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 unbounded generator of the system by a compact operator $W_\tau$ on a reproducing kernel Hilbert space (RKHS). A key element of this approach is that $W_\tau$ is skew-adjoint (unlike regularization approaches based on the addition of diffusion), and thus can be characterized by a unique projection-valued measure, discrete by compactness, and an associated orthonormal basis of eigenfunctions. These eigenfunctions can be ordered in terms of a measure of roughness (Dirichlet energy) on the RKHS, 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, the regularized generator has a well-defined Borel functional calculus allowing the construction of an associated unitary evolution group $\{e^{itW_\tau}\}_{t \in \mathbb{R}}$ on the RKHS by exponentiation, approximating the unitary Koopman evolution group of the original system. We establish convergence results for the spectrum and Borel functional calculus of the regularized generator to those of the original system in the limit $\tau \to 0^+$. Convergence results are also established for a data-driven formulation, where these operators are approximated using finite rank operators obtained from observed time series. An advantage of working in spaces of observables with an RKHS structure is that one can perform pointwise evaluation and interpolation through bounded linear operators, which is not possible in $L^p$ spaces. This enables the evaluation of data-approximated eigenfunctions on previously unseen states, as well as data-driven forecasts initialized with pointwise initial data (as opposed to probability densities in $L^p$). The pattern extraction and prediction framework developed here is numerically applied to a number of ergodic dynamical systems with atomic and continuous spectra, with promising results.

Keywords: Koopman operators, Perron-Frobenius operators, reproducing kernel Hilbert spaces, spectral estimation

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 \cite{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22} is the operator-theoretic formulation of ergodic theory \cite{23, 24, 25}, where instead of directly studying the properties of the dynamical flow on the 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 when defined on appropriate spaces of functions and measures, respectively.

*Corresponding author
Email address: dimitris@cims.nyu.edu (Dimitrios Giannakis)
It is a remarkable fact, realized in the work of Koopman in the 1930s [26], that the action of a general non-linear 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 underlying dynamical flow, 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 stable and convergent approximation techniques.

The Koopman operator $U^t$ associated with a continuous-time, continuous flow $\Phi^t : M \mapsto M$ on a manifold $M$ acts on functions by composition, $U^t f = f \circ \Phi^t$. It is a norm-preserving 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 Borel invariant 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. 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 seek to address the following two broad classes of problems:

1. **Coherent pattern extraction**: that is, identification of a collection of observables in $L^2(\mu)$ having high regularity and a natural temporal evolution under $U^t$.
2. **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 $F : M \mapsto Y$ taking values in a data space $Y$, sampled finitely many times along an orbit of the dynamics.

**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 $V : D(V) \mapsto L^2(\mu)$ with $D(V) \subseteq L^2(\mu)$ and

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

In particular, associated with $V$ is a unique projection-valued measure (PVM) $E : \mathcal{B}(\mathbb{R}) \mapsto \mathcal{L}(L^2(\mu))$ 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

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

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

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

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 $H$ admits the $U^t$-invariant orthogonal splitting

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

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. In particular, on these subspaces there exist unique PVMs $E_p : \mathcal{B}(\mathbb{R}) \mapsto \mathcal{L}(H_p)$ and $E_c : \mathcal{B}(\mathbb{R}) \mapsto \mathcal{L}(H_c)$, respectively, where $E_p$ is atomic and $E_c$ is continuous, yielding the spectral decomposition

$$E = E_p \oplus E_c.$$
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 flow $\Phi^t$. As a result, $H_p$ admits an orthonormal basis $\{z_j\}$ satisfying

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

(4)

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 evolve under the dynamics by multiplication by a periodic phase factor at a distinct frequency $\omega_j$, even if the underlying dynamical flow is nonlinear and aperiodic. In contrast, observables $f \in H_d$ do not exhibit an analogous quasiperiodic evolution, and are characterized instead by a weak-mixing property (decay of correlations),

$$\frac{1}{T} \int_0^T \| (g, U^s f) \|_\mu \, ds \xrightarrow{t \to \infty} 0, \quad \forall g \in L^2(\mu).$$

This is characteristic of chaotic dynamics.

**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 former it is sufficient to perform pointwise (or even weak) approximations of the operators, whereas the latter 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 $\{\phi_0, \phi_1, \ldots\}$ of $L^2(\mu)$ with associated orthogonal projection operators $\Pi_i : L^2(\mu) \mapsto \text{span}\{\phi_0, \ldots, \phi_i\}$, the finite-rank operator $U^t_i = \Pi_i U^t \Pi_i$ is fully characterized by the matrix elements $U^t_{ij} = \langle \phi_i, U^t \phi_j \rangle_\mu$ with $0 \leq i, j \leq l - 1$, and by continuity of $U^t$, the sequence of operators $U^t_i$ converges pointwise to $U^t$. Thus, if one has access to data-driven approximations $U^t_{ij,N}$ of $U^t_i$ determined from $N$ measurements of $F$ taken along an orbit of the dynamics, and these approximations converge as $N \to \infty$, then, as $l \to \infty$ and $N \gg l$, the corresponding finite-rank operators $U^t_{ij,N}$ converge pointwise to $U^t$. This property was employed in [12] in a technique called diffusion forecasting, whereby the approximate matrix elements $U^t_{ij,N}$ are computed in a data-driven basis constructed from samples of $F$ using the diffusion maps algorithm (a kernel algorithm for manifold learning) [23]. By spectral convergence results for kernel integral operators established in [30] and ergodicity, as $N \to \infty$, 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 $U^t_{ij,N}$ converge pointwise to $U^t$ as described above. It was demonstrated that diffusion forecasts of observables of the L63 system have skill approaching that of ensemble forecasts with 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 [31, 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 in cases where the support of $\mu$ is an unknown, measure-zero subset of the state space manifold $M$).

Of course, this is not to say that the spectral decomposition in [3] 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, which 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, 2, 3, 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 [4] 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 [5], 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.

Yet, despite these and other attractive theoretical properties of evolution operators, the design of data-driven approximation methods with rigorous convergence guarantees that can naturally handle both the point and continuous spectra of the operators 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, as, 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 necessitates 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 generally 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), with difficult to control asymptotic behavior. 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 data 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 [72], who establish weak spectral convergence results implied by strongly convergent approximations of the Koopman operator derived through EDMD. This approach, however, 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 [5, 3, 21, 22]. Among these, [3, 4] consider a spectral decomposition of the Koopman operator closely related to (3), 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 eigenfunctions. 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] 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 [38]. While this method also suffers from a singular behavior in $\omega$, it was found to significantly outperform conventional harmonic averaging techniques, 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. 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 appears difficult 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, 19] 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 [28, 30, 39, 40, 41] as in diffusion forecasting. A key ingredient of these techniques is a family $K_1, K_2, \ldots$ 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 $K_\infty : L^2(\mu) \mapsto L^2(\mu)$ 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, 19], a well-posed Galerkin method 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 $\Delta : D(\Delta) \mapsto L^2(\mu)$ on a suitable domain $D(\Delta) \subset D(V)$. This leads to an advection-diffusion operator

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

whose eigenvalues and eigenfunctions can be computed through provably convergent Galerkin schemes based on classical approximation theory for variational eigenvalue problems [42]. The diffusion operator in (5) is constructed so as to compute 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 data under realistic assumptions on the dynamical system and observation map.

Advection-diffusion operators as in (5) 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 obtained in [13], 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 $\Delta$ commutes with $V$, meaning that $L$ is necessarily non-normal. As is well known, 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 understand 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 restriction precludes the method from identifying coherent observables in $H_c$. Put together, these facts call for a different regularization approach to (5) that can seamlessly handle both the point and continuous spectra of $V$.

Contributions of this work. In this paper, we propose a data-driven framework for pattern extraction and prediction in 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 (5) 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 $W_\tau : \mathcal{H}_\tau \mapsto \mathcal{H}_\tau$, $\tau > 0$, 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 Hilbert-Schmidt integral operators with continuous kernels. Moreover, the spaces $\mathcal{H}_\tau$ employed in this framework are universal, i.e., lie dense in the space of continuous functions on $M$ [14], 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 $\tau \to 0$ 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 spectral measures $\tilde{E}_\tau : \mathcal{B}(\mathbb{R}) \mapsto \mathcal{L}(\mathcal{H}_\tau)$. The existence of $\tilde{E}_\tau$ 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 $\tilde{E}_\tau$ are purely atomic, have bounded support, and thus characterized by a countable set of bounded, real-valued eigenfrequencies with a corresponding orthonormal eigenbasis of $\mathcal{H}_\tau$.

(ii) For systems that do possess nontrivial Koopman eigenfunctions, there exists a subset of the eigenfunctions of $W_\tau$ converging to them as $\tau \to 0$. 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 coherently 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 predicting 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 $L^2(\mu)$).

(iv) Our framework is well-suited for data-driven approximation using techniques from statistics and machine learning [30, 38, 45]. 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 Corollary 3 we prove the spectral convergence of $W_\tau$ to $V$ in an appropriate sense by defining auxiliary compact operators acting on $L^2(\mu)$. In Theorem 16 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.

Outline of the paper. In Section 2 we make our assumptions on the underlying system precise, and also state our main results, Theorems 1, 2, and Corollary 3 there. This is followed by results on compactification of operators in RKHS, Theorems 4–9 in Section 3 which will be useful for the proofs of the main results. Before proving the 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 4, 7 and 8, 9 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 16). 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 Rossler and Lorenz 63 (L63) systems. The paper ends in Section 9 with concluding remarks.

2. Main results

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

Assumption 1. $\Phi : \mathcal{M} \mapsto M$, $t \in \mathbb{R}$ is a continuous-time, continuous flow on a metric space $M$. There exists a forward-invariant, $n$-dimensional $C^r$ 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 a Borel, ergodic, invariant 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 Lorenz 63 (L63) system [46], where \( \mathcal{M} = \mathbb{R}^3 \), \( M \) is an appropriate absorbing ball [47], and \( X \) a fractal attractor [48]), and certain dissipative partial equations with inertial manifolds [49] (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 reproducing kernel Hilbert spaces (RKHSs).

**Kernels and their associated integral operators.** In the context of interest here, a kernel will be a continuous function \( k : M \times M \to \mathbb{C} \), 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 \( K : L^2(\nu) \to C^0(\mathcal{M}) \), acting on \( f \in L^2(\nu) \) as

\[
Kf := \int_M k(\cdot,y)f(y)\,d\nu(y). 
\]

If, in addition, \( k \) lies in \( C^r(M \times M) \), then \( K \) inherits this smoothness to \( Kf \), i.e., \( Kf \in C^r(\mathcal{M}) \). 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 will be called positive-definite if for every \( x_1, \ldots, x_n \in M \) and \( a_1, \ldots, a_n \in \mathbb{C} \), the sum \( \sum_{i,j=1}^n a_i^* a_j k(x_i, x_j) \) is non-negative, and strictly positive-definite if the sum is zero iff each of the \( a_1, \ldots, a_n \) equals zero. 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 \( C^r(\mathcal{M}) \), a kernel \( k \) also induces an operator \( G = \iota K \) on \( L^2(\mu) \). This kernel will be called a Markov kernel with respect to \( \nu \) if the associated integral operator \( G : L^2(\nu) \to L^2(\nu) \) is Markov, i.e., (i) \( Gf \geq f \) if \( f \geq 0 \); (ii) \( Gf = f \) if \( f \) is constant; and (iii) \( \|Gf\|_{L^2(\nu)} \leq \|f\|_{L^2(\nu)} \), for all \( f \in L^2(\nu) \).

**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 \( \delta_x : \mathcal{H} \to \mathbb{C} \), \( \delta_x f = f(x) \), is bounded, and thus continuous, linear functional. By the Riesz representation theorem, every RKHS has a unique reproducing kernel, i.e., a kernel \( k : M \times M \to \mathbb{C} \) such that for every \( x \in M \) the kernel section \( k(x,\cdot) \) lies in \( \mathcal{H} \), and for every \( f \in \mathcal{H} \),

\[
f(x) = \delta_x f = (k(x,\cdot),f)_{\mathcal{H}},
\]

where \( (\cdot,\cdot)_{\mathcal{H}} \) is the inner product of \( \mathcal{H} \), assumed conjugate-linear in the first argument. It then follows that \( k \) is Hermitian. Moreover, according to the Moore-Aronsajn theorem [50], given a symmetric, positive-definite kernel \( k : M \times M \to \mathbb{C} \), there exists a unique RKHS \( \mathcal{H} \) for which \( k \) is the reproducing kernel. If \( k \) is continuous and strictly positive-definite, \( \mathcal{H} \) is a dense subset of \( C^0(\mathcal{M}) \) [44]. In fact, for every \( r \geq 0 \), if \( k \in C^r(M \times M) \), then \( \mathcal{H} \) is a dense subset of \( C^r(\mathcal{M}) \) [37]. Moreover, the range of \( K \) from \( \mathcal{H} \) to \( \mathcal{H} \), so we can view \( K \) as an operator \( K : L^2(\nu) \to \mathcal{H} \) between Hilbert spaces. With this definition, \( K \) is compact, and the adjoint operator \( K^* : \mathcal{H} \to L^2(\nu) \) 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 of \( \nu \), the range of \( K \) is a dense subspace of \( \mathcal{H}(\text{supp}(\nu)) \). In particular, if \( k \) is strictly positive-definite, \( \mathcal{H}(\text{supp}(\nu)) \), and thus ran \( K|_{\text{supp}(\nu)} \), are dense subspaces of \( C^0(\text{supp}(\nu)) \).

**Nyström extension.** Let \( \mathcal{H} \) be an RKHS on \( M \) with reproducing kernel \( k \). Then, the Nyström extension operator \( \mathcal{N} : D(\mathcal{N}) \to \mathcal{H} \) acts on a subspace \( D(\mathcal{N}) \) of \( L^2(\nu) \), mapping each element \( f \) in its domain to a...
function $N f \in \mathcal{H}$, such that $N f$ lies in the same $L^2(\nu)$ equivalence class as $f$. In other words, $N f(x) = f(x)$ for $\nu$-a.e. $x \in M$, and $K^* N$ is the identity on $D(N)$. It can also be shown that $N K^*$ is the identity on $\mathcal{H}(\text{supp}(\nu))$. If the kernel $k$ is strictly positive-definite, then $D(N)$ is a dense subspace of $L^2(\nu)$. We will give a precise constructive definition of $N$ in Section 4.

