mu_N)})^2}{\|f\|_{L^2(\mu_N)}^2} \right)^{-1}, \quad \forall f \in L^2(\mu_N) \setminus \{0\}, \quad \text{and} \quad \mathcal{D}_{N, \Delta t}(0) = 0. \quad (45)$$

By construction, this functional takes small values on functions with low roughness (in the sense of Dirichlet energy  $\mathcal{D}_N(f)$ ), and thus reduced risk of sensitivity to sampling errors. Moreover, the term  $\|V_{N,\Delta t}f\|_{L^2(\mu_N)}^2$  can be thought of as a spectral energy in the frequency domain for the time series  $f(x_n)$  sampled discretely at times  $t_n = n\Delta t$ . The term  $(1 - \|V_{N,\Delta t}f\|_{L^2(\mu_N)}^2/\|f\|_{L^2(\mu_N)}^2)^{-1}$  then penalizes  $f$  whose frequency spectral energy is comparable to  $1/\Delta t^2$ . In what follows, we will order by convention all data-driven eigenfunctions  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  in order of increasing  $\mathcal{D}_{N,\Delta t}(P_N^* \zeta_{\tau,N,\Delta t,j}^{(L)})$ . We end this section with a discussion on how to obtain the kernel  $\kappa$  on data space.

*Choice of kernel.* First, note that the injectivity assumption on the observation map  $F$  is with minimal loss of generality. In particular, according to the theory of delay-coordinate maps of dynamical systems [36], under mild assumptions, the map  $F_Q : M \rightarrow Y^Q$ ,  $Q \in \mathbb{N}$ , defined as

$$F_Q(x) = \left( F(x), F(\Phi^{-\Delta t}x), \dots, F(\Phi^{-(Q-1)\Delta t}x) \right),$$

is injective for large-enough  $Q$ . Moreover,  $F_Q(x_n)$  can be evaluated for all states  $x_n$  with  $n \in \{Q-1, \dots, N\}$  given the values of  $F$  on a finite trajectory  $x_0, x_1, \dots, x_{N-1}$ . Thus, delay-coordinate maps are a useful remedy when the observation map is non-injective, which is frequently the case with experimental or observational data acquired from high-dimensional systems (e.g., engineering or geophysical fluid flows).

Assuming then that the observation map  $F$  is injective, one can implement the techniques described in this section with any  $C^1$  strictly positive-definite kernel on  $Y$ . As a concrete example for the case  $Y = \mathbb{R}^m$ , we mention here the radial Gaussian kernels,

$$\kappa(y, y') = \exp\left(-\frac{d^2(y, y')}{\epsilon}\right), \quad (46)$$

where  $d : Y \times Y \rightarrow \mathbb{R}$  is the standard Euclidean metric on  $\mathbb{R}^m$ , and  $\epsilon$  a positive bandwidth parameter. Such kernels are popular in manifold learning techniques [29,42] due to their ability to approximate heat kernels and the spectrum of the Laplace-Beltrami operator in the  $\epsilon \rightarrow 0^+$  limit. Here, we do not assume that the support  $X$  of the invariant measure has manifold structure, so generally we do not have a heat kernel interpretation. Nevertheless, radial Gaussian kernels are known to be strictly positive-definite on arbitrary subsets of  $\mathbb{R}^m$  [75], which is sufficient for our purposes. The numerical experiments in Section 9 will be carried out with a variable-bandwidth variant of (46), whose construction and basic properties are described in Appendix A.

## 9. Examples and discussion

In this section, we apply the procedure described in Section 8 to ergodic dynamical systems with different types of spectra. The objective is to illustrate the results of Theorems 1, 2, 21 and Corollaries 3, 4, and demonstrate that the framework is effective in identifying coherent observables and performing prediction in quasiperiodic and mixing systems. We consider the following three examples:

- (i) A linear, ergodic flow  $\Phi^t : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  on the 2-torus,

$$\Phi^t(\theta_1, \theta_2) = (\theta_1 + \alpha_1 t, \theta_2 + \alpha_2 t) \mod 2\pi,$$

where  $\alpha_1$  and  $\alpha_2$  are rationally independent frequencies, set to 1 and  $30^{1/2}$ , respectively. The observation map  $F : \mathbb{T}^2 \rightarrow \mathbb{R}^3$  is given by the standard embedding of the 2-torus into  $\mathbb{R}^3$ ,

$$F(\theta_1, \theta_2) = (F_1, F_2, F_3) = ((1 + R \cos \theta_1) \cos \theta_1, (1 + R \cos \theta_2) \sin \theta_1, \sin \theta_2),$$

where we set  $R = 1/2$ .

(ii) The L63 system [31], generated by the  $C^\infty$  vector field  $\vec{V}$  on  $\mathbb{R}^3$ , whose components  $(V_1, V_2, V_3)$  at  $(x, y, z) \in \mathbb{R}^3$  are given by

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

We use the standard parameter values  $\beta = 8/3$ ,  $\rho = 28$ ,  $\sigma = 10$ , and take  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  to be the identity map.

(iii) The Rössler system [76] on  $\mathbb{R}^3$ , generated by the smooth vector field  $\vec{V}$ , with components  $(V_1, V_2, V_3)$  at  $(x, y, z)$  given by

$$V_1 = -y - z, \quad V_2 = x + ay, \quad V_3 = b + z(x - c).$$

We use the standard parameter values  $a = 0.1$ ,  $b = 0.1$ ,  $c = 14$ , and as in the case of the L63 system, set  $F$  to the identity map.

*Methodology.* The following steps describe sequentially the entire numerical procedure carried out for each system. Additional algorithmic details, including pseudocode, are included in Appendices A and B.

1. Numerical trajectories  $x_0, x_1, \dots, x_{N-1}$  of length  $N$ , with  $x_n = \Phi^{n\Delta t}(x_0)$ , were generated using a sampling interval  $\Delta t > 0$ . In the case of the torus rotation,  $\Delta t$  was set to  $2\pi/500 \approx 0.013$ . The sampling interval in the L63 and Rössler experiments was 0.01 and 0.04, respectively. In all three experiments, the nominal number of samples was  $N = 64,000$ . In the torus case, we also show eigenfunction results for  $N = 6400$  to assess sensitivity to sampling errors. The trajectories for the torus experiments were computed analytically. The L63 and Rössler experiments utilize numerical trajectories generated in MATLAB, using the `ode45` solver. These trajectories start from arbitrary initial conditions in  $\mathbb{R}^3$ , followed by a spinup period of  $N \Delta t$  time units before collecting the actual “production” data.
2. The observation map  $F$  described for each system was used to generate the respective time series  $F(x_0), F(x_1), \dots, F(x_{N-1})$ . For our choices of  $F$ , all of these time series take values in  $\mathbb{R}^3$ . In addition, we generated time series  $f(x_0), f(x_1), \dots, f(x_{N-1})$  for various other continuous, real-valued observables for use in forecasting experiments (described below).
3. Data-driven eigenpairs  $(\lambda_{N,j}, \phi_{N,j})$  with  $j \in \{0, \dots, L-1\}$  were computed by applying Algorithm 1 to the dataset  $F(x_0), \dots, F(x_{N-1})$ . Throughout, we used the variable-bandwidth Gaussian kernel described in Appendix A, in conjunction with the bistochastic normalization procedure from Section 4.1. In addition, we tuned the kernel bandwidth  $\epsilon$  using an automatic procedure; see Appendix A for further details and references. The number of eigenfunctions employed in our experiments ranged from  $L = 500$  to 1000; i.e.,  $L \ll N$  in all cases. As described in Appendix A, the eigenpairs  $(\lambda_{N,j}, \phi_{N,j})$  for the bistochastic kernels employed here can be determined from the singular values and left singular vectors of a non-symmetric  $N \times N$  kernel matrix, without explicit formation of the Markov kernel matrix itself. We followed that approach here, using MATLAB’s `svds` iterative solver to perform the singular value decomposition (SVD). All pairwise distances in data space required for kernel evaluation were computed by brute force (as opposed to using approximate nearest-neighbor search) in MATLAB, retaining  $5000 \ll N$  nearest neighbors per datapoint.
4. Using the eigenpairs  $(\lambda_{N,j}, \phi_{N,j})$  from Step 3 as inputs, Algorithm 3 was applied to form the  $L \times L$  operator matrices for  $W_{\tau, N, \Delta t}^{(L)}$ , and compute the corresponding eigenfrequencies  $\omega_{\tau, N, \Delta t, j}^{(L)}$  and eigenfunctions  $\zeta_{\tau, N, \Delta t, j}^{(L)} \in \mathcal{H}_{\tau, N}$ . Throughout, we used the central finite-difference scheme in (33) (which, in this case, is  $O((\Delta t)^2)$ -accurate) to compute the matrix elements of  $W_{\tau, N, \Delta t}^{(L)}$ , and MATLAB’s `eig` solver to compute the  $(\omega_{\tau, N, \Delta t}^{(L)}, \zeta_{\tau, N, \Delta t}^{(L)})$  eigenpairs. In order to investigate the dependence of the spectra of  $W_{\tau, N, \Delta t}^{(L)}$  on

$\tau$  (particularly from the perspective of Corollary 3), we computed eigenfrequencies for logarithmically spaced values of  $\tau$ , and examined the  $\tau \mapsto \omega_{j,\tau,N,\Delta t}^{(L)}$  dependence through scatterplots. Moreover, for each eigenfunction, we computed its frequency-adjusted Dirichlet energy  $\mathcal{D}_{N,\Delta t}(P_N^* \zeta_{\tau,N,\Delta t,j}^{(L)})$  from (45), and ordered the eigenpairs  $(\omega_{\tau,N,\Delta t}^{(L)}, \zeta_{\tau,N,\Delta t}^{(L)})$  in order of increasing  $\mathcal{D}_{N,\Delta t}(P_N^* \zeta_{\tau,N,\Delta t,j}^{(L)})$ .

5. Forecasting experiments were performed by constructing the data-driven functions  $f_{\tau,N,\Delta t,L,L'}^{(t_m)}$  via Algorithm 4 for lead times  $t_m := m \Delta t$ ,  $m \in \mathbb{N}_0$ , in an interval  $[0, m_{\max} \Delta t]$ . In all cases, we used  $L = L'$ . Initial conditions for the forecasts were generated from a time series  $F(\hat{x}_0), \dots, F(\hat{x}_{\hat{N}-1})$  of the observation map, sampled on a dynamical trajectory  $\hat{x}_0, \dots, \hat{x}_{\hat{N}+m_{\max}-1}$  of length  $\hat{N} + m_{\max}$ , independent of the training data. The values  $U^{t_m} f(\hat{x}_n) = f(\hat{x}_{n+m})$  of the forecast observable on this trajectory were used as verification data to assess the out-of-sample predictions  $f_{\tau,N,\Delta t,L,L'}^{(t_m)}(F(\hat{x}_n))$ . The latter, were evaluated using Algorithm 2. In all experiments,  $\hat{N}$  was equal to  $N$ . Forecast errors were assessed through the  $L^2$  norm associated with the sampling measure  $\hat{\mu}_{\hat{N}} = \sum_{n=0}^{\hat{N}-1} \delta_{\hat{x}_n} / \hat{N}$ . Specifically, for a given lead time  $t_m$  we compute a normalized root mean square error (RMSE) metric,

$$\varepsilon(t_m) = \|U^{t_m} f - f_{\tau,N,\Delta t,L,L'}^{(t_m)}\|_{L^2(\hat{\mu}_{\hat{N}})} / \|f\|_{L^2(\hat{\mu}_{\hat{N}})},$$

