Dynamic Mode Decomposition for Continuous Time Systems with the Liouville Operator

Joel A. Rosenfeld[1] · **Rushikesh Kamalapurkar[2]** · **L. Forest Gruss[3]** · **Taylor T. Johnson[4]**

Abstract

Dynamic mode decomposition (DMD) has become synonymous with the Koopman operator, where continuous time dynamics are discretized and examined using Koopman (i.e. composition) operators. Using the newly introduced "occupation kernels," the present manuscript develops an approach to DMD that treats continuous time dynamics directly through the Liouville operator. This manuscript outlines the technical and theoretical differences between Koopman-based DMD for discrete time systems and Liouville-based DMD for continuous time systems, which includes an examination of Koopman and Liouville operators over several reproducing kernel Hilbert spaces. While Liouville operators are modally unbounded, this manuscript introduces the concept of a scaled Liouville operator, which, for many dynamical systems, is a compact

✉ Joel A. Rosenfeld
rosenfeldj@usf.edu

Rushikesh Kamalapurkar
rushikesh.kamalapurkar@okstate.edu

L. Forest Gruss
forest.rosenfeld@gmail.com

Taylor T. Johnson
taylor.johnson@vanderbilt.edu

[1] Department of Mathematics and Statistics, University of South Florida, Tampa, USA

[2] School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, USA

[3] Department of Psychology, Eckerd College, St. Petersburg, USA

[4] Department of Electrical Engineering and Computer Science, Institute for Software Integrated Systems (ISIS), Nashville, USA

operator over the native space of the exponential dot product kernel. Compactness of scaled Liouville operators allows for norm convergence of Liouville-based DMD, which is a decided advantage over Koopman-based DMD.

1 Introduction

DMD has emerged as an effective method of extracting fundamental governing principles from high-dimensional time series data. The method has been employed successfully in the field of fluid dynamics, where DMD methods have demonstrated an ability to determine dynamic modes, also known as "Koopman modes," which agree with Proper Orthogonal Decomposition (POD) analyses (cf. Budišić et al. 2012; Črnjarić-Žic et al. 2020; Kutz et al. 2016; Mezić 2005, 2013; Williams et al. 2015a, b). However, DMD methods employing Koopman operators do not address continuous time dynamical systems directly. Instead, current DMD methods analyze discrete time proxies of continuous time systems (Kutz et al. 2016). The discretization process constrains Koopman-based DMD methods to systems that are forward complete (Bittracher et al. 2015). The objective of the present manuscript is to develop DMD methods that avoid discretization of continuous time dynamical systems, while providing convergence results that are stronger than Koopman-based DMD and applicable to a broader class of dynamical systems.

The connection between Koopman operators and DMD relies on the idea that a finite dimensional nonlinear dynamical system can be expressed as a linear operator over an infinite dimensional space. The linear representation enables treatment of the nonlinear system via tools from the theory of linear systems and linear operators. The idea of lifting finite dimensional nonlinear systems into infinite dimensional linear ones has been successfully utilized in the literature to achieve various identification and control objectives; however, a few fundamental limitations severely restrict the class of systems for which the connection between Koopman operators and DMD can be established via lifting to infinite dimensions. In particular, this article focuses on the following limitations.

Existence of Koopman Operators in Continuous Time Systems Consider the continuous time dynamical system given as $\dot{x} = 1 + x^2$. Discretization of this system with time step 1 yields the discrete dynamics $x_{i+1} = F(x_i) := \tan(1 + \arctan(x_i))$. It should be immediately apparent that F is not well defined over $\mathbb{R}$. In fact, through the consideration of $x_i = \tan(\pi/2 - 1)$ it can be seen that $F(x_i)$ is undefined. Since the symbol for a Koopman operator must be defined over the entire domain, there is no well-defined Koopman operator arising from this discretization. Note that the example above is not anecdotal. In addition to commonly used examples in classical works, such as Khalil (2002), mass-action kinetics in thermodynamics (Haddad 2019, Section 6.3), chemical reactions (Tóth et al. 2018, Section 8.4), and species populations (Hallam