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

**Assumption 2.** $p : M \times M \to \mathbb{R}$ is a $C^\tau$, symmetric, strictly positive-definite Markov kernel with respect to the invariant measure $\mu$, with $p > 0$.

We will later describe how kernels satisfying Assumption 2 can easily be constructed from real, symmetric, strictly positive-definite, $C^\tau$ kernels using the bistochastic kernel normalization technique proposed in [51]. It should be noted that many of our results will require $\tau = 2$ differentiability class in Assumptions 1 and 2, but in some cases that requirement will be relaxed to $\tau = 1$ or 0.

**One-parameter kernel families.** Let $P : L^2(\mu) \mapsto \mathcal{H}$ 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 $L^2(\mu)$ has positive eigenvalues, which can be ordered as $1 = \lambda_0 > \lambda_1 \geq \ldots$. Given a real, orthonormal basis $\{\phi_0, \phi_1, \ldots\}$ of $L^2(\mu)$ consisting of corresponding eigenfunctions, it is known from RKHS theory [37] that $\{\psi_0, \psi_1, \ldots\}$ with $\psi_j = \lambda_j^{-1/2} P\phi_j$ is an orthonormal basis of $\mathcal{H}(X)$. Defining

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

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 $L^2(\mu)$.

**Theorem 1 (Markov kernels).** Let Assumptions 1 and 2 hold. Then, the series expansion for $p_\tau(x,y)$ in (7) converges to the values of $C^\tau$, symmetric, strictly positive-definite Markov kernels on $M$, and the convergence is uniform on $X \times X$. Moreover, the following hold:

(i) The RKHS $\mathcal{H}_\tau$ on $M$ corresponding to $p_\tau$ lies dense in $C^0(M)$, and for every $0 < \tau_1 < \tau_2$, the inclusions $\mathcal{H}_{\tau_2}(X) \subseteq \mathcal{H}_{\tau_1}(X) \subseteq \mathcal{H}(X)$ hold. Moreover, $\{\psi_{\tau,0,\tau,1,\ldots}\}$ is an orthonormal basis of $\mathcal{H}_{\tau}(X)$.

(ii) For every $\tau > 0$, the operator $G_\tau = P_\tau^* P_\tau$, where $P_\tau : L^2(\mu) \mapsto \mathcal{H}_{\tau}$ is the integral operator associated with $p_\tau$, is a positive-definite, self-adjoint, compact Markov operator on $L^2(\mu)$.

(iii) Define $G_0 := I_{L^2(\mu)}$. Then, the family $\{G_{\tau}\}_{\tau \geq 0}$ forms a strongly continuous Markov semigroup; i.e., $G_{\tau_1 + \tau_2} = G_{\tau_1} G_{\tau_2}$ for every $\tau_1, \tau_2 \geq 0$, and $G_\tau$ converges pointwise to the identity operator as $\tau \to 0^+$.

We will now use Markov operators $G_\tau$ from Theorem 1 to compactify the generator $V$, and then establish various ways that these compactifications converge to $V$. In what follows, $\mathcal{H}_\tau := D(N_\tau) \mapsto \mathcal{H}_\tau$ will be the Nyström operator associated with $\mathcal{H}_\tau$. We also let $\mathcal{H}_\infty = \bigcap_{\tau \geq 0} D(N_\tau)$ be the dense subspace of $L^2(\mu)$ whose elements have $\mathcal{H}_\tau$ representatives for every $\tau > 0$. Note that $\mathcal{H}_\infty$ is dense since it contains all finite linear combinations of the $\phi_j$. Similarly, setting $\mathcal{H}_\infty = \bigcap_{\tau \geq 0} \mathcal{H}_\tau$, it follows that $\mathcal{H}_\infty(X)$ is a dense subspace of $\mathcal{H}(X)$. Given a Borel-measurable function $Z : \mathbb{R} \mapsto \mathbb{C}$ and a densely-defined skew-adjoint operator $A$, $Z(A)$ will denote the operator-valued function obtained through the Borel functional calculus as in Section 1.

For every set $\Omega \subseteq C$, $\partial \Omega$ denotes its boundary.

**Theorem 2 (Main theorem).** Under Assumptions 1, 2 with $\tau = 2$, and the definitions in (7), the following hold for every $\tau > 0$:

(i) The operator $W_\tau P_\tau V P_\tau^* : \mathcal{H} \to \mathcal{H}$ is a well-defined, skew-adjoint, Hilbert-Schmidt operator.

(ii) The operator $G_\tau V : D(V) \mapsto L^2(\mu)$ extends to a Hilbert-Schmidt integral operator $B_\tau : L^2(\mu) \mapsto L^2(\mu)$. Moreover, the restrictions of $G_\tau V$ and $B_\tau$ on the dense subspace $D(N_\tau) \subseteq D(V)$ coincide with the operator $P_\tau^* W_\tau N_\tau$.  

8
(iii) The operators \( B_\tau \) and \( W_\tau \) have purely atomic spectral measures \( E_\tau : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(L^2(\mu)) \) and \( E_\tau : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(H_\tau) \), respectively, with the same eigenvalues.

(iv) For every bounded Borel measurable set \( \Omega \subset \mathbb{R} \) such that \( E(\partial \Omega) = 0 \), \( E_\tau(\Omega) \) and \( P_\tau^* E_\tau(\Omega) N_\tau \) converge strongly to \( E(\Omega) \), respectively on \( L^2(\mu) \) and \( H_\tau \), as \( \tau \to 0^+ \). Moreover, for every \( \tau > 0 \), there exists a unitary operator \( U_\tau : L^2(\mu) \to H_\tau \) such that \( U_\tau^* E_\tau(\Omega) U_\tau \) converges to \( E(\Omega) \), strongly on \( L^2(\mu) \).

(v) For every bounded continuous function \( Z : \mathbb{R} \to \mathcal{C} \), as \( \tau \to 0^+ \), \( P_\tau^* Z(W_\tau) N_\tau \) and \( U_\tau^* Z(W_\tau) U_\tau \) converge to \( Z(V) \), in the strong operator topologies of \( H_\infty \) and \( L^2(\mu) \), respectively.

(vi) For every element \( i\omega, \omega \in \mathbb{R} \), of the spectrum of the generator \( V \), there exists a continuous curve \( \tau \mapsto \omega_\tau \) such that \( i\omega_\tau \) is an eigenvalue of \( B_\tau \) and \( W_\tau \), and \( \lim_{\tau \to 0^+} \omega_\tau = \omega \).

**Remark.** In general, the operators \( B_\tau \) are non-normal, and have non-orthogonal eigenspaces. As a result, unlike PVMs associated with skew-adjoint operators (including the measures \( E_\tau \) associated with \( W_\tau \)), the corresponding spectral measures \( E_\tau \) take values of non-orthogonal projection operators on \( L^2(\mu) \). A complete definition of \( E_\tau \) will be given in Section 3.

The skew-adjoint operator \( W_\tau \) from Theorem 2 can be viewed as a compact approximation to the generator \( V \), which is unbounded and exhibits complex spectral behavior (see Section 1). This approximation has a number of advantages for both coherent extraction and precision. In particular, regardless of the spectrum of \( V \), \( W_\tau \) has a complete orthonormal basis of eigenfunctions, which are \( C^2 \) functions lying in \( H_\tau \). These eigenfunctions recover Koopman eigenfunctions as special cases, in the sense that for every Koopman eigenfunction, there exists a sequence of eigenfunctions of \( W_\tau \) that converges to it as \( \tau \to 0 \) (in \( L^2(\mu) \) norm).

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 prediction 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.

With regards to prediction, let \( \{i\omega_{\tau,0},i\omega_{\tau,1},\ldots\} \) be the set of eigenvalues of \( W_\tau \) (and \( V_\tau \)), ordered so that \( |\omega_{\tau,0}| \geq |\omega_{\tau,1}| \geq \cdots \geq 0 \). Note that the \( \omega_{\tau,j} \) come in complex-conjugate pairs, and 0 is the only accumulation point of the sequence \( \omega_{\tau,0}, \omega_{\tau,1}, \ldots \) by skew-adjointness and compactness of \( W_\tau \), respectively. Let also \( \{\zeta_{\tau,0},\zeta_{\tau,1},\ldots\} \) be an orthonormal basis of \( H_\tau \) 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 \( H_\infty \) under \( e^{tW_\tau} \).

**Corollary 3 (Prediction).** For every \( \tau > 0 \), \( W_\tau \) generates a norm-continuous group of unitary operators \( e^{tW_\tau} : H_\tau \to H_\tau, \ t \in \mathbb{R} \). Moreover, for any observable \( f \in L^2(\mu) \) and error bound \( \epsilon > 0 \), there exists \( f_\epsilon \in H_\infty \) such that for all \( t \in \mathbb{R} \), \( \|U^t f - U^t f_\epsilon\|_\mu < \epsilon \), and

\[
\lim_{\tau \to 0^+} \|U^t f_\epsilon - e^{tW_\tau} f_\epsilon\|_\mu = 0, \quad e^{tW_\tau} f_\epsilon = \sum_{j=0}^{\infty} e^{i\omega_{\tau,j}\cdot t} \langle \zeta_{\tau,j}, f_\epsilon \rangle H_\tau, \zeta_{\tau,j}.
\]

**Remark.** The function \( e^{tW_\tau} f_\epsilon \) lies in \( H_\tau \), and is therefore a continuous function which we employ as a predictor for the evolution of the observable \( f \) under \( U^t \). Corollary 3 suggests that to obtain this predictor, we first regularize \( f \) by approximating it by a function \( f_\epsilon \in H_\infty \), 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 3 can be obtained for other operator-valued functions of the generator than the exponential functions, \( U^t = e^{tV} \). A constructive procedure for obtaining the predictor in a data-driven setting will be described in Section 8.

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. The are of independent interest, as, with appropriate modifications,
they apply for more general classes of unbounded, skew-adjoint operators obtained by extension of differentiation operators. In some cases, the following weaker analog of Assumption 3 will be sufficient.

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

Given the RKHS $\mathcal{H} \subset C^1(M)$ associated with $k$, and the corresponding integral operators $K : L^2(\mu) \to \mathcal{H}$ and $G = K^*K : L^2(\mu) \to L^2(\mu)$, we begin by formally introducing the operators $A : L^2(\mu) \to L^2(\mu)$ and $W : \mathcal{H} \to \mathcal{H}$, where

$$A = VG, \quad W = KV^*.$$  

(8)

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, under sufficient regularity conditions on the kernel, $A$ and $W$ are well-defined, and in fact compact, operators.

**Theorem 4 (Pre-smoothing).** Let Assumptions 1 and 3 hold, and define $k' : M \times M \to \mathbb{R}$ as the $C^0$ kernel with $k'(x, y) := V k(\cdot, y)(x)$. Then:

(i) The range of $k'$ lies in the domain of $V$.

(ii) The operator $A$ from (8) is a well-defined, Hilbert-Schmidt integral operator on $L^2(\mu)$ with kernel $k'$, and thus operator norm bounded 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 $GV : D(V) \to L^2(\mu)$.

**Remark.** As stated in Section 1, $V$ is an unbounded operator, whose domain is a strict subspace of $L^2(\mu)$. Theorem 4 thus shows that if we regularize this operator by first applying the smoothing operator $G$, then not only is $V$ 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 5 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 5 (Compactification in RKHS).** Let Assumptions 1 and 3 hold, and define $\tilde{k}' : M \times M \to \mathbb{R}$ as the $C^0$ kernel with $\tilde{k}'(x, y) = -k(\cdot, y)(x)$. Then:

(i) The range of $K^*$ lies in the domain of $V$, and $VK^* : \mathcal{H} \to L^2(\mu)$ is a bounded operator.

(ii) The operator $W$ from (8) is a well-defined, Hilbert-Schmidt, skew-adjoint, real operator on $\mathcal{H}$, 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 $e^{tW} : \mathcal{H} \to \mathcal{H}$, $t \in \mathbb{R}$.

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

**Theorem 6 (Post-smoothing).** Let Assumptions 1 and 3 hold. Then, the adjoint of $-A$ from (8) is a Hilbert-Schmidt integral operator $B : L^2(\mu) \to L^2(\mu)$ with kernel $\tilde{k}'$. In addition:
(i) The densely-defined operator \( GV : D(V) \to L^2(\mu) \) is bounded, and \( B \) is equal to its closure, \( \overline{GV} := (GV)^{**} \). Moreover, \( B \) is a closed extension of \( KW\mathcal{N} : D(\mathcal{N}) \to L^2(\mu) \), and if the kernel \( k \) is strictly positive-definite, 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 bounded operators \( e^{tB} : L^2(\mu) \to L^2(\mu), t \in \mathbb{R} \), 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 6(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 4–6, we have obtained the following sequence of operator extensions:
\[
KW\mathcal{N} \subseteq GV \subset B = \overline{GV} = (GV)^{**}.
\]

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-semidefinite, it has a square root \( G^{1/2} : L^2(\mu) \to L^2(\mu) \), which is the unique positive-semidefinite 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 \( G \). Using this operator, we will show in Theorem 7 below that the operator \( G^{1/2} V G^{1/2} \), 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 7 (Skew-adjoint compactification).** Let Assumptions 4 and 5 hold with \( r = 1 \). Then, \( G^{1/2} V G^{1/2} \) is a densely defined, bounded operator with a unique skew-adjoint extension to a Hilbert-Schmidt, skew-adjoint operator \( \tilde{V} : L^2(\mu) \to L^2(\mu) \). Moreover, \( \tilde{V} \) is unitarily equivalent to the operator \( W \) from Theorem 6, that is, there exists a unitary operator \( U : L^2(\mu) \to \mathcal{H} \) such that \( \tilde{V} = U G V U^* \).

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 8 and 9 below. Before stating these theorems, we introduce the spectral measures appropriate for the spectral decomposition of the non-normal, compact operators \( A \) and \( B \).

**Non-orthogonal projection-valued measures.** In what follows, a non-orthogonal projection valued measure (NPVM) on a Hilbert space \( \mathcal{H} \), will be a mapping \( E \) from the Borel sets of \( \mathbb{C} \) into the space of linear, bounded maps \( \mathcal{L}(\mathcal{H}) \) with the following properties.