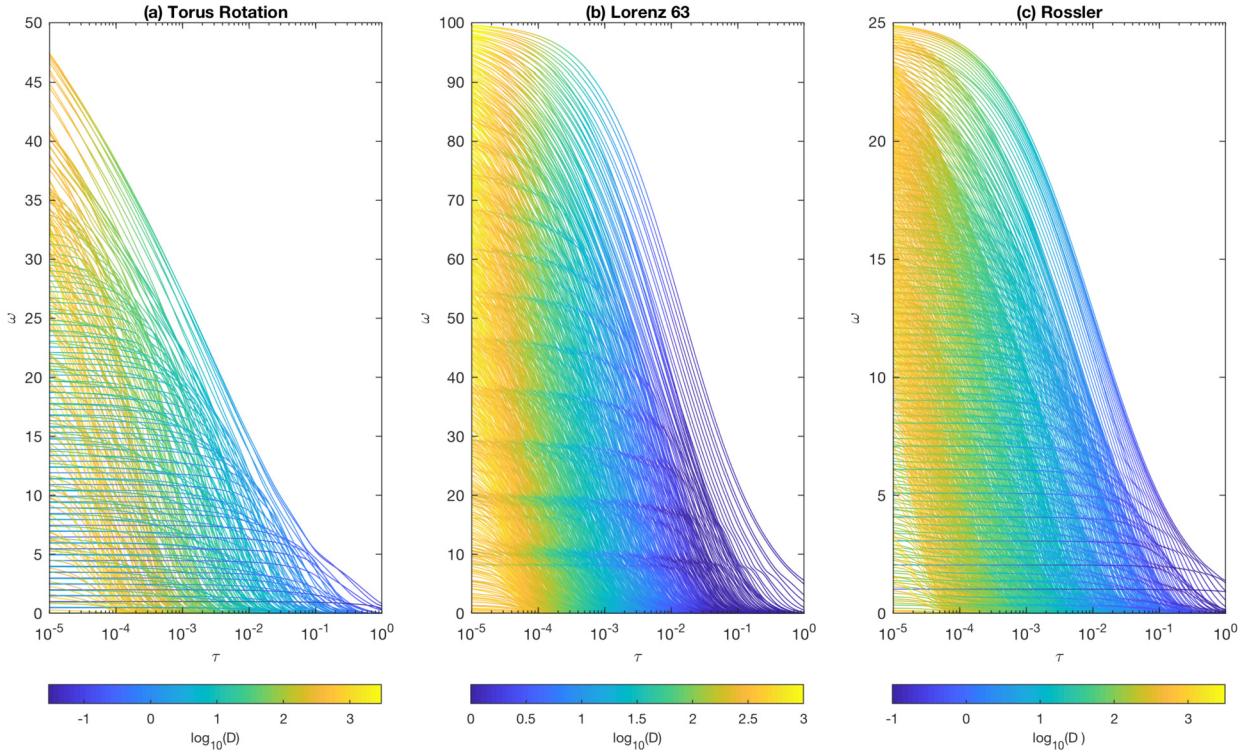
such that values  $\varepsilon(t_m) \ll 1$  correspond to skillful forecasts, whereas  $\varepsilon(t_m) \simeq 1$  indicates loss of skill. The metric  $\varepsilon(t_m)$  is an empirical estimator of the normalized expected error with respect to the invariant measure,  $\|U^{t_m} f - f_{\tau,N,\Delta t,L,L'}^{(t_m)}\|_{L^2(\mu)} / \|f\|_{L^2(\mu)}$ , to which it converges almost surely as  $\hat{N} \rightarrow \infty$ .

We now present and discuss the experimental results for each system. Hereafter, for notational simplicity, we will drop  $N$ ,  $L$ , and  $\Delta t$  subscripts and superscripts from data-driven eigenfrequencies, eigenfunctions, and operators. We will also use  $\mathcal{D}_{\tau,j}$  as a shorthand notation for the frequency-adjusted Dirichlet energy from (45) of the  $j$ -th eigenfunction of  $W_\tau$ .

*Linear flow on the 2-torus.* For any choice of rationally independent frequencies  $\alpha_1$  and  $\alpha_2$ , the system has a unique Borel ergodic invariant probability measure  $\mu$ , which coincides with the Haar measure on  $\mathbb{T}^2$ . Thus, in the notation of Assumption 1, the state space  $\mathcal{M}$ , the forward-invariant compact manifold  $M$ , and the support of the invariant measure  $X$  are all equal to  $\mathbb{T}^2$ . The basin of the invariant measure  $B_\mu$  from Section 8 is also equal to  $\mathbb{T}^2$ . For this invariant measure, the Koopman group on  $L^2(\mu)$  has pure point spectrum, consisting of eigenfrequencies of the form  $j_1 \alpha_1 + j_2 \alpha_2$ ,  $j_1, j_2 \in \mathbb{Z}$ , corresponding to the eigenfunctions  $e^{i(j_1 \theta_1 + j_2 \theta_2)}$ . The latter form an orthonormal basis of  $L^2(\mu)$ , so that the point and continuous spectrum subspaces in the invariant splitting in (3) are  $H_p = L^2(\mu)$  and  $H_c = \{0\}$ , respectively. Note that because  $\alpha_1$  and  $\alpha_2$  are rationally independent, the set of eigenfrequencies is dense in  $\mathbb{R}$ , which implies that the support of the PVM  $E$  of this system (in this case, the closure of its set of eigenvalues) is equal to the whole real line. This makes the problem of numerically distinguishing eigenfrequencies from non-eigenfrequencies non-trivial, despite the simplicity of the underlying dynamics.

Fig. 3(a) shows a scatterplot of the eigenfrequencies  $\omega_{\tau,j}$  and the corresponding Dirichlet energies  $\mathcal{D}_{\tau,j}$ , computed for  $L = 500$  and values of  $\tau$  logarithmically spaced in the interval  $[10^{-5}, 1]$ . There, the behavior of the numerically computed eigenfrequencies is broadly consistent with the results in Theorem 2 and Corollary 3. In particular, the eigenfrequencies are seen to form continuous curves parameterized by  $\tau$  (consistent with Theorem 2(vii) and Proposition 19(iv)), and the Dirichlet energy delineates the curves that have numerically converged over the examined values of  $\tau$  (i.e., as  $\tau$  approaches  $10^{-5}$ ), from those that have not (as expected from Corollary 3). Notice, in particular, that the task of visually identifying continuous eigenfrequency curves in Fig. 3(a) would be significantly more difficult without color-coding by Dirichlet energy.

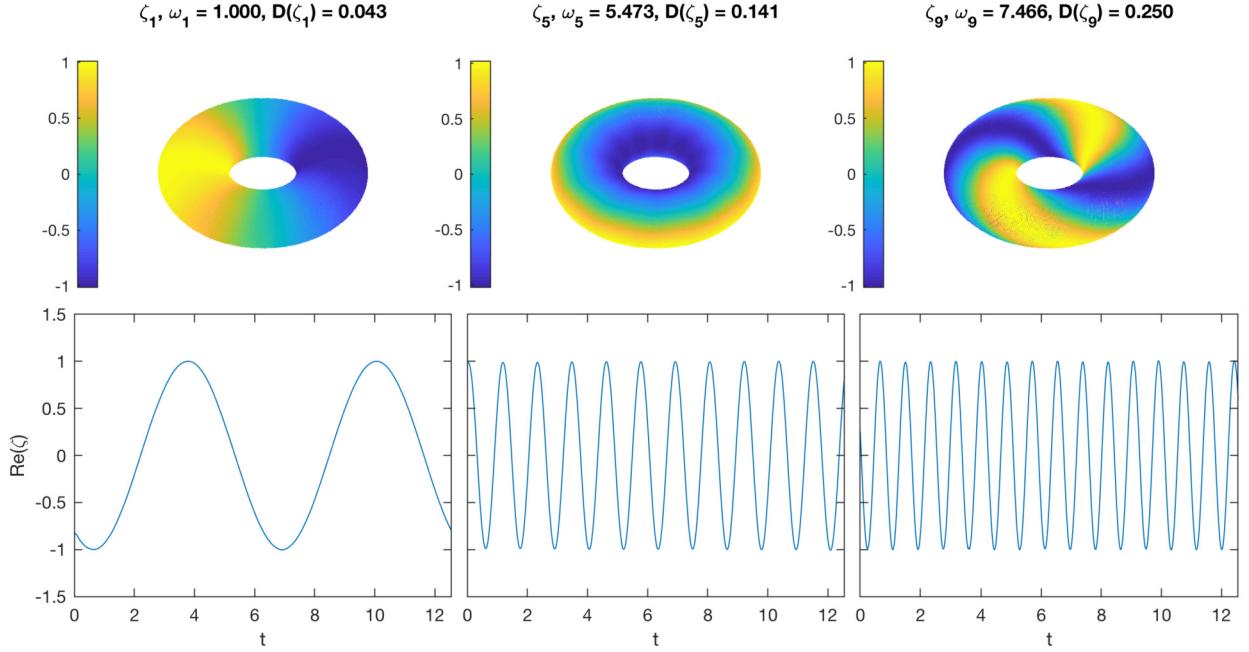
According to Corollary 3, the eigenfrequency curves of  $W_\tau$  with bounded Dirichlet energies should approximate Koopman eigenfrequencies, and the corresponding eigenfunctions should approximate Koopman



**Fig. 3.** Eigenfrequencies  $\omega_j$  of the data-driven generators  $W_\tau$  as a function of  $\tau$ , for (a) the linear torus flow; (b) the L63 system; and (c) the Rössler system. Colors represent the logarithms of the frequency-adjusted Dirichlet energies from (45) of the corresponding eigenfunctions. Only positive frequencies are shown, as the  $\omega < 0$  parts of the spectra are mirror images of the  $\omega > 0$  parts by skew-adjointness and reality of  $W_\tau$ . (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

eigenfunctions. Indeed, as illustrated in Fig. 4, the leading data-driven eigenfrequencies  $\omega_{\tau,j}$  for  $\tau = 10^{-5}$  agree with the theoretical eigenfrequencies to two to four significant figures. Moreover, the corresponding eigenfunctions agree well with the Koopman eigenfunctions of this system; that is,  $\zeta_{\tau,j}$  in Fig. 4 have the structure of Fourier functions on the 2-torus, with near-exact sinusoidal time series at the corresponding eigenfrequencies.

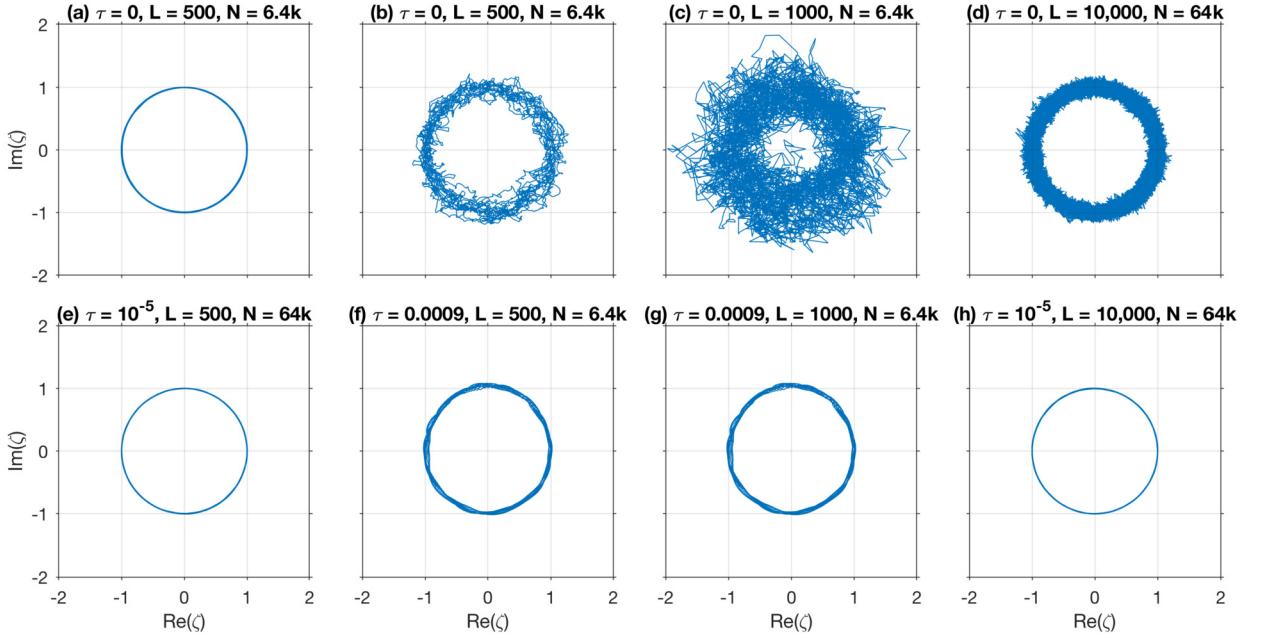
Next, to assess the significance of RKHS regularization in spectral approximation of the generator, in Fig. 5 we compare the real and imaginary parts of numerical eigenfunctions  $\zeta_{\tau,j}$  of  $W_\tau$  with eigenfunctions obtained from a “naive” approximation of the generator with  $\tau = 0$ . In all cases, we select the eigenfunction whose corresponding eigenfrequency is closest to the generating eigenfrequency  $\alpha_1 = 1$ , and plot the real and imaginary parts of  $\zeta_{\tau,j}$  as a scatterplot in the complex plane. For an exact approximation of a normalized Koopman eigenfunction, the plotted points should lie in the unit circle. According to the results in the figure, the naive approximation performs comparably to the regularized approximation for the experiment with  $N = 64,000$  and  $L = 500$  basis functions, but the quality of the approximation has considerably higher sensitivity to the number of samples and/or basis functions employed. Indeed, as is evident from Figs. 5(b) and 5(e), decreasing the number of samples to  $N = 6400$  imparts a significant amount of high-frequency noise in the eigenfunction obtained from the naive approximation, whereas the eigenfunction based on RKHS regularization is comparatively more stable. This can be understood from the fact that the quality of the data-driven basis functions  $\phi_{N,j}$  generally decreases with decreasing  $N$ , and the amount of quality degradation is higher the smaller the corresponding eigenvalue  $\lambda_{N,j}$  is. As a result, without regularization to suppress the basis functions corresponding to small  $\lambda_{N,j}$ , the quality of the naive approximation also degrades.