and Levin 2012, Section 4.2) often give rise to such models. In general, unless the solutions of the continuous time dynamics are constrained to be forward complete, (for example, by assuming that the dynamical systems are globally Lipschitz continuous (Coddington and Levinson 1955, Chapter 1)) the resultant Koopman operator cannot be expected to be well-defined. This observation is validated by Bittracher et al. (2015), but otherwise conditions on the dynamics are largely absent from the literature.

Boundedness of Koopman Operators Even in the case of globally Lipschitz models, results regarding convergence of the DMD operator to the Koopman operator rely on the assumption that the Koopman operator is bounded over a specified RKHS (cf. Korda and Mezić 2018). Boundedness of composition operators, like the Koopman operator, has been an active area of study in the operator theory community. Indeed, it turns out there are very few bounded composition operators over many function spaces. A canonical example is in the study of the Bargmann-Fock space, where only affine symbols yield bounded composition operators and of those the compact operators arise from $F(z) = az + b$ where $|a| < 1$ (Carswell et al. 2003). A similar result holds for the native RKHS of the exponential dot product kernel and the native RKHS of the Gaussian radial basis function kernel (Gonzalez et al. 2021, Theorem 1). The implication of these results is that Koopman operators arising from the discretization of continuous time nonlinear systems cannot generally be expected to be bounded.

Practical Utility of Convergence Results In the DMD literature, convergence of the DMD operator to the Koopman operator is typically established in the strong operator topology (SOT). However, as noted in Korda and Mezić (2018), since SOT convergence is the topology of pointwise convergence (Pedersen 2012), it is not sufficient to justify use of the DMD operator to interpolate or extrapolate the system behavior from a collection of samples. Furthermore, by selecting a complete set of observables and adding them one at a time to a finite rank representation of the Koopman operator, pointwise convergence is to be expected. This is a restatement of the more general result that all bounded operators may be approximated by finite rank operators in SOT, which itself is a specialization of a much broader result for topological vector spaces (cf. Pedersen 2012, pg. 172). While Korda and Mezić (2018) also provides theoretically interesting insights into convergence of the eigenvalues and the eigenvectors of the DMD operator to eigenvalues and eigenfunctions of the Koopman operator along a subsequence, without the means to identify the convergent subsequences, practical utility of subsequential convergence is limited. In contrast, norm convergence is uniform convergence for operators, and yields a bound on the error over the kernels corresponding to the entire data set. Thus, a meaningful convergence result would arise from the norm convergence of finite rank representations to Koopman operators. However, this result is only possible for *compact* Koopman operators, which are virtually nonexistent in applications of interest.

A subset of Liouville operators, called Koopman generators, have been studied as limits of Koopman operators in works such as Cvitanovic et al. (2005), Das and Giannakis (2020), Froyland et al. (2014), Giannakis (2019), Giannakis and Das (2020) and Giannakis et al. (2018). Since Koopman generators are limits of Koopman operators, they also require the assumption of forward completeness on the dynamical system. This discussion brings into question the impact of various approaches to the study of

continuous time dynamical systems through discretization and Koopman operators, which all rely on the compactness, boundedness, or existence of Koopman operators.

The present work sidesteps the limiting process, and as a result, the assumptions regarding existence of Koopman operators, through the use of "occupation kernels". Specifically, occupation kernels remove the burden of approximation from that of operators and places it on the estimation of occupation kernels from time-series data, which requires much less theoretical overhead. Consequently, Liouville operators may be directly examined via occupation kernels, while avoiding limiting relations involving Koopman operators that might not be well defined for a particular discretization of a continuous time nonlinear dynamical system. As a result, the use of Liouville operators in a DMD routine allows for the study of dynamics that are locally rather than globally Lipschitz.