(i) \( E(\mathbb{C}) = I_{\mathcal{H}} \) and \( E(\emptyset) = 0 \).

(ii) For each \( S \in B(\mathbb{C}), E(S) \) is a projection map, i.e., \( E(S)E(S') = E(S') \).

(iii) For any disjoint collection \( S_1, S_2, \ldots \in B(\mathbb{C}), E(\bigcup_{j \in \mathbb{N}} S_j) = E(S_1) + E(S_2) + \cdots \), where this countable sum of operators is meant to converge in the strong operator topology.

(iv) For any \( S_1, S_2, \ldots \in B(\mathbb{C}), E(\cap_{j \in \mathbb{N}} S_j) = E(S_1)E(S_2)\cdots \).

An NPVM \( E \) will be called a spectral measure for a (possibly unbounded) operator \( T : D(T) \to \mathcal{H}, D(T) \subseteq \mathcal{H} \)
\[
\langle f, \mathcal{T}g \rangle_{\mathcal{H}} = \int_{\mathbb{C}} zdE_{fg}, \forall u \in \mathcal{H}, \quad .
\]

where \( E_{fg} : B(\mathbb{C}) \to \mathbb{C} \) is the complex-valued Borel measure given by \( E_{fg}(S) = \langle f, g \rangle_{L^2(\mu)} \). \( E(S) \) is not necessarily an orthogonal map. Spectral measures are known to exist uniquely for normal operators.

We construct spectral measures for families of skew-adjoint as well as non-normal compact operators, in Theorem 8. It is well known [52] that if \( T \) is a normal operator, then the support of \( E \) equals \( \sigma(T) \). The spectral measure provides a decomposition of the operator \( T \), and also provides a way to define its functional calculus, as shown in [4] (2).
Theorem 8 (Spectra of the compactified generators). Let Assumptions 1 and 3 hold with \( r = 1 \), and assume further that the kernel \( k \) is strictly positive definite. Let also \( \{ \tilde{z}_0, \tilde{z}_1, \ldots \} \) 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) \( A, B, \) and \( W \) have the same spectra as \( \tilde{V} \), including multiplicities of eigenvalues.

In addition, if the kernel \( k \) is Markov:

(ii) 0 is a simple eigenvalue of each of the operators \( A, B, \tilde{V}, \) and \( W \), and the corresponding eigenfunctions are the constant functions.

(iii) Every \( \tilde{z}_j \) lies in the domain of \( G^{-1/2} \). The set \( \{ z'_j = G^{-1/2}\tilde{z}_j : j \in \mathbb{N}_0 \} \) are eigenfunctions of \( A \) corresponding to the eigenvalues \( \{ i\omega_j : j \in \mathbb{N}_0 \} \), and forms an unconditional Schauder basis of \( L^2(\mu) \).

(iv) The set \( \{ z_0, z_1, \ldots \} \) with \( z_j = G^{1/2}\tilde{z}_j \) is a (not necessarily orthogonal), unconditional Schauder basis of \( L^2(\mu) \), consisting of eigenfunctions of \( B \) corresponding to the same eigenvalues, \( \{ i\omega_0, i\omega_1, \ldots \} \).

Moreover, it is the unique dual sequence to the \( \{ z_j \} \), satisfying \( \langle z'_j, z_l \rangle_\mu = \delta_{jl} \).

(v) The set \( \{ \zeta_0, \zeta_1, \ldots \} \) with \( \zeta_j = Kz'_j \) is an orthonormal basis of \( \mathcal{H} \) consisting of eigenfunctions \( \zeta_j \) of \( W \) corresponding to the eigenvalues \( i\omega_j \).

(vi) The operators \( A, B, \tilde{V}, \) and \( W \) are respectively characterized by the projection-valued measures \( E', E, \tilde{E} : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(L^2(\mu)) \) and \( \mathcal{E} : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(\mathcal{H}) \), such that

\[
E'(U) = \sum_{j: \omega_j \in U} \langle z'_j, \cdot \rangle_{L^2(\mu)} z'_j, \quad E(U) = \sum_{j: \omega_j \in U} \langle z_j, \cdot \rangle_{L^2(\mu)} z_j, \quad \tilde{E}(U) = \sum_{j: \omega_j \in U} \langle \tilde{z}_j, \cdot \rangle_{L^2(\mu)} \tilde{z}_j,
\]

\[
\mathcal{E}(U) = \sum_{j: \omega_j \in U} \langle \zeta_j, \cdot \rangle_{\mathcal{H}} \zeta_j,
\]

and

\[
A = \int_{\mathbb{R}} i\omega \, dE'(\omega), \quad B = \int_{\mathbb{R}} i\omega \, dE(\omega), \quad \tilde{V} = \int_{\mathbb{R}} i\omega \, d\tilde{E}(\omega), \quad W = \int_{\mathbb{R}} i\omega \, d\mathcal{E}(\omega).
\]

Moreover, \( \tilde{E} \) and \( \mathcal{E} \) are unitarily equivalent under conjugation under the unitary operator \( U : L^2(\mu) \to \mathcal{H} \) from Theorem 7, i.e., \( E(U) = U^* \mathcal{E}(U) \, U \).

Remark. (i) The compactness of \( \tilde{V} \), and \( W \) allows for simple expressions for the functional calculi of these operators. For instance, for every Borel-measurable function \( Z : i\mathbb{R} \to \mathbb{C} \), we have

\[
Z(W) = \int_{\mathbb{R}} Z(i\omega) \, d\mathcal{E}(\omega) = \sum_{j=0}^{\infty} Z(i\omega_j) \langle \zeta_j, \cdot \rangle_{\mathcal{H}} \zeta_j,
\]

and analogous relationships hold for \( Z(A), Z(B), \) and \( Z(\tilde{V}) \).

(ii) Because the operators \( A \) and \( B \) are generally non-normal and have non-orthogonal eigenspaces, the corresponding projection-valued measures, respectively \( E \) and \( E' \), may take values in the set of non-orthogonal projections on \( L^2(\mu) \). While a non-orthogonal projection may be unbounded operator, in this case, \( E \) and \( E' \) always result in bounded projection operators.

(iii) The Markovianity assumption on the kernel was important to conclude that \( A, B, \tilde{V}, \) and \( W \) 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.

The results in Theorems 4–8 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 \( p_r \) associated with the Markov semigroups in our main result, Theorem 2. Specifically, we assume:
Assumption 4. \{k_\tau : M \times M \to \mathbb{R}\} with \(\tau > 0\) is a one-parameter family of \(C^1\), symmetric, strictly positive-definite kernels, such that, as \(\tau \to 0^+\), 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 skew-adjoint compactified generators \(\tilde{V}_\tau \supseteq C_{1/2} V G_{1/2}^*\) converges strongly to \(V\) on the subspace \(D(V^2) \subset D(V)\).

Under this assumption, we establish the following notion of spectral convergence for approximations of the generator \(V\) by compact operators.

Theorem 9 (Spectral convergence). Suppose that Assumptions 2 and 4 hold with \(r = 1\), and let \(W_\tau : \mathcal{H}_\tau \to \mathcal{H}_\tau\) and \(B_\tau : L^2(\mu) \to L^2(\mu)\), with \(\tau > 0\), be the Hilbert-Schmidt operators constructed via 5 and Theorem 4 respectively, applied for the kernels \(k_\tau\) from Assumption 4. Let also \(E_\tau\) and \(E_\tau^*\) be the (purely atomic) spectral measures of \(B_\tau\) and \(W_\tau\), respectively, constructed as in Theorem 4. Then:

(i) As \(\tau \to 0^+\), the operators \(A_\tau\) and \(B_\tau\) converges strongly to \(V\) on \(D(V)\).

(ii) For every bounded continuous function \(Z : i\mathbb{R} \to \mathbb{C}\) and as \(\tau \to 0^+\), \(Z(A_\tau)\), \(Z(B_\tau)\) and \(K^*_\tau Z(W_\tau)N_\tau\) converge to \(Z(V)\), in the strong operator topology of \(D(Z(V))\).

(iii) For every bounded Borel measurable set \(U \subseteq \mathbb{R}\) such that \(E(\partial U) = 0\), and as \(\tau \to 0^+\), \(E_\tau(U)\), \(E_\tau^*(U)\) and \(K^*_\tau E_\tau(U)N_\tau\) converge strongly to \(E(U)\).

(iv) For every element of the spectrum \(\omega\) of the generator \(V\), there exists a sequence of eigenvalues \(\omega_\tau\) of \(B_\tau\) (and \(A_\tau\), \(W_\tau\)) depending continuously on \(\tau\), and converging to \(\omega\) as \(\tau \to 0^+\).

Note that Theorem 9 makes several of the statements of our main result, Theorem 2. In Section 5 we will prove that theorem by invoking Theorems 4-9 for the family of Markov kernels \(p_\tau\).

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.

Nyström extension. We begin by describing the Nyström extension in RKHS. In what follows, \(\mathcal{H}\) is an RKHS on \(M\) with reproducing kernel \(k\), \(\nu\) an arbitrary finite Borel measure with compact support \(\subseteq M\), and \(K : L^2(\nu) \to \mathcal{H}\) the corresponding integral operator defined via 4. The Nyström extension operator \(\mathcal{N} : D(\mathcal{N}) \to \mathcal{H}\), with \(D(\mathcal{N}) \subseteq L^2(\nu)\), 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

\[\psi_j = \lambda_j^{-1/2} K \phi_j, \quad j \in J,\]

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

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

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 ran\(\mathcal{N} = \text{ran} \mathcal{N} = \text{span} \{\psi_j\}_{j \in J}\). Moreover, \(K^* \mathcal{N}\) and \(\mathcal{N} K^*\) reduce to the identity operators on \(D(\mathcal{N})\) and ran\(\mathcal{N}\), respectively. In fact, upon restriction to \(S\), ran\(\mathcal{N}\) coincides with the RKHS \(\mathcal{H}(S)\), and \(\{\psi_j|_S\}_{j \in J}\) forms an orthonormal basis of the latter space. If, in addition, the kernel \(k\) is strictly positive definite, 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 \(\mathcal{N}\), defined as the unique bounded operator \(\mathcal{N}^\dagger : \mathcal{H} \to L^2(\mu)\) satisfying (i) \(\ker \mathcal{N}^\dagger = \ker \mathcal{N}^\perp\); (ii) \(\text{ran} \mathcal{N}^\dagger = \ker \mathcal{N}^\perp\); 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 for the invariant measure \(\mu\) of the system, but also for discrete sampling measures encountered in data-driven approximation schemes.
Strong resolvent convergence. In order to prove the various spectral convergence claims made in Sections \ref{sec:approximation} and \ref{sec:proof} 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 $V$ appears difficult to verify. For example, in an approximation of $V$ by $\tilde{V}_\tau = G_{\tau}^{1/2} V G_{\tau}^{1/2}$, even though $G_{\tau}^{1/2} f$ converges to $f$ as $\tau \to 0^+$ for every $f \in D(V)$, $VG_{\tau}^{1/2}$ may not converge to $V f$, as $V$ is unbounded.

On the other hand, for every $\tau$, the resolvents $(\tau - B_\tau)^{-1}$ converge to $V$ weakly, but this type of convergence turns out to be too weak for the types of spectral convergence we seek to establish. Instead, for our purposes, it will be sufficient to establish strong convergence on the domain of $V$, which sits between weak and strong convergence, and is sufficient to establish our spectral convergence claims.

To wit, let $T : D(T) \to H$ be a closed operator on a Hilbert space $H$, and consider a sequence of operators $T_\tau : D(T_\tau) \to H$ indexed by a parameter $\tau > 0$. The sequence $T_\tau$ is said to converge to $T$ as $\tau \to 0^+$ in strong resolvent sense if for every complex number $\rho$ in the resolvent set of $T$, not lying on the imaginary line, the resolvents $(\rho - T)^{-1}$ converge to $(\rho - T)^{-1}$ strongly. Following \cite{[2]}, we will say that the sequence $T_\tau$ is $p$-continuous if every $T_\tau$ is bounded, and the function $\tau \mapsto \|P_2(iT_\tau)\|$ is continuous for all quadratic polynomials $P_2$ with real coefficients. In the case of skew-adjoint operators, which are necessarily densely-defined and closed, strong resolvent convergence implies the following convergence results for spectra and Borel functional calculi.

**Proposition 10.** Suppose that $T_\tau : D(T_\tau) \to H$ is a sequence of skew-adjoint operators converging in strong resolvent sense as $\tau \to 0^+$ to an operator $T : D(T) \to H$; $(T$ is necessarily skew-adjoint). Let $\Theta_\tau : B(H) \to \mathcal{L}(H)$ and $\Theta : B(H) \to \mathcal{L}(H)$ be the spectral measures of $T_\tau$ and $T$, respectively. Then:

(i) For every bounded, continuous function $Z : i\mathbb{R} \to \mathbb{C}$, $Z(T_\tau)$ converges strongly to $Z(T)$.

(ii) Let $J \subset J' \subset i\mathbb{R}$ be two bounded intervals. Then for every $\phi \in L^2(\mu)$, $\limsup_{\tau \to 0^+} \|1_{J'}(T_\tau)\phi\|_\mu \leq \|1_{J}(T)\phi\|_\mu$.

(iii) For every bounded, Borel-measurable set $U$ such that $\Theta(\partial U) = 0$, $\Theta_\tau(U)$ converges strongly to $\Theta(U)$.

(iv) For every bounded, Borel-measurable function $Z : i\mathbb{R} \to \mathbb{C}$ 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, (ii) hold for every Borel-measurable set $U \subset \mathbb{R}$, and (iii) for every bounded Borel-measurable function $Z : i\mathbb{R} \to \mathbb{C}$, and

(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 of $\theta_\tau \in i\mathbb{R}$ of eigenvalues of $T_\tau$ such that $\lim_{\tau \to 0^+} \theta_\tau = \theta$. Moreover, if the sequence $T_\tau$ is $p$-continuous, the curve $\tau \mapsto \theta_\tau$ is continuous.

**Proof.** By Proposition 10.1.9. \cite{[2]} Claim (i) is actually an equivalent characterization of strong resolvent convergence. Claims (v) and (vi) are results from the perturbation theory of normal operators, proved in Theorem 2, Chapter 8, \cite{[55]}, and Corollary 10.2.2. \cite{[3]} respectively.