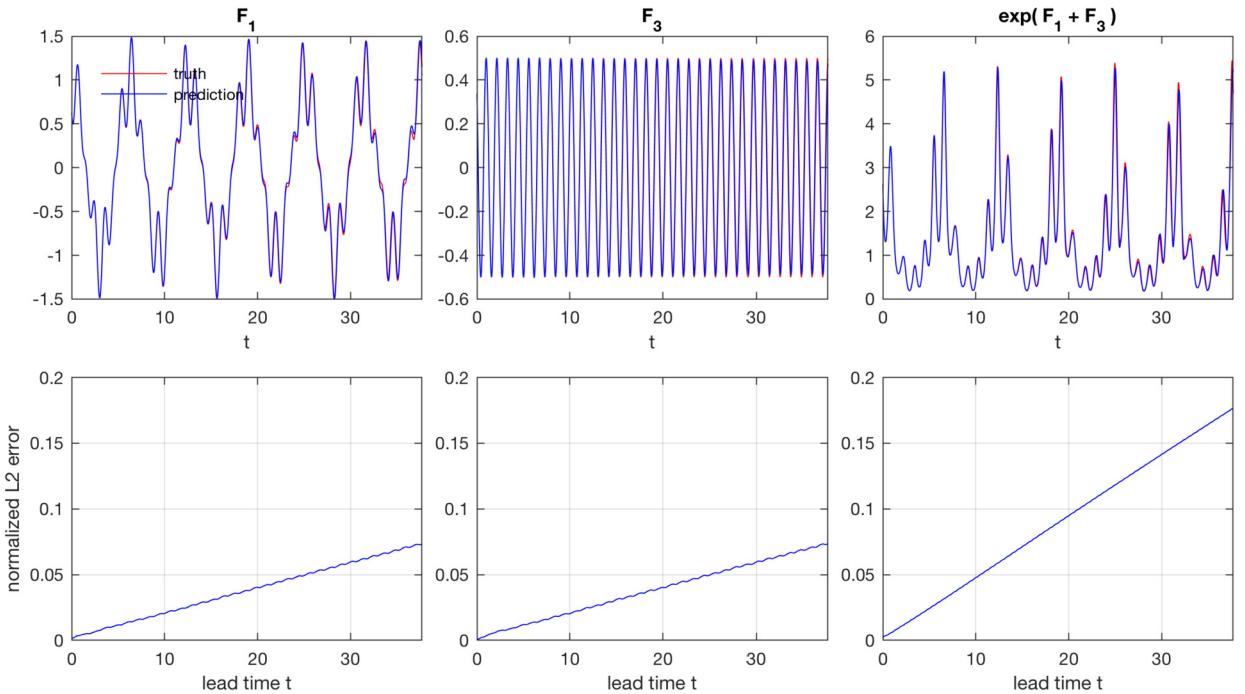
**Fig. 4.** Representative eigenfunctions  $\zeta_{\tau,j}$  of the data-driven generator  $W_{\tau}$  with  $\tau = 10^{-5}$  for the linear flow on the 2-torus. Top row: Scatterplots of  $\text{Re}(\zeta_{\tau,j})$  on the training dataset embedded in  $\mathbb{R}^3$ . Bottom row: Time series  $t_n \mapsto \text{Re}(\zeta_{\tau,j}(x_n))$  of the eigenfunctions, sampled along a portion of the dynamical trajectory in the training data. The numerical eigenfrequencies  $\omega_{\tau,j}$  and frequency-adjusted Dirichlet energies  $D_{\tau,j}$  are also indicated. The eigenfrequencies with  $j = 1, 5$ , and 9 shown here agree with the theoretical eigenfrequencies  $\alpha_1 = 1$ ,  $\alpha_2 \approx 5.477$ , and  $2\alpha_1 + 1\alpha_2 \approx 7.477$  to within four, three, and two significant figures, respectively.

Increasing  $L$  at fixed  $N$  is also expected to adversely affect the naive approximation, in this case not only due to the basis function errors just mentioned, but also because the spectrum of  $V$  is a dense subset of the imaginary line. That is, even with “perfect” basis functions, increasing  $L$  without regularization will result in the creation of near-degenerate eigenfrequencies around any reference eigenfrequency in the spectrum of the approximate generator, leading to high sensitivity to perturbations. Figs. 5(c) and 5(g) demonstrate that increasing  $L$  from 500 to 1000 in the  $N = 6400$  experiments results in considerable degradation of the eigenfunctions obtained from the naive approximation, whereas the eigenfunctions of the RKHS-regularized generator behave stably under this parameter change. In the case of the larger,  $N = 64,000$  dataset, a similar behavior is observed in Figs. 5(d) and 5(g) by increasing  $L$  to 10,000.

Turning now to forecasting, in Fig. 6 we show prediction results for the components  $F_1$  and  $F_3$  of the torus embedding into  $\mathbb{R}^3$ , as well as the observable  $e^{F_1+F_3}$ , which has a non-polynomial dependence on the components of the observation map (and in this case, the Koopman eigenfunctions). In all three cases, we use  $\tau = 10^{-5}$  and  $L = 500$ , and examine lead times  $t$  in the interval  $[0, 3000 \Delta t] \approx [0, 39]$ , which is approximately 26 times longer than the “fast” characteristic timescale  $2\pi/\alpha_2 \approx 1.15$  of the system. Over that interval, the normalized forecast errors  $\varepsilon(t)$  exhibit a linear error growth, remaining below 0.1 in the case of  $F_1, F_3$ , and below 0.2 in the case of  $e^{F_1+F_3}$ . The somewhat lower forecast skill for  $e^{F_1+F_3}$  is consistent with the fact that infinitely many Koopman eigenfrequencies are required to fully capture the dynamical evolution of this observable. As mentioned in Section 1, an advantageous aspect of the RKHS framework presented here over previous forecasting techniques operating on  $L^2$  spaces [12,17] is that it produces pointwise-evaluatable prediction functions, as opposed to expectation values with respect to probability measures with  $L^2$  densities (which must be supplied by the user as initial conditions). As illustrated in Fig. 6, the forecasts accurately reproduce the dynamical evolution of all three observables examined here.



**Fig. 5.** Real and imaginary parts of numerical Koopman eigenfunctions for the torus flow obtained from data-driven approximations of the generator without regularization (a–d) and the RKHS regularization  $W_\tau$  (e–h), for different dataset sizes  $N$  and values of the regularization and spectral resolution parameters  $\tau$  and  $L$ . The eigenfunctions depicted here are those whose corresponding eigenfrequency in the data-driven spectrum is closest to the theoretical eigenfrequency  $\alpha_1 = 1$ . For an exact approximation of a normalized Koopman eigenfunction, the numerical eigenfunctions should take values in the unit circle in the complex plane.



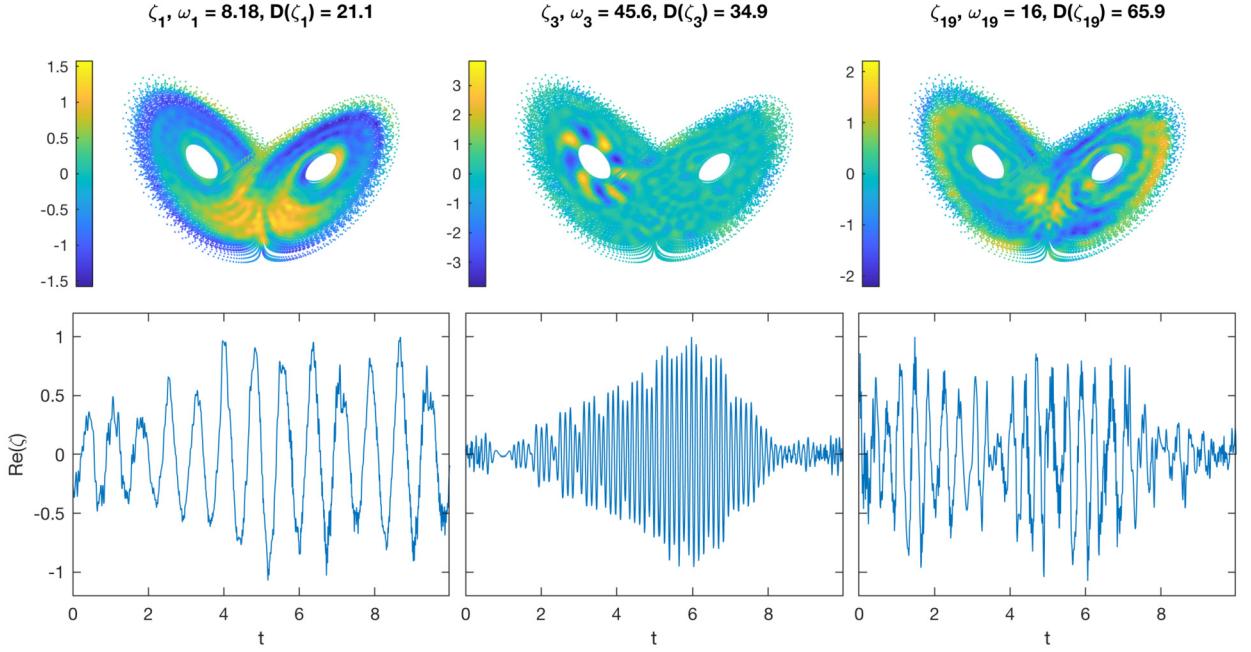
**Fig. 6.** Data-driven prediction of the components  $F_1$  and  $F_3$  of the embedding  $F$  of the 2-torus into  $\mathbb{R}^3$  (left and center columns), and the non-polynomial observable  $\exp(F_1 + F_3)$  (right column) for the linear torus flow, using the operator  $e^{tW_\tau}$  with  $\tau = 10^{-5}$ . Top row: Comparison of the true and predicted signals as a function of lead time  $t$  for a fixed initial condition in the verification dataset. Bottom row: Normalized RMSE  $\varepsilon(t)$  as a function of lead time.

*Lorenz 63 system.* For our standard choice of parameters, the L63 system is known to have a compact attractor  $X \subset \mathcal{M}$  in the state space  $\mathcal{M} = \mathbb{R}^3$  [55] with fractal dimension  $\approx 2.06$  [77], supporting a physical invariant measure  $\mu$ , which has a single positive Lyapunov exponent  $\Lambda \approx 0.91$  [78]. Due to dissipative dynamics, the attractor is contained within absorbing balls [79], playing here the role of the forward-invariant compact manifold  $M \supset X$ . The system is also rigorously known to be mixing [80], which implies that its associated Koopman unitary group on  $L^2(\mu)$  has no nonzero eigenfrequencies. Thus, the  $H_p$  subspace for this system is the one-dimensional space consisting of constant functions, while  $H_c$  contains all non-constant  $f \in L^2(\mu)$  with  $\int_M f d\mu = 0$ .

Fig. 3(b) shows the dependence of the eigenfrequencies of  $W_\tau$  for the L63 system, as well as the corresponding Dirichlet energies, on  $\tau \in [10^{-5}, 1]$ , computed using  $L = 750$  basis functions. As one might expect, the behavior of this spectrum is qualitatively different from that of the quasiperiodic torus flow in Fig. 3(a). That is, instead of the eigenfrequency curves of low Dirichlet energy interleaved with higher-Dirichlet-energy curves in Fig. 3(a), the eigenfrequencies in Fig. 3(b) exhibit an apparent continual growth in Dirichlet energy as  $\tau$  decreases to 0. This behavior is consistent with Corollary 3, according to which if the Dirichlet energy were to saturate along a sequence of eigenfrequencies of  $W_\tau$  as  $\tau \rightarrow 0^+$ , and that sequence had a nonzero limit, then that limit would necessarily be a nonzero Koopman eigenfrequency. As stated above, the latter is not possible for the L63 system. Nevertheless, upon visual inspection, one can identify in Fig. 3(b) frequency bands characterized by smaller Dirichlet energy than the surrounding frequencies; e.g., frequency bands centered at  $\omega \approx 8, 10, 20, 27$ , as well as higher frequencies. According to Corollary 3, the corresponding eigenfunctions of  $W_\tau$  are good candidates for coherent observables, evolving as approximate Koopman eigenfunctions, as we now verify.