The action of the adjoint of a Liouville operator on an occupation kernel provides the input-output relationships that enable DMD of time series data. For the adjoint of a Liouville operator to be well defined, the operator must be densely defined over the underlying RKHS (Rosenfeld et al. 2019a, b). As a result, the exact class of dynamical systems that may be studied using Liouville operators depends on the selection of the RKHS. However, the requirement that the Liouville operator must be densely defined is not overly restrictive. For example, on the real valued Bargmann-Fock space, Liouville operators are densely defined for a wide range of dynamics that are expressible as real entire functions (which includes polynomial, exponential, sine, and cosine, etc.).

Perhaps the strongest case for Liouville operators is the fact that they can be "scaled" to generate compact operators. Section 3 of this paper introduces the idea of scaled Liouville operators as variants of Liouville operators that are compact for a large class of dynamical systems over the Bargmann-Fock space. Scaled Liouville operators make slight adjustments to the data by scaling the trajectories by a single parameter $|a| < 1$. Through the selection of a close to 1, scaled Liouville operators yield compact operators that are numerically indistinguishable from the corresponding unbounded Liouville operators over a given compact workspace. More importantly, the DMD procedure performed on scaled Liouville operators yields a sequence of finite rank operators that *converge in norm* to the scaled Liouville operators (see Theorem 2).

Practical Benefits of the Developed Method In addition to the theoretical benefits of Liouville operators detailed above, there are several practical benefits that arise from the use of occupation kernels and Liouville operators. Quadrature techniques, such as Simpson's rule, allow for the efficient estimation of occupation kernels while mitigating signal noise (Rosenfeld et al. 2019a, b), and also provide a robust estimation of the action of Liouville operators on occupation kernels. Furthermore, as snapshots are being integrated into trajectories for the generation of occupation kernels, the method presented in this manuscript can naturally incorporate irregularly sampled data.

In DMD, a large finite dimensional representation of the linear operator is constructed from data (i.e., *snapshots*) using a collection of *observables*. A subsystem of relatively small rank is then determined via a singular value decomposition (SVD) and approximation of the linear operator by the small rank subsystem is supported by a direct mapping between the eigenfunctions of the former and the eigenvectors of the latter (Williams et al. 2015b). The fact that the rank of the smaller subsystem

is typically in agreement with the number of snapshots, which can be considerably smaller than the number of observables, makes DMD particularly useful when there is a small number of snapshots of a high dimensional system. However, direct application of DMD to high dimensional systems sampled at high frequencies still poses a significant computational challenge, where many snapshots may have to be discarded to produce a computationally tractable problem, as was done in Kutz et al. (2016, Example 2.3). Such systems include mechanical systems with high sampling frequencies (Cichella et al. 2015; Walters et al. 2018), and neurobiological systems recorded via electroencephalography (EEG) where the typical sampling frequencies are of the order of 500 Hz (Gruss and Keil 2019). The methods in the present manuscript replace snapshots with integrals of trajectories of the system. The use of trajectories instead of individual snapshots reduces the dimensionality of the problem without discarding any data.

The developed algorithm also obviates the need for the truncated SVD that is utilized throughout DMD literature. For example, in Williams et al. (2015b) the truncated SVD is leveraged to convert from a feature space representation of the action of the Koopman operators to an approximation of the Gram matrix and an "interaction matrix." This stands in opposition of the spirit of the "kernel trick," where kernel functions are a means to avoid any direct interface with feature space. Following (Rosenfeld and Kamalapurkar 2021), the presented algorithm is given purely with respect to the occupation kernels, and the resultant methods are considerably simpler than what is seen in Williams et al. (2015b).