To prove Claim (ii), let $f : \mathbb{R} \to \mathbb{R}$ be a piecewise linear continuous function which equals $1$ on $J'$ and with support contained in $J$. Then note that the inequality $1_{J'} \leq f \leq 1_J$ everywhere. Thus, for each $n \in \mathbb{N}$, $1_{J'}(A_n) \leq f(A_n)$ and $f(A) \leq 1_J(A)$. Since $f$ is continuous and bounded, by Claim (ii), $f(A_n)$ converges strongly to $f(A)$. The proof of Claim (ii) can now be completed by the following inequality.

$$\limsup_{n \to \infty} \|1_{J'}(A_n)\phi\|_\mu \leq \limsup_{n \to \infty} \|f(A_n)\phi\|_\mu = \|f(A)\phi\|_\mu \leq \|1_J(A)\phi\|_\mu.$$

We will now prove Claim (iii). For every subset $A \subset \mathbb{R}$, define $I_A : \mathbb{R} \to \mathbb{R}$, $I_A(x) := x1_A(x)$.

Note that if $A$ is a bounded, Borel measurable set, then $I_A$ is a bounded, Borel measurable function. For any operator $T$, the operator $I_A(T)$ is the spectral truncation of $T$ to the set $A$. Moreover, if $T$ is self-adjoint,
Lemma 11. Let $T_\tau : D(T_\tau) \to H$ and $T : D(T) \to H$ be the skew-adjoint operators from Theorem 10. 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 of $T$, 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 Theorem 5 of [53]. Claims (ii), (iii) and (v) follow from Propositions 10.1.18 and 10.1.8 respectively of [53]. There the statements are for self-adjoint operators, but they apply to skew-adjoint operators as well.
Results from analysis on manifolds. For the remainder of this section, we 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 $M$, or the embedding $F : M \to Y$ into the data space $Y$ from Section 6), 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 $\nabla^j \alpha \in C^{q-j}(M; T^{n+j})$ up to order $j = r$. 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 $\|\alpha\|_{C^q(M; T^n M)} = \sum_{j=0}^q \|\nabla^j \alpha\|_{C^0(M; T^{n+j})}$, where $\|\cdot\|_x$ denotes pointwise Riemannian norms on tensors. The case $C^q(M; T^{n} M)$ with $n = 0$ corresponds to the $C^0(M)$ 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. Given an RKHS $H$ on $M$ with a $C^r$ reproducing kernel, the inclusion map $\iota_H : H \to C^0(M)$ is bounded [67].

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

**Lemma 12.** 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 $\nabla$, the claim follows by an application of the Cauchy-Schwartz inequality for the Riemannian inner product, viz.

$\|\Xi f\|_{C^0(M)} = \|\Xi \cdot \grad f\|_{C^0(M)} \leq \|\Xi\|_{C^0(M; TM)} \|\grad f\|_{C^0(M; TM)} = \|\Xi\|_{C^0(M; TM)} \|\nabla f\|_{C^0(M; TM)} \leq \|\Xi\|_{C^0(M; TM)} \|f\|_{C^1(M)}$.

In particular, under Assumption 1, the dynamical flow $\Phi^t$ on $M$ is generated by a vector field $\tilde{V} \in C^0(M; TM)$, for which Lemma 12 applies. This vector field is related to the generator $V$ by a conjugacy with the inclusion maps $\iota : C^0(M) \to L^2(\mu); \iota_1 : C^1(M) \to L^2(\mu)$, namely, $\iota \tilde{V} = V_{\iota_1}$.

The following is a well known result from analysis, and the proof is left to the reader.

**Lemma 13 (C1 convergence theorem).** Let $M$ be a compact, connected, $C^1$ manifold equipped with a $C^0$ Riemannian metric. Let also $f_j : M \to \mathbb{R}$ be a sequence of tensor fields in $C^1(M; T^{n} M)$, such that the sequence $\{\|\nabla f_j\|_{C^0(M; T^{n} M)}\}_{j \in \mathbb{N}}$ 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 $F \in C^1(M; T^{n} M)$ 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 14.** Let $M$ be a compact, connected, $C^r$ manifold with $r \geq 1$, equipped with a $C^{r-1}$ Riemannian metric. Suppose that $f_j : M \to \mathbb{R}$ is a sequence of real-valued $C^r(M)$ functions such that the sequence $\{\|f_j\|_{C^r(M)}\}_{j \in \mathbb{N}}$ is summable, and 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 on $q \in \{1, \ldots, r\}$, invoking Lemma 13 as needed. First, note that summability of $\{\|f_j\|_{C^r(M)}\}_{j \in \mathbb{N}}$ implies summability of $\{\|\nabla^q f_j\|_{C^0(M; T^{n} M)}\}_{j \in \mathbb{N}}$ for all $q \in \{1, \ldots, r\}$. Because of this, and the fact that $\sum_{j \in \mathbb{N}} f_j(x)$ converges, it follows from Lemma 13 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 $\{\|\nabla^q f_j\|_{C^0(M; T^{n} M)}\}_{j \in \mathbb{N}}$, it follows from Lemma 13 that $\nabla^q F = \sum_{j \in \mathbb{N}} \nabla^q f_j$ converges in $C^1(M; T^{n} M)$ norm. Thus, $\nabla^{q+1} F = \sum_{j \in \mathbb{N}} \nabla^{q+1} f_j$ converges in $C^q(M; T^{n} M)$ 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.
5. Proof of Theorems 4–7

Proof of Theorem [4] By Assumption [3], \( H \) is a subspace of \( C^1(M) \), and therefore for every \( f \in H \), \( K^* f = \iota_1 f \), where \( \iota_1 \) is the \( C^0(M) \to L^2(\mu) \) inclusion map. Claim (i) then follows from the facts that \( \text{ran} \, \iota_1 \subset D(V) \), and \( K \) is bounded. To prove Claim (ii), let \( K' : L^2(\mu) \to C^0(M) \) be the kernel integral operator associated with the continuous kernel \( k' \), and \( \iota \) the \( C^0(M) \to L^2(\mu) \) inclusion map. Because \( K' \) is a Hilbert–Schmidt integral operator on \( L^2(\mu) \), with operator norm bounded above by its Hilbert–Schmidt norm, \( \|K'\|_{C^0(X \times X)} \leq \|k'\|_{C^0(X \times X)} \), 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 \( K' f(x) = \langle k'(x, \cdot), f \rangle_\mu \). Now because \( k \) lies in \( C^1(M \times M) \), for every \( x \in M \) the function \( k'(x, \cdot) \) is the \( C^0(M) \) limit \( k'(x, \cdot) = \lim_{t \to 0} g_t \), where \( g_t = (k(\Phi_t(x), \cdot) - k(x, \cdot))/t \), and by continuity of inner products,

\[
K' f(x) = \langle k'(x, \cdot), f \rangle_\mu = \langle \lim_{t \to 0} g_t, f \rangle_\mu = \lim_{t \to 0} \frac{1}{t} \left[ \langle k(\Phi_t(x), \cdot), f \rangle_\mu - \langle k(x, \cdot), f \rangle_\mu \right] = \tilde{V} K f(x).
\]

Therefore, because \( \text{ran} \, K \subset C^1(M) \), for any \( f \in L^2(\mu) \) we have

\[
i K' f = \iota \tilde{V} K f = \iota \iota_1 K f = \iota K' f = \iota VG f,
\]

proving Claim (ii). Finally, to prove Claim (iii), note by definition,

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

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

\[
\langle h, g \rangle_\mu = \langle -Af, g \rangle_\mu = -\langle VG f, g \rangle_\mu = \langle G f, V g \rangle_\mu = \langle f, GV g \rangle_\mu.
\]

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

Proof of Theorem [3] We begin with the proof of Claim (i). The inclusion ran \( K^* \subset D(V) \) holds because \( H \) is a subspace of \( C^1 \). To prove that \( VK^* \) is bounded, we make use of Using Lemma [12] and the fact that the inclusion map \( \iota_H : H \to C^1(M) \) is bounded [see Propositions 6.1-6.2, [37] ], we compute

\[
\|VK^* f\|_{L^2(\mu)} = \|\tilde{V} f\|_{L^2(\mu)} \leq \|\tilde{V}\|_{C^0(M)} \leq \|\tilde{V}\|_{C^1(M)} \leq \|\iota_H\| \|\tilde{V}\| \|f\|_H,
\]

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, note 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 \( x \in X \),

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

Moreover, \( \int_M \tilde{k}'(x, \cdot) f \, d\mu = \langle 1_M, \tilde{V} (k(x, \cdot)) f \rangle_\mu \) vanishes by skew-adjointness of \( V \). Using these results, we obtain

\[
Wf(x) = KV K^* f(x) = KV \iota_1 f(x) = K \tilde{V} f(x) = \int_M k(x, \cdot) \tilde{V} f \, d\mu = \int_M \tilde{k}'(x, \cdot) f \, d\mu. \quad \square
\]

Proof of Theorem [6] 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 [4] and therefore has a unique closed extension \( \overline{GV} : L^2(\mu) \to L^2(\mu) \) equal to \((GV)^*\). In order to verify that \( \overline{GV} = B \), it suffices to show that \( GV f = B f \) 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 $\iota f$ in this subspace, we have $B f = iK^* f$ and $G V f = \iota_1 K f$, where $K': L^2(\mu) \to C^0(M)$ is the integral operator with kernel $K'$, defined analogously to the operator $K'$ in the proof of Theorem \[4\] Employing the Leibniz rule as in the proof of Theorem \[5\] it is straightforward to verify that $B f$ is indeed equal to $G V f$, proving that $B$ is the unique closed extension of $G V$. Next, to show that $B$ is also an extension of $K^* W N$, it suffices to show that $G V \supseteq K^* W N$. To that end, note first that $K^* W N$ is a well defined operator by Theorem \[5\] and thus, substituting the definition for $W$ in \[5\], and using the fact that $K^* N$ is the identity on $D(N)$, we obtain

$$K^* W N = G V K^* N = G V |_{D(N)}.$$  

This shows that $K^* W N \subseteq G V \subset B$, confirming that $B$ is a closed extension of $K^* W N$. If $k$ is strictly positive, then $D(N)$ is dense, and by the bounded linear transformation theorem, $B$ is the unique closed extension of $K^* W N$. This completes the proof of Claim (i).

Next, to prove Claim (ii), note that because $B$ is compact, 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 $n \in \mathbb{N}$, $K^* W^n = B^n$, we arrive at the claimed identity,

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

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

**Proof of Theorem 7.** 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 \[7\] (with some abuse of notation as $J$ may be infinite). Let $U : L^2(\mu) \to \mathcal{H}$ be the unitary operator mapping $\phi_j$ to $\psi_j$. To prove the theorem, it suffices to show that $G^{1/2} V G^{1/2}$ is well-defined on a dense subspace of $L^2(\mu)$, and on that subspace, $G^{1/2} V G^{1/2}$ and $U^* U$ are equal. To verify that $G^{1/2} V G^{1/2}$ is densely defined, note first that $G^{1/2} \phi_j$ trivially vanishes for $j \notin J$, and therefore $G^{1/2} V G^{1/2} \phi_j$ is well-defined and vanishes too. Moreover if $j \in J$, $G^{1/2} \phi_j = K^* \psi_j$, and $G^{1/2} V G^{1/2} \phi_j$ is again well defined since ran $K^* \subset D(V)$. As a result, the domain of $G^{1/2} V G^{1/2}$ 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 $U^* U$ and $G^{1/2} V G^{1/2}$ are equal on this subspace, we observe that they have the same matrix elements in the $\{ \phi_j \}$ basis of $L^2(\mu)$, i.e., that $\langle \phi_i, G^{1/2} V G^{1/2} \phi_j \rangle_{L^2(\mu)}$ is equal to $\langle \phi_i, U^* U \phi_j \rangle_{\mathcal{H}}$ for all $i, j \in \mathbb{N}_0$. To verify this, note first that $U \phi_j = \psi_j$ for any $j \notin \mathbb{N}_0$, but because ker $K^* = \mathbb{R} K^1$, $K^* \psi_j$, and therefore $K^* U \phi_j$ vanish when $j \notin J$. Because $G^{1/2} \phi_j$ also vanishes in this case, we deduce that if either of $i$ and $j$ does not lie in $J$, the matrix elements $\langle \phi_i, U^* U \phi_j \rangle_{\mu}$ and $\langle \phi_j, G^{1/2} V G^{1/2} \phi_j \rangle_{\mu}$ both vanish. On the other hand, if $i, j \in J$, we have

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

We have thus shown that $U^* U$ and $G^{1/2} V G^{1/2}$ have the same matrix elements in an orthonormal basis of $L^2(\mu)$, and because the former operator is defined on the whole of $L^2(\mu)$ and the latter is densely defined, this implies that $V = U^* U$ is the unique closed extension of $G^{1/2} V G^{1/2}$. That $V$ is skew-adjoint and Hilbert-Schmidt follows immediately. \[\square\]

**6. Proof of Theorems 8 and 9.** We will need the following lemma, describing how to convert between eigenfunctions of the operators $A$, $B$, $\tilde{V}$, and $W$. The proof follows directly from the definitions of these operators, so the details will be omitted.
Lemma 15. Let Assumptions [7] and [8] hold with \( r = 1 \). Then,

(i) If \( \zeta \in \mathcal{H} \) is an eigenfunction of \( W \) at eigenvalue \( i\omega \), then \( K^*\zeta \) is an eigenfunction of \( B \) at eigenvalue \( i\omega \).

(ii) \( z' \) is an eigenfunction of \( A \) at eigenvalue \( i\omega \) iff \( Kz' \) is an eigenfunction of \( W \) at eigenvalue \( i\omega \).

(iii) If \( z' \in L^2(\mu) \) is an eigenfunction of \( A \) with at eigenvalue \( i\omega \), then \( G^{1/2}z' \) is an eigenfunction of \( \tilde{V} \) at eigenvalue \( i\omega \).