Representative eigenfunctions  $\zeta_{\tau,j}$  chosen from these frequency bands for  $\tau = 10^{-4}$ , and visualized as scatterplots on the L63 attractor, as well as time series on the sampled dynamical trajectory, are displayed in Fig. 7. At least at the level of time series, the qualitative features of these eigenfunctions can be interpreted as generalizations of the Koopman eigenfunctions associated with the point spectra of measure-preserving ergodic dynamical systems. That is, similarly to Koopman eigenfunctions, the eigenfunctions of  $W_\tau$  in Fig. 7 are narrowband signals, evolving at a characteristic frequency determined from the corresponding eigenvalue  $\omega_{\tau,j}$ , and with  $\approx 90^\circ$  phase difference between their real and imaginary (not shown) parts. However, unlike true Koopman eigenfunctions, the oscillatory signals associated with  $\zeta_{\tau,j}$  exhibit pronounced amplitude modulations, giving them the appearance of wavepackets. If single-frequency, constant-amplitude, sinusoidal time series are to be thought of as hallmark features of Koopman eigenfunctions in measure-preserving systems, it appears that the eigenfunction time series of  $W_\tau$  shown in Fig. 7 lose the constancy of the amplitude, while maintaining a narrowband frequency character with high phase coherence between real and imaginary parts. In other words, these eigenfunctions reveal observables of the L63 system with an approximately cyclical behavior, despite mixing dynamics.

Despite the qualitative similarities of the corresponding time series, it is evident from Fig. 7 that the geometrical structure of the eigenfunctions of  $W_\tau$  on the L63 attractor may exhibit significantly different characteristics. For example, eigenfunction  $\zeta_{\tau,3}$  shown there (which corresponds to fairly high eigenfrequency,  $\omega_{\tau,3} \approx 46$ ) appears to be strongly localized on one of the two lobes of the L63 attractor, whereas eigenfunctions  $\zeta_{\tau,1}$  and  $\zeta_{\tau,19}$  (corresponding to lower eigenfrequencies,  $\omega_{\tau,1} \approx 8.2$  and  $\omega_{\tau,19} \approx 16$ , respectively) are supported on both lobes. Moreover, the level sets of  $\zeta_{\tau,3}$  are arranged in predominantly transverse directions to the dynamical flow, whereas those of  $\zeta_{\tau,1}$  and  $\zeta_{\tau,19}$  appear to be more aligned with the orbits of the dynamics. These differences are consistent with the fact that  $\omega_{\tau,3}$  is appreciably larger than  $\omega_{\tau,1}$ , as a more transverse arrangement of level sets relative to the orbits of the dynamics means that more contour crossings per unit time take place. It should be noted that an analogous eigenfunction to  $\zeta_{\tau,3}$ , but supported in the opposite lobe of the L63 attractor is also present in the spectrum of  $W_\tau$  (not shown here). It is also worthwhile noting that eigenfunction  $\zeta_{\tau,1}$  bears some qualitative similarities with the pattern depicted



**Fig. 7.** As in Fig. 4, but for eigenfunctions of the data-driven generator  $W_\tau$  with  $\tau = 10^{-4}$  for the L63 system. The eigenfunction time series in the lower panels have been scaled by their maximum absolute values so as to fit within the same axis limits. Observe the qualitatively different geometrical structure of the eigenfunctions on the Lorenz attractor. Despite these differences, the corresponding eigenfunction time series have the structure of amplitude-modulated wavetrains with a fairly distinct carrier frequency and lower-frequency modulating envelopes.

in [21, Figure 13]. Based on its corresponding eigenfrequency and level-set structure, eigenfunction  $\zeta_{\tau,19}$  resembles a second harmonic of  $\zeta_{\tau,1}$ .

Next, we consider forecasting experiments for the three components  $(F_1, F_2, F_3)$  of the observation map  $F$ , which coincide with the components of the L63 state vector in  $\mathbb{R}^3$ . We evaluate data-driven forecast functions for these observables at lead times in the interval  $[0, 500 \Delta t] = [0, 5]$ , using the regularization parameter  $\tau = 10^{-5}$  and  $L = 750$  basis functions. Representative forecast trajectories and the corresponding normalized  $L^2$  errors are displayed in Fig. 8. Unlike the linear error growth seen in Fig. 6 for the torus experiments, the L63 forecasts exhibit an exponential-like initial error growth, lasting for lead times up to  $t \simeq 0.7$ , and followed by a more gradual increase. The initial error growth period is somewhat shorter, though of the same order of magnitude, than the  $e$ -folding timescale associated with the system's positive Lyapunov exponent, i.e.,  $1/\Lambda \approx 1.1$ . In the case of observables  $F_1$  and  $F_2$ , the normalized  $L^2$  error  $\varepsilon(t)$  is seen to saturate around 1.4 as  $t$  approaches 5. Observable  $F_3$  exhibits a somewhat slower error growth than  $F_1$  and  $F_2$ , which may be a manifestation of dynamical symmetry of the L63 system under the transformation  $(x, y, z) \in \mathbb{R}^3 \mapsto (-x, -y, z)$ , making  $F_3$  a more predictable observable.

To interpret the long-time behavior of the error  $\varepsilon(t)$ , note that the Koopman operator of a mixing dynamical system such as L63 has the property that, as  $t \rightarrow \infty$ ,  $\langle g, U^t f \rangle_\mu$  converges to  $\langle g, 1 \rangle_\mu \langle 1, f \rangle_\mu$ . Based on this, it is possible to verify that, in this limit, the normalized  $L^2$  error  $\|(U^t - e^{t\tilde{V}_\tau})f\|_{L^2(\mu)} / \|f\|_{L^2(\mu)}$  associated with the quasiperiodic, unitary evolution group generated by  $\tilde{V}_\tau := \mathcal{U}_\tau^* W_\tau \mathcal{U}_\tau$  converges to  $\sqrt{2}$ . Now, our RKHS-based prediction scheme does not employ  $\tilde{V}_\tau$  directly, and as follows from Corollary 4, its error is governed by the non-unitary group generated by  $B_\tau$ , viz.,  $\|U^t f - e^{tB_\tau} P_\tau^* f_\epsilon\|_{L^2(\mu)} / \|f\|_{L^2(\mu)}$ . Nevertheless, for sufficiently small  $\tau$  and  $\epsilon$ ,  $e^{tB_\tau} P_\tau^* f_\epsilon$  can be made arbitrarily close to  $e^{t\tilde{V}_\tau} f$ , uniformly over compact time intervals. In that case,  $\varepsilon(t)$  would saturate close to  $\sqrt{2}$ , as observed in Fig. 8. It is worthwhile noting that, in the presence of mixing, the forecast functions derived from expectation values must necessarily converge to a constant equal to the mean with respect to the invariant measure of the dynamics. For instance, as  $t \rightarrow \infty$ , a forecast of the form  $\mathbb{E}_\rho U^t f = \langle \rho, U^t f \rangle_\mu$  [12,17], where  $\rho$  is a probability density in  $L^2(\mu)$ ,

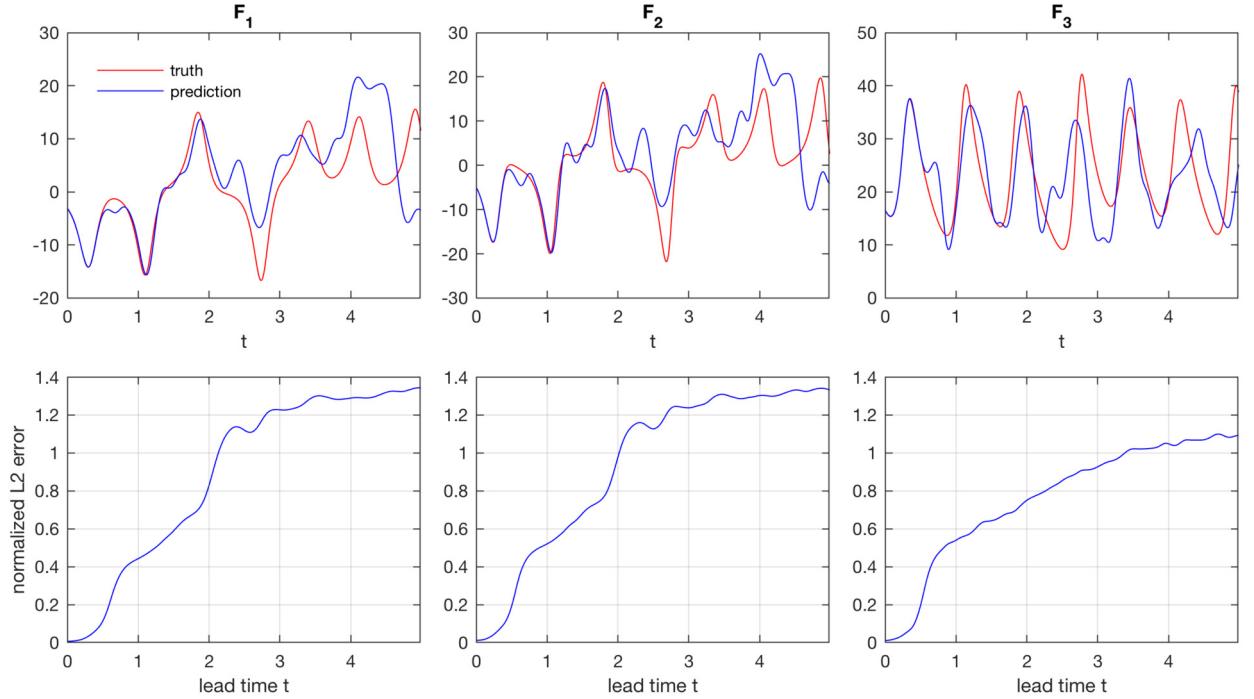
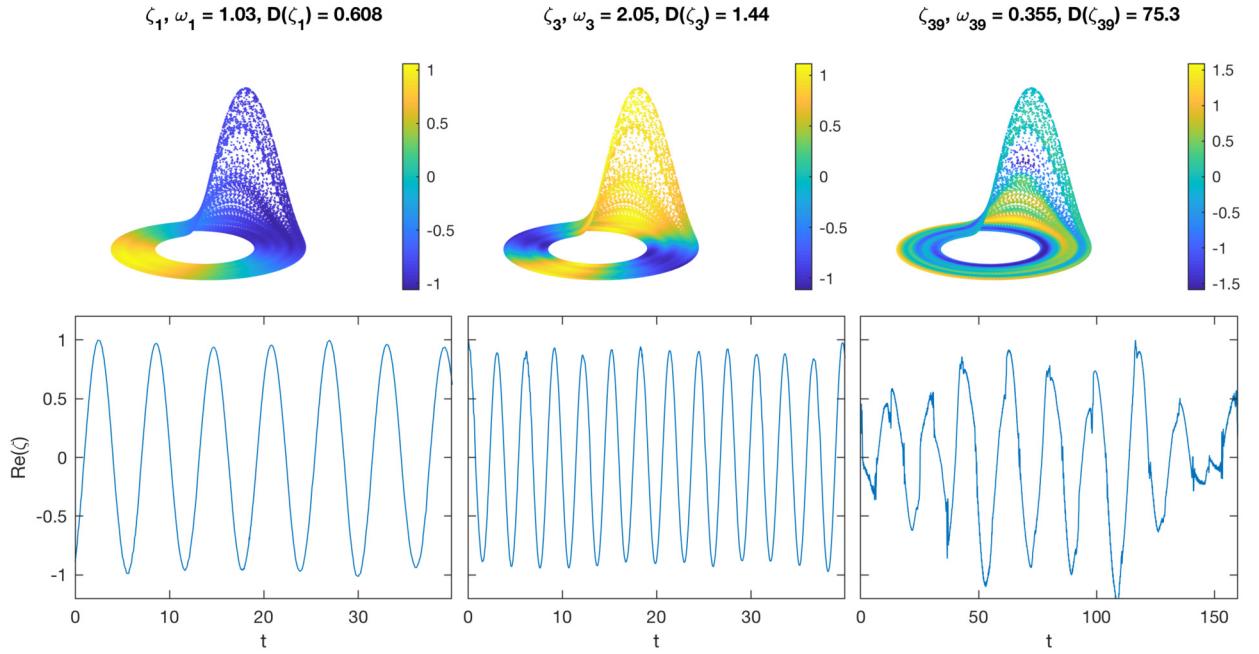


Fig. 8. As in Fig. 6, but for data-driven prediction of the components  $F_1$ ,  $F_2$ , and  $F_3$  of the L63 state vector.

satisfies  $\mathbb{E}_\rho U^t f \rightarrow \langle \rho, 1 \rangle_\mu \langle 1, U^t f \rangle_\mu = \int_M f d\mu$ . Such forecasts have asymptotic relative error  $\epsilon(t)$  equal to 1, i.e., smaller relative error than the quasiperiodic unitary evolution models constructed here, but arguably a constant prediction does not provide a realistic representation of the underlying dynamics. Indeed, as illustrated by the forecast trajectories in Fig. 8, the RKHS-based framework produces non-trivial, L63-like dynamics even at late times, when initial-value predictability has been lost. In that regard, the data-driven forecasts presented here are more akin to a “simulation” of L63 dynamics, as opposed to estimation of expectation values and/or other statistics.