A Comparison with Similar Literature Liouville operators are studied in the context of DMD procedures using limiting definitions in works such as Klus et al. (2020). The manuscript (Klus et al. 2020), which was posted to arXiv around the same time as the first draft of this manuscript, approaches the Koopman generator through Galerkin methods. While the signs that the field is expanding beyond Koopman operators is encouraging, the authors of Klus et al. (2020) still adopt the limiting definitions of the Koopman generator in their work, which is an artifact from ergodic theory. Quantities similar to occupation kernels have been studied in the literature previously, in the form of occupation measures and time averaging functionals. Occupation kernels and occupation measures both represent the same functional over different spaces. Occupation measures are in the dual space of the Banach space of continuous functions, while occupation kernels are functions in a RKHS. As such, functions in the RKHS may be estimated through projections onto the span of occupation kernels, and this fact is leveraged in Sect. 4 where finite rank representations of the Liouville operators arise from the matrix representation of a projection operator. Occupation kernels are also distinct from time average functionals, where the latter is the average of a sum of iterated applications of the Koopman operator to an observable. As a result, in contrast with occupation kernels, whose definition is independent of Koopman and Liouville operators, time average functionals are only useful for the study of globally Lipschitz dynamics.

The relevant preliminary concepts for the theoretical underpinnings of the approach taken in the present manuscript are reviewed in Sect. 2.1. This includes definitions and properties of RKHSs as well as densely defined operators and their adjoints.

2 Technical Preliminaries

2.1 Reproducing Kernel Hilbert Spaces

Definition 1 A reproducing kernel Hilbert space (RKHS) over a set X is a Hilbert space of functions from X to $\mathbb{R}$ such that for each $x \in X$, the evaluation functional $E_x g := g(x)$ is bounded.

By the Riesz representation theorem, corresponding to each $x \in X$, there is a function $k_x \in H$, such that for all $g \in H$, $\langle g, k_x \rangle_H = g(x)$. The kernel function corresponding to H is given as $K(x, y) = \langle k_y, k_x \rangle_H$. The kernel function is a positive definite function in the sense that for any finite number of points $\{c_1, c_2, \ldots, c_M\} \subset X$, the corresponding Gram matrix $[K(c_i, c_j)]_{i,j=1}^M$ is positive semi-definite. The Gram matrix arises in many contexts in machine learning, such as in support vector machines (cf. Hastie et al. 2005). Particular to the subject matter of this manuscript, the Gram matrix plays a pivotal role in the construction of the kernel-based extended DMD method of Williams et al. (2015b) and the occupation kernel approach presented herein.

The Aronszajn-Moore theorem states that there is a unique correspondence between RKHSs and positive definite kernel functions (Aronszajn 1950). That is the RKHS may be constructed directly from the kernel function itself or the kernel function may be determined by a RKHS through the Riesz representation theorem. When the RKHS is obtained from the kernel function, it is frequently referred to as the native space of that kernel function (Wendland 2004).

RKHSs interact with function theoretic operators, such as Koopman (composition) operators (Jury 2007; Luery 2013; Williams et al. 2015b), multiplication operators (Rosenfeld 2015a, b), and Toeplitz operators (Rosenfeld 2016), in many nontrivial ways. For example, the kernel functions themselves play the role of eigenfunctions for the adjoints of multiplication operators Szafraniec (2000), and when the function corresponding to a Koopman operator has a fixed point at $c \in X$, the kernel function centered at that point (i.e. $K(\cdot, c) \in H$) is an eigenfunction for the adjoint of the Koopman operator (Cowen and MacCluer 1995). The kernel functions can also be demonstrated to be in the domain of the adjoint of densely defined Koopman operators as will be demonstrated in Sect. 2.2.