(iv) If \( \tilde{z} \) 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 8. In what follows, \( \sigma_a(T) \) will denote the set of eigenvalues of a linear operator \( T \), including multiplicities. Starting from Claim (i), because all of \( A, B, \tilde{V}, \) and \( W \) are compact operators, in order to verify equality of their spectra, it is sufficient to show that \( \sigma_a(A), \sigma_a(B), \sigma_a(\tilde{V}), \) and \( \sigma_a(W) \) are equal. First, note that \( \sigma_a(W) = \sigma_a(\tilde{V}) \) follows from the fact that \( W \) and \( \tilde{V} \) are unitarily equivalent. Moreover, by Lemma 15 \( \sigma_a(A) \subseteq \sigma_a(W) \subseteq i\mathbb{R} \), and because \( A \) is a real operator, it follows that \( \sigma_a(A) \) is symmetric about the origin of the imaginary line \( i\mathbb{R} \), so that

\[
\sigma_a(A) = -\sigma_a(A) = -\sigma_a(A^*) = -\sigma_a(-B) = \sigma_a(B).
\]

Thus, the equality of \( \sigma_a(A), \sigma_a(B), \sigma_a(\tilde{V}) \), and \( \sigma_a(W) \) will follow if it can be shown that \( \sigma_a(A) = \sigma_a(\tilde{V}) \). Indeed, it follows from Lemmas 15(iii) and 15(iv) that \( \sigma_a(A) \subseteq \sigma_a(\tilde{V}) \) and \( \sigma_a(\tilde{V}) \subseteq \sigma_a(B) \). These relationships, together with the fact that \( \sigma_a(A) = \sigma_a(B) \), imply that \( \sigma_a(\tilde{V}) = \sigma_a(A) \), and thus \( \sigma(A) = \sigma(B) = \sigma(\tilde{V}) = \sigma(W) \), as claimed. This completes the proof of Claim (i).

To prove Claim (ii), note that under Markovianity and strict positive-definiteness of \( k, Gf = f \) implies that \( f \) is \( \mu \)-a.e. constant. In addition, by ergodicity, \( Vf = 0 \) 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 \( \sigma_a(A) = \sigma_a(B) = \sigma_a(\tilde{V}) = \sigma_a(W) \), 0 is also a simple eigenvalue of \( B, \tilde{V}, \) and \( W \), and the constancy of the corresponding eigenfunctions follows directly from the definition of these operators.

Next, for Claim (iii), fix a nonzero eigenvalue \( i\omega \) 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 15(iii) and Claim (i) this set is actually a basis of the eigenspace of \( \tilde{V} \) at eigenvalue \( i\omega \). As result, every eigenfunction of \( \tilde{V} \) at nonzero corresponding eigenvalue lies in the range of \( G^{1/2} \). Moreover, it follows from Claim (iii) 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 \( V \), and thus in the domain of \( G^{-1/2} \), as claimed. Next, since \( \tilde{z}_j \in \text{ran} G^{1/2} \),

\[
\tilde{V}\tilde{z}_j = \tilde{V}G^{1/2}G^{-1/2}\tilde{z}_j = G^{1/2}VGG^{-1/2}\tilde{z}_j = G^{1/2}VGG^{-1/2}\tilde{z}_j = G^{1/2}A\tilde{z}_j.
\]

It then follows that \( A\tilde{z}_j = i\omega \tilde{z}_j \). It remains to be proved that the \( \tilde{z}_j \) form an unconditional Schauder basis.

First note that in Claim (iv), each of the \( \tilde{z}_j \) is an eigenfunction of \( \tilde{B} \) with eigenvalue \( i\omega_j \), by Lemma 15(iv). Moreover, for every \( j, k \in \mathbb{N}_0 \), we have

\[
\langle \tilde{z}_j, \tilde{z}_k \rangle_\mu = \langle G^{-1/2}\tilde{z}_j, G^{1/2}\tilde{z}_k \rangle_\mu = \langle \tilde{z}_j, \tilde{z}_k \rangle_\mu = \delta_{jk},
\]

which shows that \( \{\tilde{z}_0, \tilde{z}_1, \ldots\} \) is a dual sequence to \( \{z_0, z_1, \ldots\} \). Fix the \( \phi_j \)s from (9) as an orthonormal basis for \( L^2(\mu) \), and in what follows, this basis will be used to represent operators and collections of vectors as \( \mathbb{N} \times \mathbb{N} \) matrices. Let \( Z' \), \( L \) and \( U \) denote the matrices

\[
Z'_{ij} := \langle \phi_i, \phi_j \rangle_\mu, \quad L_{ij} := \langle z_i, \phi_j \rangle_\mu, \quad U_{ij} := \langle z_i, \phi_j \rangle_\mu, \quad i, j \in \mathbb{N}_0
\]

Then note that \( Z' \) and \( L \) have \( \ell^2 \) summable columns and rows respectively. Moreover,

\[
LZ' = \text{Id}, \quad L = U^*\Lambda^{1/2}, \quad Z' = \Lambda^{-1/2}U, \quad \text{where } \Lambda = \text{diag}(\lambda_0, \lambda_1, \ldots).
\]
Now define $I_k$ to be the the diagonal matrix with the first $k$ entries 1 and the rest 0. Then,
\[ Z' I_k L = \Lambda^{-1/2} U I_k U^* \Lambda^{1/2} = \Lambda^{-1/2} I_k \Lambda^{1/2} = I_k, \]
\[ \lim_{k \to \infty} Z' I_k L \xrightarrow{\text{strong}} I_d \tag{12} \]
By Lemma 2.1, \[ (12) \] implies that the columns of $Z'$ correspond to a Schauder basis in the chosen basis. This means that the $z_j$ form a Schauder basis. To see that it is also unconditional, not that if the $z_j$ are permuted, \[ (12) \] would still hold, but with the rows and columns of $U$, $Z'$, $L$ and $\Lambda$ permuted. Now every Schauder basis has a unique dual sequence, which is also a Schauder basis (e.g., \[ [58] \]). Thus \{ $z_j : j \in \mathbb{N}_0$ \} is an unconditional Schauder basis too, proving claims (iii and (iv).

In Claim (v), the fact that the $\zeta_j$ are eigenfunctions of $W$ follow from Lemma [15](ii). We also have
\[ \langle \zeta_i, \zeta_j \rangle_H = \langle K z_i', K z_j' \rangle_H = \langle z_i', G z_j' \rangle_{L^2(\mu)} = \langle z_i', z_j \rangle_{L^2(\mu)} = \delta_{ij}, \]
establishing orthonormality of the $\zeta_j$, and proving the claim.

Finally, in Claim (vi), first note that all of the summations are well defined and independent of ordering due to the unconditionality of all the bases involved. The results for $\bar{V}$ and $W$ follow from standard properties of compact, skew-adjoint operators. we will only prove the representation of $B$, and omit the proofs for the representations of $A$ as it is exactly analogous to the proof for $B$. Every \( f \in L^2(\mu) \) has an expansion $f = \sum_{j=0}^{\infty} a_j z_j$, with the summation holding in $L^2(\mu)$-sense. Then since $B z_j = i \omega_j z_j$ and since $B$ is a bounded operator,
\[ B(f) = B \sum_{j=0}^{\infty} a_j z_j = \sum_{j=0}^{\infty} a_j B(z_j) = \sum_{j=0}^{\infty} a_j i \omega_j z_j. \]
The fact that $B(f) = \sum_{j=0}^{\infty} a_j z_j$ will follow from this identity for the coefficients $a_j$ ,
\[ \langle z_j', f \rangle_{\mu} = \langle z_j', \sum_{k \in \mathbb{N}} a_k z_k \rangle_{\mu} = \sum_{k \in \mathbb{N}} a_k \langle z_j', z_k \rangle_{\mu} = \sum_{k \in \mathbb{N}} a_k \delta_{j,k} = a_j. \]
We have thus shown that $B = \int_{\Omega} i \omega d E(\omega)$, and $E(\emptyset) = 0$. For a fixed Borel set $U \subset \mathbb{C}$, $\pi_B(U)$ is a bounded linear operator. The countable additivity $\pi_B$ follows from directly from the definition. Note that for each $i, j, \in \mathbb{N}_0$, each of the maps $E_j$ is a projection, satisfying
\[ E_i \circ E_j = \delta_{i,j} E_j. \]
Thus, $E(U)E(V) = E(U \cap V)$. Thus, $\pi_B$ is a projection-valued spectral measure, with a discrete support \{ $i \omega_j : j \in \mathbb{N}_0$ \}. This completes the proof of the claim, and also of Theorem 8.

**Proof of Theorem 4**. We will only prove the claim for $B_\tau$, as the proofs for $A_\tau$ are exactly analogous. We begin with the proof of Claim (i). Since $G_\tau \to I_d$ strongly by Assumption 3,
\[ \lim_{\tau \to 0^+} \|(B_\tau - V)(f)\|_{\mu} = \lim_{\tau \to 0^+} \|(G_\tau V - V)(f)\|_{\mu} = \lim_{\tau \to 0^+} \|(G_\tau - I_d)(V f)\|_{\mu} = 0, \quad \forall f \in D(V). \]

Let $\bar{E}_\tau$ denote the spectral measure of $\bar{V}_\tau$. Now note that for every $\tau > 0$,
\[ B_\tau \circ G_\tau^{1/2} = G_\tau^{1/2} \circ \bar{V}_\tau, \quad E_\tau \circ G_\tau^{1/2} = G_\tau^{1/2} \circ \bar{E}_\tau, \]
\[ K^* \circ W_\tau = B_\tau \circ K^*, \quad K^* \circ \bar{E}_\tau = E_\tau \circ K^*. \tag{13} \]

The first equality follows from definition, the second follows from the first and from the representation formulae in Theorem 3. To prove Claim (ii), we use the condition in Assumption 3 that $\bar{V}_\tau$ converges strongly to $V$ on the space $D(V^2)$, as $\tau \to 0^+$. Since $\bar{V}_\tau$ is skew-adjoint, by Lemma [11](i), $\bar{V}_\tau$ converges to $V$ in the strong resolvent sense. Thus by Proposition [10](i), for every continuous, bounded function $Z : i \mathbb{R} \to \mathbb{R}$, $Z(\bar{V}_\tau)$ converges strongly to $Z$($V$). Since $G_\tau^{1/2}$ is a uniformly bounded family of operators converging to the identity, we have
\[ G_\tau^{1/2} Z(\bar{V}_\tau) \xrightarrow{\text{strong}} Z(V), \quad \text{as } \tau \to 0^+. \]
Thus by (13), $Z(B_\tau)G_{\tau}^{1/2}$ also converges strongly to $Z(V)$. Again, by the Uniform boundedness principle, $Z(B_\tau)$ is a uniformly bounded family of operators, hence $Z(B_\tau)(Id - G_{\tau}^{1/2})$ converges strongly to zero. Adding these two limits gives that $Z(B_\tau)$ converges strongly to $Z(V)$, as claimed. Claim (ii) now follows from the fact that for every bounded Borel-measurable function $\phi : i\mathbb{R} \to \mathbb{C}$, $\phi(B)$ and $K^*\phi(W)N$ are equal on $D(N)$. Claim (iii) follows in a similar manner from Proposition (10)(iii). Claim (iv) follows from Proposition (10)(vi).

7. Proof of Theorems [1, 2] and Corollary [3].

Proof of Theorem [7]. Since $p_\tau$ if it is well defined, is symmetric in $x$ and $y$, it is enough to prove that for fixed $y \in M$, $p_\tau(\cdot, y)$ is a $C^r$ function. Note that

$$r_{\tau,j} := \lambda_j^{(\tau)}/\lambda_j, \quad \{r_{\tau,j}\}_{j=0}^{\infty} \in \ell^1, \quad \forall \tau > 0.$$  

By Mercer’s theorem [59], $p(x, y) := \sum_{j=0}^{\infty} \psi_j(x)\psi_j(y)$ converges absolutely and uniformly on $X \times X$. Since $r_{\tau,j}$ converges to zero, the series for $p_\tau(x, y)$ also converges absolutely and uniformly on $X \times X$. Thus condition (i) of Lemma [14] is satisfied. For every $j \in \mathbb{N}_0$ and $\alpha \in \{1, \ldots, r\}$,

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

Thus $\|\psi_{\tau,j}\|_{C^r} \leq r_{\tau,j}\lambda_j^{-1/2}\|p\|_{C^r}$. Let $y$ be fixed, and let $f_j = \psi_{\tau,j}(y)\psi_j$. Then

$$\|f_j\|_{C^r} \leq \|\psi_{\tau,j}(y)\|_{C^r(M)}\|\psi_{\tau,j}\|_{C^r} \leq \|p\|_{C^r}r_{\tau,j}\lambda_j^{-1}.$$  

Thus $\{\|f_j\|_{C^r}\}_{j=0}^{\infty} \in \ell^1$, and condition (ii) of Lemma [14] is satisfied. So Lemma [14] applies and therefore,

$$p_\tau(x, y) = \sum_{j=0}^{\infty} \psi_{\tau,j}(x)\psi_{\tau,j}(y) = \sum_{j=0}^{\infty} r_{\tau,j}\psi_j(x)\psi_j(y) = \sum_{j=0}^{\infty} f_j,$$

converges in $C^r(M)$ norm, and therefore, the limit $p_\tau(x, y)$ is defined everywhere and forms a $C^r$ sum as claimed. The symmetry of $p_\tau$ follows again from its Mercer representation. Moreover, the functions $\phi_j$ are still eigenfunctions of $P_\tau$, with eigenvalues $\lambda_j^\tau$. This in turn implies that $P_\tau$ is also positive definite and $p_\tau$ is a Markov kernel.

In Claim (i), the set $\{\psi_{\tau,j} : j \in \mathbb{N}_0\}$ is analogous to the $\psi_j$-basis from [9] and is therefore an orthonormal basis. For every $j \in \mathbb{N}_0$, the function $1/\sqrt{\lambda_j}\psi_j$ equals $\lambda_j^{-1/2}\psi_j$, and thus lies in $H_\tau$. However, this is in the same $L^2(\mu)$ equivalence class as $\phi_j$, and the $\phi_j$s form an orthonormal basis for $L^2(\mu)$. Thus $H_\tau$ is dense in $L^2(\mu)$. The show that $H \supset H_{\tau_1} \supset H_{\tau_2}$, we need the following inequality.

$$\lambda_j > \lambda_j^{(\tau_2)} = \exp \left(\tau_2(\lambda_j^{-1} - 1)\right) > \exp \left(\tau_1(\lambda_j^{-1} - 1)\right) = \lambda_j^{(\tau_1)}.$$  

The claim now follows from the fact that $H = \{\sum_{j=0}^{\infty} a_j/\sqrt{\lambda_j}\phi_j : \sum_{j=0}^{\infty} |a_j|^2 < \infty\}$ and $H_\tau = \{\sum_{j=0}^{\infty} a_j/\sqrt{\lambda_j^{(\tau)}}\phi_j : \sum_{j=0}^{\infty} |a_j|^2 < \infty\}$. This proves Claim (i).