*Rössler system.* The Rössler system is sometimes viewed as a simplified analog of the L63 system, as it only has a single quadratic nonlinearity, as opposed to two nonlinearities in the L63 system. Yet, despite the simplicity of its governing equations, it exhibits complex dynamical characteristics, some of which are not seen in the L63 system. For the standard choice of parameters listed above, one well known such feature is an outward spiraling motion in the  $z = 0$  plane about an unstable fixed point at  $(x, y, z) = (0, 0, 0) \in \mathbb{R}^3$ , which undergoes intermittent bursts to large positive  $z$  values when the radial coordinate  $r = \sqrt{x^2 + y^2}$  has become sufficiently large. This behavior produces a stiff signal in the  $z$  coordinate, as well as banding of trajectories in state space, which are challenging to model with data-driven approaches. Another notable aspect of the Rössler system is that chaotic behavior predominantly takes place in the  $(r, z)$  coordinates, whereas the evolution of the azimuthal angle in the  $z = 0$  plane proceeds at a near-constant angular frequency, approximately equal to 1 in natural time units. The Rössler system is also known to possess a single positive Lyapunov exponent, approximately equal to 0.071 [78]. While, to our knowledge, theorems on the existence and measure-theoretic mixing properties of the Rössler system analogous to [77,80] for the L63 system have not been established, the system has been studied extensively through analytical and numerical techniques, supporting the hypothesis that the Rössler system is indeed mixing, albeit at a slow rate [81].

In light of the above, it is perhaps not too surprising that the dependence of the eigenfrequencies of  $W_\tau$  for this system, depicted in Fig. 3(c) for  $L = 750$ , exhibits features reminiscent of both the torus and L63



**Fig. 9.** As in Fig. 4, but for eigenfunctions of  $W_\tau$ ,  $\tau = 10^{-5}$ , for the Rössler system.

spectra in Figs. 3(a) and 3(b), respectively. That is, the spectrum of  $W_\tau$  for the Rössler system exhibits bands of eigenfrequency curves with an apparent continual increase of Dirichlet energy with decreasing  $\tau$ , as in L63, but superposed to these curves is a set of eigenfrequencies at approximately integer multiples of a base frequency  $\alpha \simeq 1$ , and with near-constant corresponding Dirichlet energies, as in the linear torus flow. A visualization of corresponding eigenfunctions from the latter group, e.g., eigenfunctions  $\zeta_{\tau,1}$  and  $\zeta_{\tau,3}$  in Fig. 9 computed for  $\tau = 10^{-5}$ , reveals that these frequencies are indeed associated with highly coherent observables, which are predominantly functions of the azimuthal phase angle, and evolve near-periodically at integer multiples of the base frequency  $\alpha$ . Meanwhile, another group of eigenfrequencies of  $W_\tau$ , whose corresponding Dirichlet energies undergo a moderate increase with decreasing  $\tau$ , exhibit manifestly radial variability in state space and amplitude-modulated time series, reminiscent of the eigenfunctions recovered in the L63 system. Such an eigenfunction is  $\zeta_{\tau,39}$  shown in Fig. 9, whose corresponding eigenfrequency,  $\omega_{\tau,39} \approx 0.36$ , is smaller than the base frequency  $\alpha$ . On the basis of the eigenfrequency  $\omega_{\tau,39}$ , we can identify a characteristic timescale  $2\pi/\omega_{\tau,39} \approx 17.5$  of coherent radial oscillations of the Rössler system.

Next, Fig. 10 shows forecasting results for the components  $(F_1, F_2, F_3)$  of the Rössler state vector over lead times  $t \in [0, 2000 \Delta t] = [0, 80]$ , computed for  $\tau = 10^{-5}$  and  $L = 1000$ . Due to the dynamical behavior of the Rössler system outlined above, one would expect that predicting  $F_3$  is significantly more challenging than predicting  $F_1$  or  $F_2$ , and this is indeed reflected in the results in Fig. 10. In particular, consistent with the near-linear evolution of the azimuthal phase angle in the  $z = 0$  plane, prediction of the observation map components  $F_1$  and  $F_2$  remains skillful for the entire forecast interval examined, with the normalized error  $\varepsilon(t)$  exhibiting a gradual increase to 0.25 by  $t \simeq 80$ . An inspection of the individual forecast trajectories shown in Fig. 10 indicates that the errors in these forecasts are predominantly amplitude errors (as opposed to phase errors), likely caused by chaotic dynamics of the radial coordinate  $r$ . On the other hand, forecasts of the  $F_3$  component exhibit a significantly more rapid error growth, reaching  $\varepsilon(t) \simeq 1.15$  as  $t$  approaches 80. This error can be understood from the highly stiff, intermittent nature of  $F_3$ , exhibiting infrequent excursions to large positive values and virtually no negative values. As is evident from the forecast trajectory in Fig. 10, the data-driven forecasts are generally successful in capturing the timing of the  $F_3$  bursts (likely aided by the high coherence of the azimuthal phase angle), but for lead times  $t \gtrsim 20$ , they struggle to reproduce the

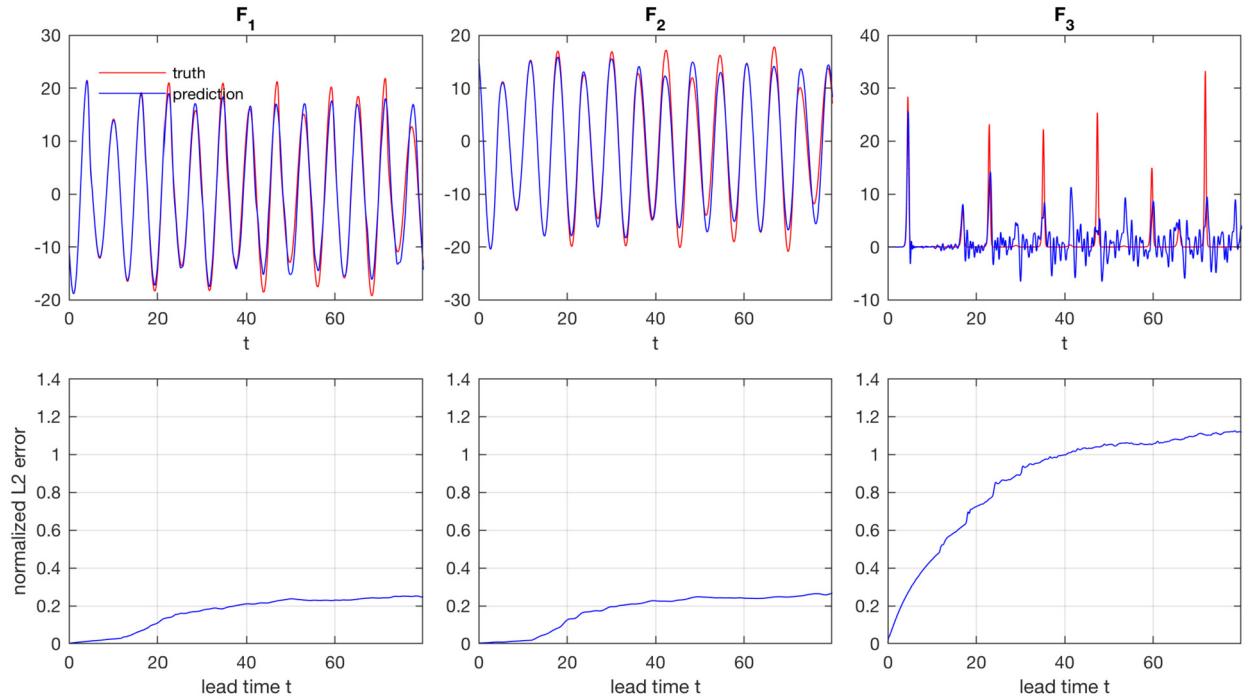


Fig. 10. As in Fig. 8 but for forecasts of the state-vector components of the Rössler system, using  $\tau = 10^{-5}$  and  $L = 7000$ .

amplitude of the bursts and the non-negativity of the  $F_3$  signal. In separate calculations, we have verified that the non-negativity of the forecast signal over a given time interval can be improved by increasing the number of basis functions  $L$ .

## 10. Concluding remarks

In this paper, we have developed a data-driven framework for spectral analysis of measure-preserving, ergodic dynamical systems, using ideas from RKHS theory. A central element of our approach has been to regularize the unbounded, skew-adjoint generator of the unitary Koopman group of the system by pre- and post-composing it with integral operators associated with reproducing kernels of RKHSs, rendering it into a compact operator. We showed that if this procedure is carried out using a one-parameter family of Markov kernels of appropriate ( $C^1$ ) regularity, the resulting regularized generators form a one-parameter family of trace-class, skew-adjoint integral operators  $W_\tau$  on RKHS, converging to the Koopman generator in strong resolvent sense in a limit of vanishing regularization parameter  $\tau$ . As a result, at every  $\tau > 0$ ,  $W_\tau$  can be spectrally decomposed in terms of a purely atomic projection-valued measure (PVM), with an associated discrete set of eigenfrequencies and an orthonormal basis of eigenfunctions, converging to the PVM of the Koopman generator as  $\tau \rightarrow 0^+$  in an appropriate sense. Notably, this result holds for measure-preserving ergodic systems of arbitrary spectral characteristics (pure point, continuous, mixed), and further allows consistent approximation of the functional calculus of the Koopman generator for bounded continuous functions. In particular, exponentiation of the regularized generator leads to a unitary, quasiperiodic evolution group,  $e^{tW_\tau}$ , which can be used as an approximation of the Koopman group of the system to perform forecasting of observables with convergence guarantees. We also showed that the eigenfunctions associated with this group form coherent observables lying in the approximate point spectrum of the Koopman operator, generalizing the coherent patterns associated with Koopman eigenfunctions and the point spectrum of the system.

Another advantageous aspect of the RKHS framework is that it naturally lends itself to data-driven approximation from time-ordered measurements of the system state taken through injective observation maps, requiring little structural modification of the continuous formulation. In particular, the data-driven approximation schemes employ properties of physical measures to consistently approximate integrals with respect to the invariant measure by time averages, and take advantage of RKHS regularity to approximate the action of the generator on functions by temporal finite differences. Coupled with the ability afforded by RKHSs to perform interpolation and out-of-sample evaluation, this approach leads to data-driven forecast functions for the evolution of observables, as well as coherent eigenfunctions, whose robustness can be assessed a posteriori through a Dirichlet energy criterion.

We demonstrated the efficacy of this approach through a suite of coherent pattern extraction and forecasting experiments in the setting of a quasiperiodic flow on the 2-torus and the chaotic L63 and Rössler systems. In the case of the torus rotation, the eigenfrequencies of the RKHS-regularized operator  $W_\tau$  correctly identify generating eigenfrequencies of the system, as well as integer combinations of such eigenfrequencies. Meanwhile, in the L63 and Rössler settings, eigenfunctions of  $W_\tau$  identified via the Dirichlet energy criterion exhibit an approximately cyclical evolution, behaving as approximate Koopman eigenfunctions. These eigenfunctions reveal coherent oscillatory observables of these systems with characteristic timescales determined from the corresponding eigenvalues, despite potentially mixing dynamics. Forecasting using the evolution group generated by  $W_\tau$  was found to perform well in these systems, with skill likely aided by the presence of approximately periodic eigenfunctions in the respective spectra.

Areas of future research stemming from this work include improved representations of the generator through alternative schemes to finite differences, as well as extensions to partially observed systems (i.e., non-injective observation maps). In addition, the fact that the spectral convergence results in Theorem 2 require pointwise convergence of the approximating operators only on a core of the generator  $V$ , yet in Proposition 19 we were able to establish pointwise convergence on the full domain  $D(V)$ , suggests that it may be possible to weaken the  $C^1$  regularity assumptions on the kernels and their associated RKHSs underlying Theorem 2. It would also be fruitful to explore formulations of the framework presented here utilizing methods for kernel learning [82,83] to optimize prediction skill of prescribed observables. Meanwhile, the approximately cyclical nature of the identified eigenfunctions in the L63 and Rössler systems suggests possible connections between the spectral properties of  $W_\tau$  and periodic orbits of the underlying flow in state space expected for non-uniformly hyperbolic dynamics. Finally, a topic of significant interest in both the Koopman and transfer operator literature is spectral analysis and forecasting of dissipative and/or non-ergodic systems [4,10,84,85]. While some of the spectral approximation techniques employed in this work make use of the skew-adjoint structure of the generator of measure-preserving systems (e.g., strong convergence in a core in Lemma 14), and we have also made use of ergodicity to establish correspondences between the spectra of various types of regularized generators (e.g., Lemma 18), it would nevertheless be fruitful to explore applications of RKHS theory to spectral analysis of such “open dynamical systems”, extending the framework developed here beyond the measure-preserving, ergodic setting.