For machine learning applications kernel functions are frequently used for dimensionality reduction by expressing the inner product of data cast into a high-dimensional feature space as evaluation of the kernel function itself (Steinwart and Christmann 2008; Hastie et al. 2005). Specifically, a feature map corresponding to a RKHS is given as the mapping $x \mapsto \Psi(x) := (\Psi_1(x), \Psi_2(x), \ldots)^T \in \ell^2(\mathbb{N})$ for $x \in X$ such that $K(x, y) = \langle \Psi(y), \Psi(x) \rangle_{\ell^2(\mathbb{N})}$. That is, kernel function may be expressed as

$$K(x, y) = \sum_{m=1}^{\infty} \Psi_m(x) \Psi_m(y).$$

The feature space expression for a function $g \in H$ is given as $\mathbf{g} = (g_1, g_2, \ldots)^T \in \ell^2(\mathbb{N})$ so that $g(x) = \langle \mathbf{g}, \Psi(x) \rangle_{\ell^2(\mathbb{N})} = \langle g, K(\cdot, x) \rangle_H$. This representation of inner

products of vectors in a feature space as evaluation of a kernel function is central to the usage of kernel methods in data science, where the feature space is generally unknown but may be accessed through the kernel function. The approach taken in Williams et al. (2015b) uses the feature space as the fundamental basis for their representation and obtains kernel functions through a truncated SVD, whereas the present work avoids the invocation of the feature space and the truncated SVD.

The most frequently employed RKHS in machine learning applications is the native space of the Gaussian radial basis function kernel. The Gaussian radial basis function kernel is given as $K(x, y) = \exp\left(-\frac{1}{\mu}\|x - y\|_2^2\right)$, and it is a positive definite function over $\mathbb{R}^n$ for all n.

Another important kernel is the exponential dot product kernel, $K(x, y) = \exp\left(\frac{1}{\mu}x^T y\right)$, which is also a positive definite function over $\mathbb{R}^n$. What is significant concerning the exponential dot product kernel is that its native space is the Bargmann-Fock space, where bounded Koopman operators have been completely classified. Another significant feature, which will be leveraged in this manuscript, is that polynomials are dense inside the Bargmann-Fock space with respect to the Hilbert space norm.

2.2 Adjoints of Densely Defined Liouville Operators

In the study of operators, the theory concerning bounded operators is the most complete (cf. Pedersen 2012; Folland 2013). A bounded operator over a Hilbert space is a linear operator $W : H \rightarrow H$ such that $\|Wg\|_H \leq C\|g\|_H$ for some $C > 0$. The smallest C that satisfies $\|Wg\|_H \leq C\|g\|_H$ for all $g \in H$ is the norm of W and written as $\|W\|$. A classical theorem in operator theory states that the collection of bounded operators is precisely the collection of continuous operators over a Hilbert space (or more generally a Banach space) (Folland 2013, Chapter 5).

Unbounded operators over a Hilbert space are linear operators given as $W : \mathcal{D}(W) \rightarrow H$, where $\mathcal{D}(W)$ is the domain contained within H on which the operator W is defined (Pedersen 2012, Chapter 5). When the domain of W is dense in H, W is said to be a densely defined operator over H. While unbounded operators are by definition discontinuous, closed operators over a Hilbert space satisfy weaker limiting relations. That is, an operator is closed if, whenever $\{g_m\}_{m=1}^{\infty} \in \mathcal{D}(W)$, both $\{g_m\}_{m=1}^{\infty}$ and $\{Wg_m\}_{m=1}^{\infty}$ are convergent sequences, $g_m \rightarrow g \in H$, and $Wg_m \rightarrow h \in H$, we have that $g \in \mathcal{D}(W)$ and $Wg = h$ (Pedersen 2012, Chapter 5). The Closed Graph Theorem states that if W is a closed operator such that $\mathcal{D}(W) = H$, then W is bounded.

Lemma 1 *Given a RKHS, H, consisting of continuously differentiable functions, a Liouville Operator with symbol f, $A_f : \mathcal{D}(A_f) \rightarrow H$, is defined as $A_f g := \nabla g \cdot f$ where g resides in the canonical domain*

$$\mathcal{D}(A_f) := \{g \in H : \nabla g \cdot f \in H\}.$$

With this domain, A_f is closed over RKHSs that are composed of continuously differentiable functions.