Note that the orthonormal basis of $\phi_j$s are also eigenfunctions of $G_\tau$, with eigenvalues $\lambda_{\tau,j}$. Thus $G_\tau$ is strictly positive definite and self adjoint and compact. The Markov property follows from the observation that $1 = \lambda_{\tau,0} > \lambda_{\tau,1} \geq \ldots$, proving Claim (ii).

The semi-group property of $G_\tau$ follows from the fact the $\phi_j$s form an orthonormal eigenbasis for all the $G_{\tau,j}$ with eigenvalues $\lambda_{\tau,j}$, and for each $j \in \mathbb{N}_0$, $\lambda_{\tau,j}^{(\tau_1+\tau_2)} = \lambda_{\tau,j}^{(\tau_1)}\lambda_{\tau,j}^{(\tau_2)}$. To prove strong continuity of this semi-group, it is enough to prove continuity as $\tau \to 0^+$. Let $f \in L^2(\mu)$, then $f$ can be written as the sum $f = \sum_{j=0}^{\infty} a_j\phi_j$. For simplicity, assume that $\|f\|_{\mu} = \sum_{j=0}^{\infty} |a_j|^2 = 1$. Let $\epsilon > 0$ be fixed, it is enough to
show that \( \lim_{\tau \to 0^+} ||(P_\tau - \text{Id})(f)|| < 2\epsilon \). Let \( f_N \) be the partial sum \( \sum_{j=0}^{N-1} a_j \phi_j \). Then for \( N \) large enough, \( ||f - f_N||_\mu < \epsilon \). Then,

\[
(P_\tau - \text{Id})(f) = (P_\tau - \text{Id})(f) + (P_\tau - \text{Id})(f - f_N) = \sum_{j=0}^{N} a_j \left( \lambda_j^\tau - 1 \right) \phi_j + (P_\tau - \text{Id})(f - f_N)
\]

Since \( P_\tau \) is a positive definite Markov operator, \( ||P_\tau|| \leq 1 \) and therefore, the last term can be bounded as \( ||(P_\tau - \text{Id})(f - f_N)||_\mu < 2\epsilon \). Now note that for each \( j \), \( \lambda_j^\tau - 1 = \exp\left( (1 - \lambda_j^{-1}) \tau \right) - 1 \) converges to 0 as \( \tau \to 0^+ \). This concludes the proof of Theorem 4.

**Proof of Theorem 5** Claim (i) of the theorem follows from Theorem 5 (ii); Claim (ii) follows from theorem 6 and Claims (iii)-(vi) will follow from Theorems 8 and 9 if we can show that \( p_\tau \) satisfies Assumption 4. The condition in this assumption that \( G_\tau \) converges pointwise to the identity has already been proven to hold, in Theorem 4 (iii). It thus remains to be shown that \( V_\tau \) converges pointwise to \( V \) on \( D(V^2) \), as \( \tau \to 0^+ \).

Since \( G_\tau \) is a semigroup, it is equivalent to show that \( G_\tau V G_\tau \) converges pointwise to \( V \) on \( D(V^2) \). Since \( G_\tau \) is uniformly bounded and converges to the identity, it is equivalent to show that \( A_\tau = V G_\tau \) converges pointwise to \( V \) on \( D(V^2) \). We will prove this using three small observations (A1)-(A3). Fix an \( f \in D(V^2) \) and define \( f_\tau = P_\tau f \).

(A1) \( P_\tau f_\tau \in D(V^2) \cap \mathcal{H} \), for every \( \tau > 0 \) : Note that \( f_\tau \in \text{ran} P_\tau \), and since \( p_\tau \) is a \( C^2 \) kernel, \( \text{ran}(P_\tau) \subset C^2(M) \subset D(V^2) \). Since the range of \( P_\tau \) lies in \( \mathcal{H}, f_\tau \in \mathcal{H}, \) and \( H \subset \mathcal{H} \) by Theorem 4 (i).

(A2) \( f_\tau \) converges to \( f \) in \( \mathcal{H} \)-norm as \( \tau \to 0^+ \) : Expanding \( f \) in the \( \phi_j \) basis gives \( f = \sum_{j=0}^{\infty} a_j \phi_j \) and \( f_\tau = \sum_{j=0}^{\infty} a_j \lambda_j \phi_j \), thus

\[
||f_\tau - f||^2_\mathcal{H} = \sum_{j=0}^{\infty} (1 - \lambda_j) a_j^2 ||\phi_j||^2 \leq \sum_{j=0}^{N} (1 - \lambda_j) a_j^2 ||\phi_j||^2 + \sum_{j=N+1}^{\infty} (1 - \lambda_j) a_j^2 ||\phi_j||^2.
\]

\[
\lim_{\tau \to 0^+} ||f_\tau - f||^2_\mathcal{H} \leq \sum_{j=0}^{N} ||a_j||^2 ||\phi_j||^2 + \sum_{j=N+1}^{\infty} ||a_j||^2 ||\phi_j||^2
\]

The inequality on the second line follows from the fact that \( \lambda_j \in (0, 1) \). Since \( ||f||_\mathcal{H} = \sum_{j=0}^{\infty} ||a_j||^2 ||\phi_j||^2 \), the last term converges to 0 as \( N \to \infty \). Since \( N \) was arbitrary, the limit must be 0, as claimed.

(A3) \( V P_\tau f_\tau \) is a Cauchy sequence in \( L^2(\mu) \) : \( V P_\tau f_\tau : \mathcal{H} \to L^2(\mu) \) is a bounded operator by Theorem 5 (i), thus \( V P_\tau f_\tau \) is also bounded as an operator \( \mathcal{H} \to L^2(\mu) \). Therefore, since \( f_\tau \) is a Cauchy sequence in \( \mathcal{H}, V P_\tau f_\tau = V P_\tau f = A_\tau f \) is a Cauchy sequence in \( L^2(\mu) \).

Thus, \( f_\tau \) is a family of functions in \( \mathcal{H} \) and must converge to \( V f \), since \( f \in D(V^2) \) and \( D(V^2) \) is a core for \( D(V) \). This proves the claim and completes the proof of Theorem 2.

**Proof of Corollary 3** The functions \( \phi_j \) form an orthonormal basis of \( L^2(\mu) \) and therefore, for \( L \subset Z \) large enough, if \( f_\tau \) is defined to be the projection of \( f \) to the subspace spanned by \( \{\phi_0, \ldots, \phi_{L-1}\} \), then \( ||f - f_\tau||_\mu < \epsilon \). Then since \( U^t \) is an isometry on \( L^2(\mu) \), for all \( t \in \mathbb{R}, ||U^t f - U^t f_\tau||_\mu < \epsilon \). Secondly, each \( \phi_j \) lies in \( H_\infty \), and since \( f_\tau \) is a finite linear combination of the \( \phi_j, f_\tau \in H_\infty \). This proves the first half of Corollary 3. The limit involving \( \tau \to 0^+ \) follows from Theorem 2 (v), and the formula for \( e^{W_\tau} f_\tau \) follows from the functional calculus for \( W_\tau \). This completes the proof of Corollary 3.

8. Data-driven approximations and convergence

We now take up the problem of approximating the operators in Theorem 2 from a finite time series of observed data and without prior knowledge of the dynamical flow \( \Phi^t \). As already alluded to in Section 1 besides a lack of knowledge of the underlying dynamics, this problem has the following issues:
1. The support $X$ of the ergodic invariant measure $\mu$ is generally a non-smooth subset of the state space manifold $L$, of zero Lebesgue measure (e.g., a fractal attractor of a dissipative dynamical system). As a result, it is not possible to construct bases of $L^2(\mu)$ (to be used for function and/or operator approximation) by restriction of smooth basis functions defined on $L$. Moreover, one does not have direct access to the invariant measure $\mu$ and the associated $L^2(\mu)$ space, but is limited to working with the sampling measure $\mu_N := \frac{1}{N} \sum_{n=0}^N \delta_{x_n}$ supported on the finite trajectory $\{x_0, \ldots, x_{N-1}\}$. Here, $\delta_y$ is the Dirac delta measure, supported on the point $y$.

2. In realistic experimental scenarios, the sampled states will not lie exactly on $X$.

3. Measurements are not taken continuously in time, preventing direct evaluation of the action of the generator $V$ on functions.

This is called a data-driven setting and the general assumptions are the following.

Assumption 5. There is a time-ordered dataset consisting consisting of the values $F(x_0), F(x_1), \ldots, F(x_{N-1})$, of an injective $C^2$ observation map $F : M \mapsto Y$ into a $C^2$ manifold $Y$, evaluated on the trajectory $x_n = \Phi^n(x_0)$ of the discrete time map $\Phi^n : L \mapsto L$. Here $\Delta t > 0$ is a a fixed sampling interval. $\rho : Y \times Y \mapsto \mathbb{R}$ is a $C^2$ symmetric, strictly positive-definite kernel on $Y$.

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 $Y = S^2$).

Data-driven Hilbert spaces. Associated with the sampling measure $\mu_N$ is an $N$-dimensional Hilbert space, $L^2(\mu_N)$, equipped with the inner product $(f, g)_{\mu_N} := \frac{1}{N} \sum_{n=0}^{N-1} f(x_n)g(x_n)$. This space consists of equivalence classes of complex-valued functions on $X$ having common values at the sampled states $x_0, \ldots, x_{N-1}$ (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. While, in general, there is no correspondence between the elements of $L^2(\mu_N)$ and $L^2(\mu)$ (allowing one, e.g., to perform approximation of functions and operators on $L^2(\mu)$ in subspaces constructed via $L^2(\mu_N)$ functions), the fact that our approximations of $\mu$ are based on operators acting on RKHSs allows us to construct data-driven subspaces for operator approximation through integral operators on $L^2(\mu_N)$ without invoking the data - inaccessible, $L^2(\mu)$ space.

We provide below a numerical procedure that has as input the assumptions in Assumption 5, and outputs a predictor for an arbitrary $C^0$ observable $f$. We then prove in Theorem 16 that the prediction converges to the true observable, in the limit of infinitely many data-points, i.e., as $N \to \infty$.

Input. A continuous observable $f \in C^0(M)$; a continuous, bounded function $Z : i\mathbb{R} \mapsto \mathbb{R}$; a parameter $L \in \mathbb{Z}$; and the data set of size $N$ mentioned in Assumption 5. Assume that the underlying dynamical system satisfies Assumption 1.

Step 1. The first step would be to construct a pull-back kernel from the kernel $\kappa$ on the data-space. This is done as follows.

$$k(x, y) := \rho(F(x), F(y)), \quad \forall x, y \in L.$$  

The assumptions on $\rho$ and $F$ in Assumption 5 ensure that $k$ is also a $C^2$ symmetric, strictly positive-definite kernel on $L$. See our discussion later on how to construct $\rho$ when the injectivity conditions on $F$ is not satisfied.

Step 2. Next, we construct a $C^2$, symmetric kernel $p_N$ from the kernel $k$ using Coifman and Hirn’s method to obtain bistochastic kernels, and then use $p$ to construct a data-driven family of kernels $p_{\tau,N}$ mentioned in Theorems 1 and 2.

$$d_N(x) := \int_X k(x, y)d\mu_N(y), \quad q_N(y) := \int_X k(x, y)/d_N(x)d\mu_N(x).$$

$$p_N(x, y) := \int_X k(x, z)k(y, z)/d_N(x)d\mu_N(y)d\mu_N(z).$$
Note that $p_N$ is a $C^2$ function on $M \times M$, and forms a strictly positive definite kernel, which is positive valued everywhere on $X \times X$. It therefore induces an RKHS $\mathcal{H}_N$ on $M$. Let $P_N : L^2(\mu_N) \to \mathcal{H}_N$ be the integral operator with kernel $p_N$. Then by construction, $P_N$ is a Markov operator, i.e., $P_N 1_N = 1_N$, where $1_N$ is the constant function on $L^2(\mu_N)$ equal to 1 everywhere. $P_N$ is symmetric and positive definite, it thus has $N$ positive eigenvalues $\lambda = \lambda_{N,0} > \lambda_{N,1} \geq \ldots \geq \lambda_{N,N-1} > 0$, and corresponding eigenfunctions $\phi_{N,j}$, which form an orthonormal basis for $L^2(\mu_N)$.

**Step 3.** We will use the kernel $p_N$ to define a new family of kernels similar to [7]. For each $\tau > 0$,

$$
\lambda_{\tau,N,j} := e^{\tau (1 - \lambda_{N,j}^2)}, \quad \psi^{(L)}_{\tau,N,j} := \lambda_{\tau,N,j}^{-1/2} \lambda_{N,j}^{-1/2} \psi_{N,j}, \quad 0 \leq j < N.
$$

$$
\psi_{N,j} := \lambda_{N,j}^{-1} \frac{1}{N} \sum_{n=0}^{N-1} \phi_{N,j}(x_n)p_N(x_n, \cdot), \quad p_{\tau,N}(x,y) := \sum_{j=0}^{N-1} \psi^{(L)}_{\tau,N,j}(x)\psi^{(L)}_{\tau,N,j}(y).
$$

For each $0 \leq j < N$, $\phi_{N,j} \in \mathcal{H}_N$. The kernel $p_{\tau,N}$ is symmetric and so induces an RKHS $\mathcal{H}_{\tau,N}$.