## Acknowledgments

Dimitrios Giannakis acknowledges support by ONR YIP grant N00014-16-1-2649, ONR MURI grant N00014-19-1-2421, NSF grants DMS-1521775, DMS-1854383, 1842538, and DARPA grant HR0011-16-C-0116. Sudhhasattwa Das was supported as a postdoctoral research fellow from the first two grants. Joanna Slawinska acknowledges support from NSF grants 1551489 and 1842538. The authors would like to thank Igor Mezić for pointing out a possible connection between the results of Theorem 2 and the pseudospectrum of the Koopman operator, which led to Corollary 3.

## Appendix A. Variable-bandwidth kernels

The numerical experiments in Section 9 were performed using variable-bandwidth Gaussian kernels  $\kappa_N : Y \times Y \rightarrow \mathbb{R}$  of the form

$$\kappa_N(y, y') = \exp\left(-\frac{d^2(y, y')}{\epsilon\sigma_N(y)\sigma_N(y')}\right). \quad (\text{A.1})$$

Here,  $\sigma_N : Y \rightarrow \mathbb{R}$  is a strictly-positive,  $C^1$  function on  $Y$ , which generally depends on the training dataset  $\{y_0, \dots, y_{N-1}\}$ . We indicate this dependence with  $N$  subscripts. Intuitively, the role of the bandwidth function  $\sigma_N$  is to correct for variations in the “sampling density” of the data. In particular, for a well conditioned kernel integral operator  $G_N$ , the number of datapoints lying within radius  $O(\epsilon^{1/2})$  balls centered at each datapoint should not exhibit significant variations across the dataset, yet, the standard radial Gaussian kernel from (46) has no mechanism for preventing this from happening. For appropriately chosen  $\sigma_N$ , the variable-bandwidth kernel in (A.1) can, in effect, vary the radii of these balls to help improve conditioning. The different bandwidth functions proposed in the literature include near-neighbor distances [86] and kernel density estimates [43]. In the numerical experiments of Section 9, we will employ the latter approach, defining

$$\sigma_N(y) = \rho_N^{-1/\tilde{m}}(y), \quad \rho_N(y) = \frac{1}{(\pi\tilde{\epsilon})^{\tilde{m}/2}} \int_Y e^{-d^2(y, y')/\tilde{\epsilon}} d\tilde{\mu}_N(y). \quad (\text{A.2})$$

Here,  $\tilde{\mu}_N = \sum_{n=0}^{N-1} \delta_{y_n}/N$  is the sampling measure in data space,  $\tilde{\epsilon}$  a positive bandwidth parameter (different from  $\epsilon$  in (A.1)), and  $\tilde{m}$  a positive parameter approximating the dimension of  $F(X)$ . The parameters  $\epsilon$ ,  $\tilde{\epsilon}$ , and  $\tilde{m}$  are all determined from the data automatically; see [12,17] for descriptions of this procedure.

If  $F(X)$  has the structure of a Riemannian submanifold of  $Y$ , and the pushforward  $\tilde{\mu}$  on of the invariant measure on  $Y$  has a smooth density, the functions  $\rho_N$  from (A.2) are estimates of the sampling density  $\rho = d\tilde{\mu}/d\text{vol}$ , which converge in the limit of  $N \rightarrow \infty$  followed by  $\tilde{\epsilon} \rightarrow 0$ . Thus, with this choice of bandwidth function, the bandwidth of the kernel  $\kappa_N$  from (A.1) will be large (small) when the sampling density is small (large), achieving the desired balancing of the kernel. More quantitatively, with this choice of bandwidth functions and after suitable normalization,  $\kappa_N$  approximates the heat kernel of a conformally transformed Riemannian metric on  $F(X)$ , whose volume form has uniform density relative to  $\tilde{\mu}$  [17]. Of course, if  $F(X)$  does not have manifold structure, or  $\rho$  is not smooth, this Riemannian geometric interpretation is not applicable, but the balancing effect of the bandwidth functions on local balls still holds. It should be noted that one can prove spectral convergence results analogous to Lemma 23(i) for the class of  $N$ -dependent kernels on  $M$  induced by  $\kappa_N$ ; see [87] for such a result. Here, we will omit a proof of spectral convergence for the integral operators associated with  $\kappa_N$  in the interest of brevity. It is also important to note that, to our knowledge, it has not been established whether the kernels on  $M$  induced by  $\kappa_N$ , and the kernel that they converge to as  $N \rightarrow \infty$ , are  $L^2(\mu_N)$ - and  $L^2(\mu)$ -strictly-positive, respectively. That being said, we did not find evidence of zero eigenvalues of  $G_N$  in the experiments of Section 9.

## Appendix B. Pseudocode

In this appendix, we provide pseudocode listings for the techniques described in Section 8. We have split the entire process into four algorithms, the first two of which describe the construction of the data-driven eigenpairs  $(\lambda_{N,j}, \phi_{N,j})$  from Lemma 23 and pointwise evaluation of the corresponding basis functions  $\psi_{N,j}$  of  $\mathcal{H}_N$ , respectively. Algorithm 3 describes the construction of the data-driven generator  $W_{\tau, N, \Delta t}^{(L)}$  from

Theorem 21(i) and computation of its associated eigenvalues and eigenfunctions. Algorithm 4 describes the construction and pointwise (out-of-sample) evaluation of the data-driven forecast function  $f_{\tau, N, \Delta t, L, L'}^{(t)}$  from Corollary 22. In what follows,  $\vec{1}$  will denote the  $N$ -dimensional column vector whose elements are all equal to 1. Moreover, the indexing of all vector and matrix elements will start from 0.

We begin by listing Algorithm 1 for a general kernel  $\kappa$  on data space  $Y$  of the form in (34), evaluated on a time series of the values of the observation map  $F$  on a dynamical trajectory  $x_0, \dots, x_{N-1}$  in  $M$ . As stated in Sections 8 and 9, we work with the variable-bandwidth Gaussian kernel described in Appendix A. Evaluation of this kernel requires a kernel density estimation step, summarized in [17, Algorithm 1]. The variable-bandwidth Gaussian kernel also requires specification of the bandwidth parameter  $\epsilon$ , as well as the bandwidth and dimension parameters  $\tilde{\epsilon}$  and  $\tilde{m}$ , respectively, in (A.2). We set these parameters automatically via the procedure described in [12, Appendix A] and [17, Algorithm 1]. The main outputs of Algorithm 1 are the eigenpairs  $(\lambda_{N,j}, \phi_{N,j})$  of the Markov operator  $G_N$  associated with the Markov kernel  $p_N$ , obtained via the bistochastic normalization procedure from Section 4.1. Due to the  $L^2(\mu_N) \simeq \mathbb{C}^N$  isomorphism,  $G_N$  can be represented by an  $N \times N$  bistochastic matrix  $\mathbf{G}$  with elements  $G_{ij} = p_N(x_i, x_j)/N$ , and the eigenvectors  $\phi_{N,j}$  (of unit  $L^2(\mu_N)$  norm) by  $N$ -dimensional column vectors  $\vec{\phi}_j = (\phi_{N,j}(x_0), \dots, \phi_{N,j}(x_{N-1}))^\top$  with 2-norm  $\|\vec{\phi}_j\|_2 = \sqrt{N}$ . We will abbreviate  $\lambda_{N,j}$  by  $\lambda_j$ . The eigenpairs  $(\lambda_j, \vec{\phi}_j)$  can be computed without explicit formation of  $\mathbf{G}$ , owing to the fact that  $\mathbf{G} = \tilde{\mathbf{K}} \tilde{\mathbf{K}}^\top$ , where  $\tilde{\mathbf{K}}$  is a non-symmetric  $N \times N$  kernel matrix to be defined in Algorithm 1. In particular, the  $\lambda_j$  are equal to the squared singular values of  $\tilde{\mathbf{K}}$ , and the  $\vec{\phi}_j$  are equal to the corresponding left singular vectors. Algorithm 1 also outputs as auxiliary outputs the corresponding right singular vectors  $\vec{\gamma}_j \in \mathbb{R}^N$  of  $\tilde{\mathbf{K}}$  and a degree vector  $\vec{q} \in \mathbb{R}^N$  associated with that matrix; these outputs will be used for pointwise evaluation in Algorithm 2.

**Algorithm 1** (Data-driven basis).

- Inputs
  - Time series  $F(x_0), \dots, F(x_{N-1})$  in data space  $Y$ .
  - Number  $L \leq N$  of eigenpairs to be computed.
- Outputs
  - Leading  $L$  eigenvalues  $\lambda_0, \dots, \lambda_{L-1}$  of  $\mathbf{G}$  and the corresponding eigenvectors  $\vec{\phi}_0, \dots, \vec{\phi}_{L-1} \in \mathbb{R}^N$ .
  - Degree vector  $\vec{q} \in \mathbb{R}^N$ .
  - Right singular vectors  $\vec{\gamma}_0, \dots, \vec{\gamma}_{L-1} \in \mathbb{R}^N$ .
- Steps
  1. Compute the  $N \times N$  kernel matrix  $\mathbf{K}$  with  $K_{ij} = \kappa(F(x_i), F(x_j))$ .
  2. Compute the  $N$ -dimensional degree vectors  $\vec{d} = \mathbf{K} \vec{1}$  and  $\vec{q} = \mathbf{K} \mathbf{D}^{-1} \vec{1}$ , where  $\mathbf{D} = \text{diag } \vec{d}$ .
  3. Form the  $N \times N$  kernel matrix  $\tilde{\mathbf{K}} = \mathbf{D}^{-1} \mathbf{K} \mathbf{Q}^{-1/2}$ , with  $\mathbf{Q} = \text{diag } \vec{q}$ .
  4. Compute the  $L$  largest singular values  $\sigma_0, \dots, \sigma_{L-1}$  of  $\tilde{\mathbf{K}}$ , and set  $\lambda_j = \sigma_j^2$ . Set  $\vec{\phi}_j$  and  $\vec{\gamma}_j$  to the corresponding left and right singular vectors, respectively, normalized such that  $\|\vec{\phi}_j\|_2 = \sqrt{N}$ .

Next, Algorithm 2 carries out the task of evaluating the RKHS functions  $\psi_{N,j} \in \mathcal{H}_N$  at an arbitrary collection  $\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{\hat{N}-1}$  of points in  $M$ , given the corresponding values  $F(\hat{x}_0), F(\hat{x}_1), \dots, F(\hat{x}_{\hat{N}-1})$  of the observation map  $F$ . As with Algorithm 1, this computation can be performed without explicit formation of a kernel matrix associated with  $p_N$ , using instead the singular vectors  $\vec{\phi}_0, \dots, \vec{\phi}_{L-1}$  and  $\vec{\gamma}_0, \dots, \vec{\gamma}_{L-1}$ . In what follows, we use the column vectors  $\vec{\psi}_j = (\psi_{N,j}(\hat{x}_0), \dots, \psi_{N,j}(\hat{x}_{N-1}))^\top \in \mathbb{R}^{\hat{N}}$  to represent the values of the  $\psi_{N,j}$  at the desired points. Note that in the case of the variable-bandwidth Gaussian kernels from Appendix A, the computation of the  $\vec{\psi}_j$  requires an additional density estimation step for the out-of-sample data  $F(\hat{x}_n)$ , which is carried out analogously to [17, Algorithm 1]. Moreover, all kernel parameters  $\epsilon$ ,  $\tilde{\epsilon}$ , and  $\tilde{m}$  are the same as those used in Algorithm 1.

**Algorithm 2** (*Pointwise evaluation in RKHS*).

- Inputs
  - Values  $F(\hat{x}_0), \dots, F(\hat{x}_{\hat{N}-1})$  of the observation map at the evaluation points.
  - Eigenvalues  $\lambda_0, \dots, \lambda_{L-1}$ , eigenvectors  $\vec{\phi}_0, \dots, \vec{\phi}_{L-1}$ , right singular vectors  $\vec{\gamma}_0, \dots, \vec{\gamma}_{L-1}$ , and degree vector  $\vec{q}$  from Algorithm 1.
- Outputs
  - Vectors  $\vec{\psi}_0, \dots, \vec{\psi}_{L-1} \in \mathbb{R}^{\hat{N}}$  with the values of the RKHS functions  $\psi_{N,0}, \dots, \psi_{N,j}$  at the evaluation points.
- Steps
  1. Compute the  $\hat{N} \times N$  kernel matrix  $\hat{\mathbf{K}}$  with  $\hat{K}_{ij} = \kappa(F(\hat{x}_i), F(x_j))$ .
  2. Compute the  $\hat{N}$ -dimensional degree vector  $\hat{d} = \hat{\mathbf{K}}\vec{1}$ .
  3. Form the  $\hat{N} \times N$  kernel matrix  $\bar{\mathbf{K}} = \hat{\mathbf{D}}^{-1} \hat{\mathbf{K}} \mathbf{Q}^{-1/2}$ , where  $\hat{\mathbf{D}} = \text{diag } \hat{d}$  and  $\mathbf{Q} = \text{diag } \vec{q}$ .
  4. Output  $\vec{\psi}_j = \bar{\mathbf{K}} \vec{\gamma}_j$ .