**Step 4.** We now compute data-driven versions of $W_{\tau,N}$, let $V_{\Delta t}$ be any finite difference operator on $L^2(\mu_N)$. For the integer parameter $L$ taken as input, define the following $L \times L$ matrices

$$
\left( V_{\tau,N,\Delta t}^{(L)} \right)_{i,j} := \langle \phi_{\tau,N,i}, V_{\Delta t} \phi_{\tau,N,j} \rangle_{\mu_N}, \quad \left( W_{\tau,N,\Delta t}^{(L)} \right)_{i,j} := \langle \lambda_{\tau,N,i} \lambda_{\tau,N,j} \rangle^{1/2} \left( V_{\tau,N,\Delta t}^{(L)} \right)_{i,j}.
$$

The matrices $\left( V_{\tau,N,\Delta t}^{(L)} \right)$ and $\left( W_{\tau,N,\Delta t}^{(L)} \right)$ are both $L \times L$ skew-symmetric matrices. Hence the former has $L$ eigenpairs which are indexed below by $j = 0, \ldots, L-1$.

$$
\left( W_{\tau,N,\Delta t}^{(L)} \right)_{\tau,N,\Delta t,j} = i \omega_{\tau,N,\Delta t,j} \left( \gamma_{\tau,N,\Delta t,j} \right)_{\tau,N,\Delta t,j} \in \mathbb{C}^{L \times 1}, \quad |\omega_{\tau,N,\Delta t,j}| \geq \ldots \geq |\omega_{\tau,N,\Delta t,L-1,j}|.
$$

The $L$-dimensional vector $\gamma_{\tau,N,\Delta t,j}$ represents a combination of the functions $\psi_{\tau,N,0,j}, \ldots, \psi_{\tau,N,L-1,j}$ and thus corresponds to a vector $\zeta_{\tau,N,\Delta t,j} \in \mathcal{H}_{\tau,N}$. Note that this is a continuous function. These vectors will be the basis in which we perform the functional calculus $Z(W_{\tau,N,\Delta t}^{(L)})$.

**Step 5.** Finally we construct a continuous, data-driven predictor for $Z(V)f$. Let $\Pi_N : C^0(M) \to L^2(\mu_N)$ be the inclusion map. Define,

$$
\hat{f}_{\tau,N,\Delta t,L} := \sum_{j=0}^{L-1} Z(i \omega_{\tau,N,\Delta t,j}^{(L)}) \langle \phi_{\tau,N,j}, \Pi_N f \rangle_{\mu_N} \gamma_{\tau,N,\Delta t,j}^{(L)}.
$$

**Theorem 16** (Pointwise forecasting). Let Assumptions 1, 2 and 7 hold. Let $f \in C^0(M)$ be a continuous observable and $Z : i\mathbb{R} \to \mathbb{R}$ is some bounded, bounded function. Then

$$
\lim_{\tau \to 0^+, L \to \infty, N, \Delta t \to \infty} \|Z(V)f - \hat{f}_{\tau,N,\Delta t,L}\|_{\mu} = 0. \quad (14)
$$

(i) Recall the eigenfunctions $\zeta_{\tau,j}$ and eigenvalues $\omega_{\tau,j}$ of $W_{\tau}$, from Corollary 3. Then,

$$
\lim_{\Delta t \to 0^+, N \Delta t \to \infty} \|\zeta_{\tau,N,\Delta t,j}^{(L)} - \zeta_{\tau,j}\|_{C^0(X)} = 0, \quad \lim_{\Delta t \to 0^+, N \Delta t \to \infty} |\omega_{\tau,N,\Delta t,j}^{(L)} - \omega_{\tau,j}| = 0, \quad \forall \tau > 0, L \in \mathbb{N}.
$$

(ii) In particular, taking $Z(it) = e^{it}$ makes $\hat{f}_{\tau,N,\Delta t,L}$ a predictor for $U^t f$

$$
\lim_{\tau \to 0^+, L \to \infty, \Delta t \to 0^+, N \Delta t \to \infty} \|U^t f - \hat{f}_{\tau,N,\Delta t,L}\|_{\mu} = 0.
$$
Remark. The function $\hat{f}_{\tau,N,\Delta t,L}$ is a continuous function which can be constructed from data, and which can be evaluated at points outside the sampled set $\{x_0, \ldots, x_{N-1}\}$. This latter property is known as \textit{out of sample evaluation}. $\hat{f}_{\tau,N,\Delta t,L}$ plays the role of a data-driven approximation of $Z(V)f$, and in the case when $Z(t) = e^{it}$, it is the time-$t$ predictor for the observable $f$.

Remark. The order in which the limits are taken in Theorem 16 is important and cannot be changed. The parameter $L$ is the number of eigenfunctions $z_{\tau,j}$ used in our prediction and is therefore like a resolution parameter. The first limits taken are of $N$ and $\Delta t$, this is the data-driven limit, that is, the limit at infinity many samples taken at arbitrarily small sampling intervals. For the spectral convergence of the data-driven operators to hold, the data-driven limits must be taken at a fixed resolution $L$. After this limit has been taken, $L$ may be increased to facilitate a finite rank approximation of the compact operator $Z(W_\tau)$. Finally, the limit $\tau \to 0^+$ is taken as the 0-limit of the Markov semi-group $P_\tau$.

Remark. The choice of $L$ is independent of the prediction time $t$, which is an important requirement for a numerical implementation of $\hat{f}_{\tau,N,\Delta t,L}$, since one can only compute only a finite number of eigenfunctions with reasonable accuracy. Notice that as $\tau$ is changed, the eigenfunctions $z_{\tau,j}$ and $z_{\tau,j}'$ just get scaled, so they need not be computed separately for every $\tau > 0$.

Proof of Theorem 16. First note that Claim (ii) follows directly from (14). We will prove (14) by performing a series of simplifications. Let $f_L$ denote the projection onto the span of the functions $\phi_0, \ldots, \phi_{L-1}$, the eigenfunctions of the integral operator $P$ (see (7)). For every $L \in \mathbb{N}$, $f_L \in D(N_\tau)$. Now observe,

$$\lim_{L \to \infty} \|Z(V)f - Z(V)f_L\|_\mu = 0, \quad \text{for fixed } L : \lim_{\tau \to 0^+} \|Z(V)f_L - P_\tau Z(W_\tau)N_\tau f_L\|_\mu = 0. \quad (15)$$

The first limit follows from Theorem 2(v), and the second follows from the fact that $Z(V)$ is a bounded operator. Let $\pi_{\tau,L}$ be the orthogonal projection onto the subspace of $H_\tau$, spanned by $\{N_\tau \phi_j : j = 0, \ldots, L - 1\}$. Define $W_{\tau,L} := \pi_{\tau,L} W_\tau \pi_{\tau,L}$, $\lim_{L \to \infty} \|W_{\tau,L} - W_\tau\|_{H_\tau} = 0$, $\lim_{L \to \infty} \|Z(W_\tau)N_\tau f_L - Z(W_{\tau,L})N_\tau f_L\|_{C^0(X)} = 0$. \quad (16)

The second equality follows from the fact that $W_\tau$ is a compact operator, so it is limit in the operator norm topology on $H_\tau$ of these finite rank approximations. The third limit follows from this operator norm convergence. Recall the RKHS $H_{\tau,N}$ defined after Step 3, the kernel $p_{\tau,N}$ induces an integral operator $P_{\tau,N} : L^2(\mu_N) \to H_{\tau,N}$. Similarly, one can define a Nystrom extension $N_{\tau,N} : L^2(\mu_N) \to H_{\tau,N}$. Now analogously to (8), define $W_{\tau,N,\Delta t} : H_{\tau,N} \to H_{\tau,N}$ as $W_{\tau,N,\Delta t} = P_{\tau,N}V_{\Delta t}P_{\tau,N}$. The relations between these various operators is summarized below.

$$\begin{align*}
C^0(M) & \xrightarrow{B_N} L^2(\mu_N) \xrightarrow{\iota} L^2(\mu_N) \xrightarrow{N_{\tau,N}} H_{\tau,N} \xrightarrow{P_{\tau,N}} L^2(\mu_N) \xrightarrow{V_{\Delta t}} L^2(\mu_N) \xrightarrow{P_{\tau,N}} H_{\tau,N}.
\end{align*}$$

Similarly to $W_{\tau,L}^{(L)}$, define $W_{\tau,N,\Delta t}$ to be the restriction of $W_{\tau,N,\Delta t}$ to the $L$-dimensional subspace of $H_{\tau,N}$ spanned by $\psi_{\tau,N,0}, \ldots, \psi_{\tau,N,L-1}$. An important observation now is that

$$\hat{f}_{\tau,N,\Delta t,L} = W_{\tau,N,\Delta t,N_{\tau,N}f}. \quad (17)$$

Now note that (14) is proved by the following limit, by combining it with the limits in (15), (16) and (17).

$$\lim_{\Delta t \to 0^+, \Delta \to \infty} \|Z(W_{\tau,N,\Delta t,N_{\tau,N}f} - Z(W_{\tau,N}N_{\tau,N}f)\|_{C^0(X)} = 0. \quad (18)$$

To prove (18), we have to compare the two $L$-dimensional operators $W_{\tau,L}^{(L)}$ and $W_{\tau,N,\Delta t}^{(L)}$. We need two steps.

(a) Compare the matrix representations of these operators in the bases $\{\psi_{\tau,0}, \ldots, \psi_{\tau,L-1}\}$ and $\{\psi_{\tau,N,0}, \ldots, \psi_{\tau,N,L-1}\}$.
respectively. 

(b) Prove that the basis functions themselves converge in $C^0(\mathcal{X})$ norm as $N \to \infty$. 

This will not only prove [13] but also claim (i) of the theorem. To prove (a), note that $\psi_{\tau,j}$ and $\psi_{\tau,N,j}$ are respectively, the $j$-th eigenfunctions of the kernel integral operators $P_{\tau}$ and $P_{\tau,N}$. The kernel $P_{\tau,N}$ is the data driven version of the kernel $P_{\tau}$, and it was shown in [30] that

$$\lim_{N \to \infty} \lambda_{\tau,N,j} = \lambda_{\tau,j}, \quad \lim_{N \to \infty} \| \psi_{\tau,N,j} - \psi_{\tau,j} \|_{C^0(\mathcal{X})} = 0, \quad \forall j \in \mathbb{N}_0.$$ 

It remains to prove (b). In the bases specified, the $(i,j)$-th entry of the matrix representing $W^{(L)}_{\tau}$ is

$$W^{(L)}_{\tau} \begin{pmatrix} i,j \end{pmatrix} = (\lambda_{\tau,i} \lambda_{\tau,j})^{1/2} \langle \phi_{\tau,i}, V_{\Delta t} \phi_{\tau,j} \rangle_{\mu}, \quad W^{(L)}_{\tau} \begin{pmatrix} i,j \end{pmatrix} = (\lambda_{\tau,i} \lambda_{\tau,j})^{1/2} \langle \phi_{\tau,N,i}, V_{\Delta t} \phi_{\tau,N,j} \rangle_{\nu}$$

The inner products were shown to converge as $N \Delta t \to \infty, \Delta t \to 0^+$ in Proposition 36, [18], and we have already established convergence of the $\lambda$s. This completes the proof of Theorem 10. 

**Kernels on data space.** To make the reproducing kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ strictly positive definite, the kernel $\kappa$ on the data-space $\mathcal{Y}$ itself must be strictly positive definite. There are many ways to chose such kernels, we use the following class of kernels called radial Gaussian kernels,

$$\kappa(y_1,y_2) = \exp \left( - \frac{d^2(y_1,y_2)}{\epsilon} \right),$$

where $d : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the Euclidean metric and $\epsilon$ a positive bandwidth parameter, are positive definite. Another requirement for the pull-back $k$ (Step 1) to be strictly positive definite is that the observation map $F : \mathcal{X} \to \mathcal{Y}$ be injective. This is often not satisfied in real-world applications, for example when $F$ is a low-dimensional observation map providing only partial information. The remedy for such cases is to incorporate delays into the map, i.e., using the following as a new observation map.

$$F_Q : \mathcal{X} \to \mathcal{Y}^Q, \quad F_Q(x) = \left( F(x), F(\Phi^{-1} \Delta t), \ldots, F(\Phi^{-Q-1} \Delta t) \right), \quad Q \in \mathbb{N}.$$ 

It can be shown [36] that under mild genericity assumptions on $F$ and $\Phi^{-\Delta t}$, for $Q$ large enough, $F_Q$ is a diffeomorphism onto its image.

9. Examples and discussions

In this section, we apply the procedure described in Section 8 to ergodic dynamical systems with different types of spectra. The goal is to demonstrate that the results of Theorems 2 and 16 hold, and that the method is effective in (i) identifying Koopman eigenfunctions and eigenfrequencies, and (ii) forecasting in chaotic systems. We consider the following two systems, whose spectra are respectively pure point and absolutely continuous.

1. A linear quasiperiodic flow $R_{\alpha_1,\alpha_2}$ on $\mathbb{T}^2$, defined as

$$dR_{\alpha_1,\alpha_2}^t(\theta)/dt = (\alpha_1, \alpha_2), \quad \theta = (\theta_1, \theta_2) \in \mathbb{T}^2, \quad \alpha_1 = 1, \quad \alpha_2 = \sqrt{30}, \quad (19)$$

and observed through the an embedding $F : \mathbb{T}^2 \to \mathbb{R}^2$ given by,

$$F(\theta_1, \theta_2) = (\sin \theta_1 \cos \theta_2 + \sin \theta_2, - \cos \theta_1 \cos \theta_2, \sin \theta_2). \quad (20)$$

This system has a pure point Koopman spectrum, consisting of eigenfrequencies of the form $n_1 \alpha_1 + n_2 \alpha_2$ with $n_1, n_2 \in \mathbb{Z}$. Because $\alpha_1$ and $\alpha_2$ are rationally independent, the set of eigenfrequencies lies dense in $\mathbb{R}$, which makes the problem of numerically distinguishing eigenfrequencies from non-eigenfrequencies non-trivial despite the simplicity of the underlying dynamics. 

26
2. The Lorenz 63 (L63) flow \( \Phi_{t63} : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \), generated by the \( \infty \) vector field \( \vec{V} \) with components \((V(x), V(y), V(z))\) at \((x, y, z) \in \mathbb{R}^3\) given by
\[
V(x) = \sigma(y - x), \quad V(y) = x(\rho - z) - y, \quad V(z) = xy - \beta z,
\]
where \( \beta = 8/3, \rho = 28, \) and \( \sigma = 10. \) The system is sampled through the observation map \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) with
\[
F(x, y, z) := (x, y, z).
\]
The L63 flow is known to have a chaotic attractor \( X_{l63} \subset \mathbb{R}^3 \) with fractal dimension \( \approx 2.06 \) \[60\], supporting a physical invariant measure \[48\] with a corresponding purely continuous spectrum of the generator \[61\]. That is, there exist no nonzero Koopman eigenfrequencies for this system.