Note that when working with Gaussian kernels, as done throughout this paper, we approximate the kernel matrices  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$ ,  $\hat{\mathbf{K}}$  and  $\bar{\mathbf{K}}$  in Algorithms 1 and 2 by sparse matrices (as is common practice), retaining in each case the  $k_{\text{nn}}$  largest entries per row. In the numerical experiments of Section 9,  $k_{\text{nn}}$  was approximately 8% of  $N$ .

We now describe how to construct an  $L \times L$  matrix  $\mathbf{W}$  representing the data-driven generator  $W_{\tau,N,\Delta t}^{(L)}$  in the  $\psi_{\tau,N,j}$  basis of  $\mathcal{H}_{\tau,N}$ , and use that matrix to compute the  $(\omega_{\tau,N,\Delta t,j}^{(L)}, \zeta_{\tau,N,\Delta t,j}^{(L)})$  eigenpairs. We represent each eigenvector  $\zeta_{\tau,N,\Delta t,j}^{(L)} \in \mathcal{H}_{\tau,N}$  by a column vector  $\vec{\xi}_j = (\xi_{0,j}, \dots, \xi_{L-1,j})^\top \in \mathbb{C}^L$  storing the expansion coefficients of  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  in the  $\psi_{\tau,N,j}$  basis, i.e.,  $\zeta_{\tau,N,\Delta t,j}^{(L)} = \sum_{i=0}^{L-1} \xi_{i,j} \psi_{\tau,N,j}$ . Given a set  $\{\hat{x}_0, \dots, \hat{x}_{\hat{N}-1}\}$  of evaluation points in  $M$ , the values  $\zeta_{\tau,N,\Delta t,j}^{(L)}(\hat{x}_n)$  will be represented by the column vectors  $\vec{\zeta}_j = (\zeta_{\tau,N,\Delta t,j}^{(L)}(\hat{x}_0), \dots, \zeta_{\tau,N,\Delta t,j}^{(L)}(\hat{x}_{\hat{N}-1}))^\top \in \mathbb{C}^{\hat{N}}$ . In Algorithm 3 below, we describe the construction of  $\mathbf{W}$  and the computation of the  $\omega_j$ ,  $\vec{\xi}_j$ , and  $\vec{\zeta}_j$ , using the central finite-difference scheme from (35) to approximate the action of the generator. The algorithm can also be implemented using any skew-adjoint finite-difference scheme of appropriate regularity. Moreover, we employ the basis functions and pointwise evaluation procedures from Algorithms 1 and 2, associated with the bistochastic kernel normalization in Section 4.1, but Algorithm 3 can be implemented using any other Markov operator meeting the conditions of Theorem 21. Algorithm 3 also returns the frequency-adjusted Dirichlet energies  $\mathcal{D}_{N,\Delta t}(\zeta_{\tau,N,\Delta t,j}^{(L)})$  of the eigenfunctions from (45), abbreviated  $\mathcal{D}_j$ . We also abbreviate  $\omega_{\tau,N,\Delta t,j}^{(L)}$  by  $\omega_j$ .

**Algorithm 3** (*Data-driven generator and its eigendecomposition*).

- Inputs
  - RKHS regularization parameter  $\tau > 0$ .
  - Time step  $\Delta t > 0$ .
  - Eigenvalues  $\lambda_0, \dots, \lambda_{L-1}$ , eigenvectors  $\vec{\phi}_0, \dots, \vec{\phi}_{L-1}$ , right singular vectors  $\vec{\gamma}_0, \dots, \vec{\gamma}_{L-1}$ , and degree vector  $\vec{q}$  from Algorithm 1.
  - Pointwise-evaluated RKHS functions  $\vec{\psi}_0, \dots, \vec{\psi}_{L-1}$  from Algorithm 2.
- Outputs
  - Eigenfrequencies  $\omega_0, \dots, \omega_{L-1} \in \mathbb{R}$ , the corresponding eigenvectors  $\vec{\xi}_0, \dots, \vec{\xi}_{L-1} \in \mathbb{C}^L$ , and the Dirichlet energies  $\mathcal{D}_0, \dots, \mathcal{D}_{L-1} \geq 0$ .
  - Vectors  $\vec{\zeta}_0, \dots, \vec{\zeta}_{L-1} \in \mathbb{C}^{\hat{N}}$  with the values of the eigenfunctions  $\zeta_{\tau,N,\Delta t,j}^{(L)}$  at the evaluation points.
- Steps

1. Construct the  $L \times L$  diagonal matrix  $\tilde{\Lambda}$ , with  $\tilde{\Lambda}_{jj} = e^{\tau(1-\lambda_j^{-1})}$ , and the  $N \times L$  matrix  $\Phi$ , whose  $j$ -th column is equal to  $\vec{\phi}_j$ .
2. Form the skew-symmetric, tridiagonal,  $N \times N$  finite-difference matrix  $\mathbf{V}$  with

$$2\Delta t \mathbf{V} = \begin{pmatrix} 0 & \frac{1}{2} & & & \\ -\frac{1}{2} & 0 & 1 & & \\ & -1 & 0 & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 0 & 1 \\ & & & & -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

3. Compute the  $L \times L$  skew-symmetric matrix  $\mathbf{W} = \tilde{\Lambda}^{1/2} \Phi^\top \mathbf{V} \Phi \tilde{\Lambda}^{1/2}$ .
4. Set the eigenfrequencies  $\omega_0, \dots, \omega_{L-1}$  to the imaginary parts of the eigenvalues of  $\mathbf{W}$ . Set  $\vec{\xi}_j$  to the corresponding eigenvectors, normalized such that  $\|\vec{\xi}_j\|_2 = \sqrt{N}$ .
5. For each eigenvector  $\vec{\xi}_j$ , compute the Dirichlet energy

$$\mathcal{D}_j = \left( \frac{\|\tilde{\Lambda}^{1/2} \Phi^{-1/2} \vec{\xi}_j\|_2^2}{\|\tilde{\Lambda}^{1/2} \vec{\xi}_j\|_2^2} - 1 \right) (1 - (\omega_j \Delta t)^2)^{-1}, \quad \Lambda = \text{diag}(\lambda_0, \dots, \lambda_{L-1}).$$

6. Form the  $\hat{N} \times L$  matrix  $\Psi$ , whose  $j$ -th column is equal to  $\vec{\psi}_j$ , and set  $\vec{\zeta}_j = \Psi \vec{\xi}_j$ .

Finally, Algorithm 4 computes the values of the data-driven forecast function  $f_{\tau, N, \Delta t, L, L'}^{(t)}$  from Corollary 22 for lead time  $t \geq 0$  at a set of evaluation points  $\{\hat{x}_0, \dots, \hat{x}_{\hat{N}}\} \subset M$ , using the output of Algorithm 3 and the values  $f(x_0), \dots, f(x_{N-1})$  of the prediction observable  $f$  on the dynamical trajectory  $x_0, \dots, x_{N-1}$ . The forecast values are output as a column vector  $\hat{f} = (f_{\tau, N, \Delta t, L, L'}^{(t)}(\hat{x}_0), \dots, f_{\tau, N, \Delta t, L, L'}^{(t)}(\hat{x}_{\hat{N}-1}))^\top \in \mathbb{C}^{\hat{N}}$ . Note that a similar approach can be employed to evaluate the approximations in Theorem 21(iii) for general bounded continuous functions  $Z : i\mathbb{R} \rightarrow \mathbb{C}$ .

**Algorithm 4** (*Data-driven prediction*).

- Inputs
  - Lead time  $t \geq 0$ .
  - Number of basis functions  $L' \leq L$ .
  - Time series  $f(x_0), \dots, f(x_{N-1}) \in \mathbb{C}$  of the prediction observable.
  - Eigenvalues  $\lambda_0, \dots, \lambda_{L-1}$  and eigenvectors  $\vec{\phi}_0, \dots, \vec{\phi}_{L-1}$  from Algorithm 1.
  - Eigenfrequencies  $\omega_0, \dots, \omega_{L-1}$ , eigenvectors  $\vec{\xi}_0, \dots, \vec{\xi}_{L-1}$ , and pointwise-evaluated eigenfunctions  $\vec{\zeta}_0, \dots, \vec{\zeta}_{L-1}$  from Algorithm 3.
- Outputs
  - Column vector  $\hat{f} \in \mathbb{C}^{\hat{N}}$  with the values of the forecast function for  $U^t f$  at the evaluation points.
- Steps
  1. Form the column vector of observable values  $\vec{f} = (f(x_0), \dots, f(x_{N-1}))^\top \in \mathbb{C}^N$ .
  2. Compute the column vector of expansion coefficients  $\vec{c} = (c_0, \dots, c_{L-1})^\top \in \mathbb{C}^L$ , where

$$c_j = \begin{cases} \vec{\phi}_j^\top \vec{f} / (N \lambda_j^{1/2}), & j \leq L', \\ 0, & \text{otherwise.} \end{cases}$$

3. Form the  $L \times L$  diagonal matrix  $\mathbf{U} = \text{diag}(1, e^{i\omega_1 t}, \dots, e^{i\omega_{L-1} t})$ , the  $\hat{N} \times L$  eigenfunction matrix  $\mathbf{Z}$  whose  $j$ -th column is equal to  $\vec{\zeta}_j$ , and set  $\hat{f} = \mathbf{Z} \mathbf{U} \vec{c}$ .

## References

- [1] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal. 36 (1999) 491, <https://doi.org/10.1137/S0036142996313002>.
- [2] M. Dellnitz, G. Froyland, On the isolated spectrum of the Perron-Frobenius operator, Nonlinearity 13 (4) (2000) 1171–1188, <https://doi.org/10.1088/0951-7715/13/4/310>.
- [3] I. Mezić, A. Banaszuk, Comparison of systems with complex behavior, Physica D 197 (2004) 101–133, <https://doi.org/10.1016/j.physd.2004.06.015>.
- [4] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, Nonlinear Dyn. 41 (2005) 309–325, <https://doi.org/10.1007/s11071-005-2824-x>.
- [5] G. Froyland, K. Padberg, Almost-invariant sets and invariant manifolds – connecting probabilistic and geometric descriptions of coherent structures in flows, Physica D 238 (2009) 1507–1523, <https://doi.org/10.1016/j.physd.2009.03.002>.
- [6] C.W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D.S. Henningson, Spectral analysis of nonlinear flows, J. Fluid Mech. 641 (2009) 115–127, <https://doi.org/10.1017/s0022112009992059>.
- [7] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28, <https://doi.org/10.1017/S0022112010001217>.
- [8] G. Froyland, An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems, Physica D 250 (2013) 1–19, <https://doi.org/10.1016/j.physd.2013.01.013>.
- [9] 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, <https://doi.org/10.3934/jcd.2014.1.249>.
- [10] G. Froyland, G.A. Gottwald, A. Hammerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, SIAM J. Appl. Dyn. Syst. 13 (4) (2014) 1816–1846, <https://doi.org/10.1137/130943637>.
- [11] 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, <https://doi.org/10.3934/jcd.2014.1.391>.
- [12] T. Berry, D. Giannakis, J. Harlim, Nonparametric forecasting of low-dimensional dynamical systems, Phys. Rev. E 91 (2015) 032915, <https://doi.org/10.1103/PhysRevE.91.032915>.
- [13] D. Giannakis, J. Sławińska, Z. Zhao, Spatiotemporal feature extraction with data-driven Koopman operators, J. Mach. Learn. Res. Proc. 44 (2015) 103–115.
- [14] 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), <https://doi.org/10.1007/s00332-015-9258-5>.
- [15] S.L. Brunton, B.W. Brunton, J.L. Proctor, E. Kaiser, J.N. Kutz, Chaos as an intermittently forced linear system, Nat. Commun. 8 (2017) 19, <https://doi.org/10.1038/s41467-017-00030-8>.
- [16] H. Arbabi, I. Mezić, Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator, SIAM J. Appl. Dyn. Syst. 16 (4) (2017) 2096–2126, <https://doi.org/10.1137/17M1125236>.
- [17] D. Giannakis, Data-driven spectral decomposition and forecasting of ergodic dynamical systems, Appl. Comput. Harmon. Anal. 47 (2019) 338–396, <https://doi.org/10.1016/j.acha.2017.09.001>.
- [18] S. Das, D. Giannakis, Delay-coordinate maps and the spectra of Koopman operators, J. Stat. Phys. 175 (2019) 1107–1145, <https://doi.org/10.1007/s10955-019-02272-w>.
- [19] D. Giannakis, S. Das, Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis, Physica D 402 (2019) 132211, <https://doi.org/10.1016/j.physd.2019.132211>.
- [20] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, J. Nonlinear Sci. 28 (2) (2018) 687–710, <https://doi.org/10.1007/s00332-017-9423-0>.
- [21] M. Korda, M. Putinar, I. Mezić, Data-driven spectral analysis of the Koopman operator, Appl. Comput. Harmon. Anal. 48 (2) (2020) 599–629, <https://doi.org/10.1016/j.acha.2018.08.002>.
- [22] S. Das, D. Giannakis, Koopman spectra in reproducing kernel Hilbert spaces, Appl. Comput. Harmon. Anal. 49 (2020) 573–607, <https://doi.org/10.1016/j.acha.2020.05.008>.
- [23] M. Budisić, R. Mohr, I. Mezić, Applied Koopmanism, Chaos 22 (2012) 047510, <https://doi.org/10.1063/1.4772195>.
- [24] A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, Chaos 27 (2017) 053104, <https://doi.org/10.1063/1.4982720>.
- [25] T. Eisner, B. Farkas, M. Haase, R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, Graduate Texts in Mathematics, vol. 272, Springer, 2015.
- [26] B.O. Koopman, Hamiltonian systems and transformation in Hilbert space, Proc. Natl. Acad. Sci. 17 (5) (1931) 315–318.
- [27] M.H. Stone, On one-parameter unitary groups in Hilbert space, Ann. Math. 33 (1932) 643–648, <https://doi.org/10.2307/1968538>.
- [28] K. Schmüdgen, *Unbounded and Self-Adjoint Operators on Hilbert Space*, Graduate Texts in Mathematics, vol. 265, Springer, Dordrecht, 2012.
- [29] R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30, <https://doi.org/10.1016/j.acha.2006.04.006>.
- [30] U. von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering, Ann. Stat. 26 (2) (2008) 555–586, <https://doi.org/10.1214/009053607000000640>.
- [31] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- [32] S. Klus, P. Koltai, C. Schütte, On the numerical approximation of the Perron–Frobenius and Koopman operator, J. Comput. Dyn. 3 (1) (2016) 51–79, <https://doi.org/10.3934/jcd.2016003>.
- [33] N. Aubry, R. Guyonnet, R. Lima, Spatiotemporal analysis of complex signals: theory and applications, J. Stat. Phys. 64 (1991) 683–739, <https://doi.org/10.1007/bf01048312>.