**Methodology.** The following steps describe sequentially the entire numerical procedure carried out.

1. Numerical trajectories \( x_0, x_1, \ldots, x_{N-1} \), with \( x_n = \Phi^{n\Delta t}(x_0) \) of length \( N \) are generated, using a sampling interval \( \Delta t = 0.01 \) in all cases. In the L63 experiments, we let the system relax towards the attractor, and set \( x_0 \) to a state sampled after a long spinup time (4000 time units); that is, we formally assume that \( x_0 \) has converged to the ergodic attractor. We use the \texttt{ode45} solver of Matlab to compute the trajectories. For both the systems \( R_{\alpha_1,\alpha_2}^t \) and \( \Phi_{t63} \), \( N = 64000. \)

2. The observation map \( F \) described for each system is used to generate the respective time series \( F(x_0), F(x_1), \ldots, F(x_{N-1}) \). This dataset forms the basis of all subsequent computations. The kernel \( k \) is obtained by starting with a Gaussian kernel \( \rho \) on the data-space.

3. The eigenpairs \( (\phi_{N,j}, \lambda_{N,j}) \) are computed for \( j \in \{0, \ldots, L\} \) using Matlab’s \texttt{eigs} iterative solver, for \( L = 500, 750 \) respectively.
Figure 2: Representative eigenfunctions $\zeta_j$ of the compactified generator $W_\tau$ with $\tau = 10^{-4}$ for the L63 system [21]. Top row: Scatterplots of $\text{Re}(\zeta_j)$ on the training dataset embedded in $\mathbb{R}^3$. Bottom row: Time series of the eigenfunctions sampled along portions of the dynamical trajectory in the training data. 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. We have indicated the Dirichlet energy $\mathcal{E}_\omega$ of the computed eigenfunctions, and the value of $\omega$ for each eigenvalue. 

\[
\begin{align*}
\zeta_{11} & \quad \omega_{11} = 97.8, \quad E_{11} = 7.52 \\
\zeta_{31} & \quad \omega_{31} = 8.18, \quad E_{31} = 21 \\
\zeta_{43} & \quad \omega_{43} = 45.6, \quad E_{43} = 27.6
\end{align*}
\]
Figure 3: Eigenfrequencies $\omega_j$ of the compactified generators $W_\tau$ as a function of $\tau > 0$, for the torus rotation (20) and the L63 system (21). At each value of $\tau > 0$, we have calculated the Dirichlet energies of $L_\tau$ eigenfunctions of $W_\tau$ and then calculates the ratio of these energies to the smallest among these $L$ values. The frequencies are colored by the logarithm of the corresponding energy ratio. Only positive frequencies are shown for clarity.

**Results.** See figures 1 – 5 for the results of applying our methods.

1. The flow $R^t_{\alpha_1, \alpha_2}$ (19) has a purely discrete spectrum, and the Koopman eigenfunctions are the usual Fourier basis on the 2-torus. Fig. 1 shows that the eigenfunctions and eigenvalues of $W_\tau$ converge to those of $V$. The time-plot of these eigenfunctions show periodicity, which supports the fact that they converge to the Koopman eigenfunctions which themselves are time-periodic.

2. The flow $\Phi^t_{63}$ (21) has a purely continuous spectrum, Fig. 2 shows that the eigenfunctions of the approximation $W_\tau$ have no periodicity in time. The first and the third eigenfunctions appear to have a support which is localized in the phase space, and therefore in time since it is a continuous time flow.

3. Figure 3 checks the convergence of the spectrum of $W_\tau$, which is discrete, to the spectrum of $V$. A quantity which is a measure of the smoothness / oscillatory nature of the eigenfunctions is their *Dirichlet energy*, which is defined below.

$$E_{\text{Dir}}(f) = \| f \|_H^2 / \| f \|_\mu^2 - 1, \quad \forall f \in \mathcal{H}. \quad (23)$$

The Dirichlet energies of the $W_\tau$-eigenfunctions $\zeta_{\tau,j}$ are seen to converge for the $R^t_{\alpha_1, \alpha_2}$-flow. This seems to indicate that the $\zeta_{\tau,j}$ converge in $L^2(\mu)$ sense. The eigenvalues $\lambda_{\tau,j}$ which are continuous functions of $\tau$ are seen to converge for each $j$ as $\tau \to 0^+$, as stated in Theorem 2 (vi). For the Lorenz63 flow, for each $j$, the Dirichlet energy of $\zeta_{\tau,j}$ keeps increasing as $\tau \to 0^+$. This indicates the lack of any actual eigenfunctions to converge too.

4. Figures 4 and 5 show that the discrete spectrum of $W_\tau$, in combination with the functional calculus of compact skew-adjoint operators, can be used for prediction purposes. Corollary 3 and Theorem 15 suggest that the $L^2(\mu)$ error of the forecasts converge to 0 in the limit of $N \to \infty$, $L \to \infty$ and $\tau \to 0^+$. The forecast error for $R^t_{\alpha_1, \alpha_2}$ is seen to grow linearly with prediction time $t$. This lack of exponential growth of errors is supported by the fact that $R^t_{\alpha_1, \alpha_2}$ has zero Lyapunov exponents.

5. The forecast error grows much faster for $\Phi^t_{63}$. The graph shows that initially, there is an exponential growth in the error, as expected from the presence of a positive Lyapunov exponent for $\Phi^t_{63}$. 

29
Figure 4: 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 observable $\exp(F_1 + F_3)$ (right column) for the quasiperiodic torus rotation (19), using the operator $e^{iW\tau}$ with $\tau = 10^{-5}$. Top row: Comparison of the true and predicted signals as a function of time for a fixed initial condition. Bottom row: Normalized $L^2$ error as a function of lead time computed for an ensemble of forecasts initialized from 60,000 initial conditions sampled along an L63 trajectory independent from the training data.
Figure 5: Data-driven prediction of the L63 state vector components \( x_1 \), \( x_2 \), and \( x_3 \) using the operator \( e^{iWt} \) with \( \tau = 10^{-5} \). Top row: Comparison of the true and predicted signals as a function of time for a fixed initial condition. Bottom row: Normalized \( L^2 \) error as a function of lead time computed for an ensemble of forecasts initialized from 60,000 initial conditions sampled along an L63 trajectory independent from the training data.

6. Note that the limit in Corollary 3 holds for a fixed prediction time \( t \), so the value of the resolution parameter \( L \) which achieves a desired level of accuracy for a certain value of \( t \), may not do so for a different value of \( t \). The two flows that we investigated illustrates shows that this dependence of \( L \) on \( t \) is different for the two systems.

Summary. We have thus demonstrated that our methods are effective in approximating the functional calculus of the generator.

Acknowledgments. Dimitrios Giannakis received support from ONR YIP grant N00014-16-1-2649, NSF grant DMS-1521775, and DARPA grant HR0011-16-C-0116. Suddhasattwa Das is supported as a postdoctoral research fellow from the first grant.

References

[1] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal. 36 (1999) 491. doi:10.1137/S0036142996313002

[2] M. Dellnitz, G. Froyland, On the isolated spectrum of the Perron-Frobenius operator, Nonlinearity (2000) 1171–1188. doi:10.1088/0951-7715/13/4/310
[3] I. Mezić, A. Banaszuk, Comparison of systems with complex behavior, Phys. D. 197 (2004) 101–133. doi: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. doi:10.1007/s11071-005-2824-z

[5] G. Froland, K. Padberg, Almost-invariant sets and invariant manifolds – Connecting probabilistic and geometric descriptions of coherent structures in flows, Phys. D 238 (2009) 1507–1523. doi:10.1016/j.physd.2009.03.002

[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. doi:10.1017/s0022112009992059

[7] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28. doi:10.1017/jfm.2010.107

[8] G. Froyland, An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems, Phys. D 250 (2013) 1–19. doi: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. doi:10.3934/jcd.2014.1.249

[10] G. Froyland, G. A. Gottwald, A. Hamnerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, SIAM J. Appl. Dyn. Sys. 13 (4) (2014) 1816–1846. doi:10.1137/130943637

[11] J. H. Tu, C. W. Rowley, C. M. Lutchenburg, S. L. Brunton, J. N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dyn. 1 (2) (2014) 391–412. doi: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. doi:10.1103/PhysRevE.91.032915

[13] D. Giannakis, J. Slawinska, Z. Zhao, Spatiotemporal feature extraction with data-driven Koopman operators, J. Mach. Learn. Res. Proceedings 44 (2015) 105–115.

[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). doi:10.1007/s00332-015-9285-8

[15] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, J. N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dyn. 1 (2) (2014) 391–421. doi:10.3934/jcd.2014.1.391

[16] B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, Proc. Natl. Acad. Sci. 17 (5) (1931) 315–318. doi:10.1073/pnas.17.5.315

[17] T. Eisner, B. Farkas, M. Haase, R. Nagel, Operator Theoretic Aspects of Ergodic Theory, Vol. 272 of Graduate Texts in Mathematics, Springer, 2012.

[18] A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, Chaos 27 (2017) 053104. doi:10.1063/1.4972195

[19] M. Korda, M. Putinar, I. Mezić, Applied Koopmanism, Chaos 22 (2012) 047510. doi:10.1063/1.4772195

[20] A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, Chaos 27 (2017) 053104. doi:10.1063/1.4982720

[21] T. Eisner, B. Farkas, M. Haase, R. Nagel, Operator Theoretic Aspects of Ergodic Theory, Vol. 272 of Graduate Texts in Mathematics, Springer, 2015.

[22] B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, Proc. Natl. Acad. Sci. 17 (5) (1931) 315–318.

[23] M. H. Stone, On one-parameter unitary groups in Hilbert space, Ann. Math. 33 (1932) 643–648. doi:10.1007/s11071-005-2824-z

[24] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the koopman operator (2017). arXiv:1703.04680

[25] M. Korda, M. Putinar, I. Mezić, Data-driven spectral analysis of the Koopman operator (2017). arXiv:1710.06532

[26] S. Das, D. Giannakis, Koopman spectra in reproducing kernel Hilbert spaces (2018). arXiv:1801.07799

[27] S. Das, D. Giannakis, Data-driven spectral analysis of the Koopman operator (2017). arXiv:1710.06532

[28] U. von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering, Ann. Stat. 36 (2008) 555–586. doi:10.1214/009053607000000640

[29] S. Khas, P. Koltai, C. Schütte, On the numerical approximation of the PerronFrobenius and Koopman operator, J. Comput. Dyn. 3 (1) (2016) 51–79. doi:10.3934/jcd.2016003

[30] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, J. Nonlinear Sci. 28 (2) (2018) 687–710. doi:10.1007/s00332-017-9423-0

[31] M. H. Stone, On one-parameter unitary groups in Hilbert space, Ann. Math. 33 (1932) 643–648. doi:10.1007/s11071-005-2824-z

[32] 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. doi:10.1017/jfm.2018.297

[33] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the koopman operator (2017). arXiv:1710.06532

[34] T. Sauer, J. A. Yorke, M. Casdagli, Embedology, J. Stat. Phys. 65 (3–4) (1991) 579–616. doi:10.1016/0022-132X(91)90015-2

[35] J. C. Ferreira, V. A. Menegatto, Positive definiteness, reproducing kernel Hilbert spaces and beyond, Ann. Funct. Anal. 32
[38] R. R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, Appl. Comput. Harmon. Anal. 21 (2006) 31–52. doi:10.1016/j.acha.2005.07.005

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

[40] T. Berry, J. Harlim, Variable bandwidth diffusion kernels, J. Appl. Comput. Harmon. Anal. (2015). doi:10.1016/j.acha.2015.01.001

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

[42] I. Babuška, J. Osborn, Eigenvalue Problems, Vol. 2 of Handbook of Numerical Analysis, North Holland, Amsterdam, 1991.

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

[44] C. Micchelli, Y. Xu, H. Zhang, Universal kernels, J. Mach. Learn. Res. 7 (2006) 2651–2667. doi:10.1073/pnas.1517384113

[45] F. Cucker, S. Smale, On the mathematical foundations of learning, Bull. Amer. Math. Soc. 39 (1) (2001) 1–49. doi:10.1090/S0273-0979-01-00923-5

[46] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130–141.

[47] K. Law, A. Stuart, K. Zygalakis, Data Assimilation: A Mathematical Introduction, Vol. 62 of Texts in Applied Mathematics, Springer, 2015. doi:10.1007/978-3-319-20325-6

[48] W. Tucker, The Lorenz attractor exists, C. R. Acad. Sci. Paris, Ser. I 328 (1999) 1197–1202.

[49] P. Constantin, C. Foias, B. Nicolaenko, R. Témam, Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations, Springer, New York, 1989. doi:10.1007/978-1-4612-3506-4

[50] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 63 (1950) 337404. doi:10.1090/S0002-9947-1950-0051437-7

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

URL https://www.sciencedirect.com/science/article/pii/S1063520313000031

[52] E. Kowalski, Spectral theory in hilbert spaces, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.400.6646&rep=rep1&type=pdf.

URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.400.6646&rep=rep1&type=pdf

[53] C. R. de Oliveira, Intermediate spectral theory and quantum dynamics, Vol. 54, Springer Science & Business Media, 2008. URL https://books.google.com/books?id=FCYZ1Zf3uLQC

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

[55] N. Dunford, J. T. Schwartz, Linear operators, Vol. 2, Wiley, 1988.

[56] J. Stochel, F. Szafraniec, Domination of unbounded operators and commutativity, J. Math. Soc. Japan 55 (2) (2003) 405–437. doi:10.2969/jmsj/119149124

[57] G. Tian, Y. Ji, Y. Cao, Schauder bases and operator theory, arXiv preprint arXiv:1203.3603.

[58] R. Young, On complete biorthogonal systems, Proc. Amer. Math. Soc. 83 (3) (1981) 537–540. doi:10.1090/S0002-9939-1981-0637606-9

[59] 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 (1900) 441458. doi:10.1098/rsta.1900.0016

[60] 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. doi:10.1098/rsta.1968.0001

[61] S. Luzzatto, I. Melbourne, F. Paccaut, The Lorenz attractor is mixing, Comm. Math. Phys. 260 (2) (2005) 393–401.