[34] D. Giannakis, A. Kolchinskaya, D. Krasnov, J. Schumacher, Koopman analysis of the long-term evolution in a turbulent convection cell, *J. Fluid Mech.* 847 (2018) 735–767, <https://doi.org/10.1017/jfm.2018.297>.

[35] J. Sławińska, D. Giannakis, Spatiotemporal pattern extraction with data-driven Koopman operators for convectively coupled equatorial waves, in: A. Banerjee, W. Ding, J. Dy, V. Lyubchich, A. Rhines (Eds.), *Proceedings of the 6th International Workshop on Climate Informatics*, Boulder, Colorado, 2016, pp. 49–52.

[36] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, *J. Stat. Phys.* 65 (3–4) (1991) 579–616, <https://doi.org/10.1007/bf01053745>.

[37] J.C. Ferreira, V.A. Menegatto, Positive definiteness, reproducing kernel Hilbert spaces and beyond, *Ann. Funct. Anal.* 4 (2013) 64–88, <https://doi.org/10.15352/afa/1399899838>.

[38] V.I. Paulsen, M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, Cambridge Studies in Advanced Mathematics, vol. 152, Cambridge University Press, Cambridge, 2016.

[39] 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, <https://doi.org/10.1016/j.acha.2005.07.005>.

[40] Y. Kawahara, Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis, in: *Advances in Neural Information Processing Systems*, 2016, pp. 911–919.

[41] S. Klus, I. Schuster, K. Muandet, Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces, *J. Nonlinear Sci.* 30 (1) (2020) 283–315, <https://doi.org/10.1007/s00332-019-09574-z>.

[42] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (2003) 1373–1396, <https://doi.org/10.1162/089976603321780317>.

[43] T. Berry, J. Harlim, Variable bandwidth diffusion kernels, *Appl. Comput. Harmon. Anal.* (2015), <https://doi.org/10.1016/j.acha.2015.01.001>.

[44] T. Berry, T. Sauer, Local kernels and the geometric structure of data, *Appl. Comput. Harmon. Anal.* 40 (2016) 439–469, <https://doi.org/10.1016/j.acha.2015.03.002>.

[45] I. Babuška, J. Osborn, *Eigenvalue Problems, Handbook of Numerical Analysis*, vol. 2, North Holland, Amsterdam, 1991.

[46] B. Franke, et al., The behavior of the spectral gap under growing drift, *Trans. Am. Math. Soc.* 362 (2010) 1325–1350, <https://doi.org/10.1090/S0002-9947-09-04939-3>.

[47] C. Liverani, Fredholm determinants, Anosov maps and Ruelle resonances, *Discrete Contin. Dyn. Syst.* 13 (5) (2005) 1203–1215, <https://doi.org/10.3934/dcds.2005.13.1203>.

[48] S. Dyatlov, M. Zworski, Stochastic stability of Pollicott–Ruelle resonances, *Nonlinearity* 28 (2015) 3511–3533, <https://doi.org/10.1088/0951-7715/28/10/3511>.

[49] M. Pollicott, Meromorphic extensions of generalized zeta functions, *Invent. Math.* 85 (1) (1986) 147–164, <https://doi.org/10.1007/BF01388795>.

[50] D. Ruelle, Resonances of chaotic dynamical systems, *Phys. Rev. Lett.* 56 (5) (1986) 405–407, <https://doi.org/10.1103/PhysRevLett.56.405>.

[51] C.R. de Oliveira, *Intermediate Spectral Theory and Quantum Dynamics*, Springer, 2008.

[52] R. Alexander, D. Giannakis, Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques, *Physica D* 409 (2020) 132520, <https://doi.org/10.1016/j.physd.2020.132520>.

[53] F. Cucker, S. Smale, On the mathematical foundations of learning, *Bull. Am. Math. Soc.* 39 (1) (2001) 1–49, <https://doi.org/10.1090/S0273-0979-01-00923-5>.

[54] K. Law, A.M. Stuart, K. Zygalakis, *Data Assimilation: A Mathematical Introduction*, Texts in Applied Mathematics, vol. 62, Springer, 2015.

[55] W. Tucker, The Lorenz attractor exists, *C. R. Acad. Sci. Paris, Ser. I* 328 (1999) 1197–1202, [https://doi.org/10.1016/S0764-4442\(99\)80439-X](https://doi.org/10.1016/S0764-4442(99)80439-X).

[56] P. Constantin, C. Foias, B. Nicolaenko, R. Temam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations*, Springer, New York, 1989.

[57] N. Aronszajn, Theory of reproducing kernels, *Trans. Am. Math. Soc.* 63 (1950) 337–404, <https://doi.org/10.1090/S0002-9947-1950-0051437-7>.

[58] R. Coifman, M. Hirn, Bi-stochastic kernels via asymmetric affinity functions, *Appl. Comput. Harmon. Anal.* 35 (1) (2013) 177–180, <https://doi.org/10.1016/j.acha.2013.01.001>.

[59] S. Bögli, Local convergence of spectra and pseudospectra, *J. Spectr. Theory* 8 (2018) 1051–1098, <https://doi.org/10.4171/JST/222>.

[60] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, *Philos. Trans. R. Soc. Lond. Ser. A* 20 (1909) 441–458, <https://doi.org/10.1098/rsta.1909.0016>.

[61] S. Beckus, J. Bellissard, Continuity of the spectrum of a field of self-adjoint operators, *Ann. Henri Poincaré* 17 (2016) 3425–3442, <https://doi.org/10.1007/s00023-016-0496-3>.

[62] N. Dunford, J.T. Schwartz, *Linear Operators*, vol. 2, Wiley, 1988.

[63] E. Kowalski, Spectral theory in Hilbert spaces, <https://people.math.ethz.ch/~kowalski/spectral-theory.pdf>, 2009.

[64] J. Stochel, F. Szafraniec, Domination of unbounded operators and commutativity, *J. Math. Soc. Jpn.* 55 (2) (2003) 405–437, <https://doi.org/10.2969/jmsj/1191419124>.

[65] F. Chaitin-Chatelin, A. Harrabi, About definitions of pseudospectra of closed operators in Banach spaces, *Tech. Rep. TR/PA/98/08*, CERFACS, Toulouse, France, 1998.

[66] F. Chatelin, *Spectral Approximation of Linear Operators*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 2011.

[67] L.N. Trefethen, M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, 2005.

[68] F. Schuh, Theorem on the term by term differentiability of a series, in: KNAW, *Proceedings*, vol. 22, 1919, pp. 376–378, <http://www.dwc.knaw.nl/DL/publications/PU00011973.pdf>.

- [69] R. Young, On complete biorthogonal systems, *Proc. Am. Math. Soc.* 83 (3) (1981) 537–540, <https://doi.org/10.1090/S0002-9939-1981-0627686-9>.
- [70] G. Tian, Y. Ji, Y. Cao, Schauder bases and operator theory, arXiv:1203.3603, 2012.
- [71] M. Reed, B. Simon, *Methods of Modern Mathematical Physics, II: Fourier Analysis, Self-adjointness*, Academic Press, San Diego, 1975.
- [72] G. Dell'Antonio, *Lectures on the Mathematics of Quantum Mechanics II: Selected Topics*, Atlantis Press, Amsterdam, 2016.
- [73] L.S. Young, What are SRB measures, and which dynamical systems have them?, *J. Stat. Phys.* 108 (2002) 733–754, <https://doi.org/10.1023/A:1019762724717>.
- [74] D. Giannakis, Dynamics-adapted cone kernels, *SIAM J. Appl. Dyn. Syst.* 14 (2) (2015) 556–608, <https://doi.org/10.1137/140954544>.
- [75] M.C. Genton, *Classes of kernels for machine learning: a statistics perspective*, *J. Mach. Learn. Res.* 2 (2001) 299–312.
- [76] O.E. Rössler, An equation for continuous chaos, *Phys. Lett. A* 57 (5) (1976) 397–398, [https://doi.org/10.1016/0375-9601\(76\)90101-8](https://doi.org/10.1016/0375-9601(76)90101-8).
- [77] M.J. McGuinness, The fractal dimension of the Lorenz attractor, *Philos. Trans. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.* 262 (1968) 413–458, <https://doi.org/10.1098/rsta.1968.0001>.
- [78] J.C. Sprott, *Chaos and Time-Series Analysis*, Oxford University Press, Oxford, 2003.
- [79] K. Law, A. Shukla, A.M. Stuart, Analysis of the 3DVAR filter for the partially observed Lorenz'63 model, *Discrete Contin. Dyn. Syst.* 34 (3) (2013) 1061–10178, <https://doi.org/10.3934/dcds.2014.34.1061>.
- [80] S. Luzzatto, I. Melbourne, F. Paccaut, The Lorenz attractor is mixing, *Comment. Phys.-Math.* 260 (2) (2005) 393–401, <https://doi.org/10.1007/s00220-005-1411-9>.
- [81] M. Peifer, B. Schelter, M. Winterhalder, J. Timmer, Mixing properties of the Rössler system and consequences for coherence and synchronization analysis, *Phys. Rev. E* 72 (2005) 026213, <https://doi.org/10.1103/PhysRevE.72.026213>.
- [82] T. Berry, J. Harlim, Iterated diffusion maps for feature identification, *Appl. Comput. Harmon. Anal.* 45 (1) (2018) 84–119, <https://doi.org/10.1016/j.acha.2016.08.005>.
- [83] H. Owhadi, G.R. Yoo, Kernel flows: from learning kernels from data into the abyss, *J. Comput. Phys.* 389 (2019) 22–47, <https://doi.org/10.1016/j.jcp.2019.03.040>.
- [84] I. Mezić, Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry, *J. Nonlinear Sci.* 30 (2020) 2091–2145, <https://doi.org/10.1007/s00332-019-09598-5>.
- [85] X. Wang, J. Sławińska, D. Giannakis, Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics, *Sci. Rep.* 10 (2020) 2636, <https://doi.org/10.1038/s41598-020-59128-7>.
- [86] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, in: *Advances in Neural Information Processing Systems*, vol. 17, 2004, pp. 1601–1608.
- [87] D. Giannakis, A. Ourmazd, J. Sławińska, Z. Zhao, Spatiotemporal pattern extraction by spectral analysis of vector-valued observables, *J. Nonlinear Sci.* 29 (5) (2019) 2385–2445, <https://doi.org/10.1007/s00332-019-09548-1>.