Proof Liouville operators were demonstrated to be closed in Rosenfeld et al. (2019a).
                                                                                                                        □

The closedness of Koopman operators is well known in the study of RKHS, where they are more commonly known as composition operators (cf. Jury 2007; Luery 2013). Beyond the limit relations provided by closed operators, the closedness of an unbounded operator plays a significant role in the study of the adjoints of unbounded operators (Pedersen 2012, Chapter 5).

Definition 2 For a densely defined operator W, let

$$\mathcal{D}(W^*) := \{h \in H : g \mapsto \langle Wg, h \rangle_H \text{ is bounded on } \mathcal{D}(W)\}.$$

Since $\mathcal{D}(W)$ is dense in H, the functional $g \mapsto \langle Wg, h \rangle_H$ uniquely extends to H, and as such, for each $h \in \mathcal{D}(W^*)$ the Riesz representation theorem guarantees a function $W^*h \in H$ such that $\langle Wg, h \rangle_H = \langle g, W^*h \rangle_H$, for all $g \in \mathcal{D}(W)$. The adjoint of the operator W is thus given as $W^* : \mathcal{D}(W^*) \to H$ via the assignment $h \mapsto W^*h$.

Since the adjoint of a closed operator over a Hilbert space is densely defined (Pedersen 2012, Proposition 5.1.7), the adjoints of Liouville operators with domains given as in Lemma 1 are densely defined. Specific members of the domain of the respective adjoints may be identified, and these functions will be utilized in the characterization of the DMD methods in the subsequent sections. To characterize the interaction between the trajectories of a dynamical system and the Liouville operator, the notion of occupation kernels must be introduced (cf. Rosenfeld et al. 2019a).

Definition 3 Let X be a metric space, $\gamma : [0, T] \to X$ be an essentially bounded measurable trajectory, and let H be a RKHS over X consisting of continuous functions. Then the functional $g \mapsto \int_0^T g(\gamma(t))dt$ is bounded, and the Riesz representation theorem guarantees a function $\Gamma_\gamma \in H$ such that

$$\langle g, \Gamma_\gamma \rangle_H = \int_0^T g(\gamma(t))dt$$

for all $g \in H$. The function Γ_γ is the *occupation kernel* corresponding to γ in H.

Lemma 2 *If $f : \mathbb{R}^n \to \mathbb{R}^n$ is the dynamics for a dynamical system, and if $\gamma : [0, T] \to \mathbb{R}^n$ is a trajectory satisfying $\dot{\gamma} = f(\gamma(t))$ in the Carethèodory sense, then $\Gamma_\gamma \in \mathcal{D}(A_f^*)$ and $A_f^*\Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$.*

Proof This lemma was established in Rosenfeld et al. (2019a).   □

For Liouville operators, several examples can be demonstrated where particular symbols produce densely defined operators over the Bargmann-Fock space. In particular, since polynomials are dense in the Bargmann-Fock space, for polynomial dynamics, f, the function $A_f g = \nabla g \cdot f$ is a polynomial whenever g is a polynomial. Hence, polynomial dynamical systems correspond to densely defined Liouville

operators over the Bargmann-Fock space, and it should be noted that this is not a complete characterization of the densely defined Liouville operators over this space. For other RKHSs, different classes of dynamics correspond to densely defined operators, requiring independent evaluation for each RKHS.

3 A Compact Variation of the Liouville Operator

One of the drawbacks of employing either the Koopman operator or the Liouville operator for DMD is that the finite rank matrices produced by the method are strictly heuristic representations of the modally unbounded operators. An important question to address is whether a DMD procedure may be produced using a compact operator other than those densely defined operators discussed so far. This section presents a class of compact operators for use in DMD applied to continuous time systems. The compactness and boundedness of the operators will depend on the selection of the RKHS and the dynamics of the system. The Bargmann-Fock space will be utilized in this section, and the compactness assumption will be demonstrated to hold for a large class of dynamics.

Definition 4 Let H be a RKHS over $\mathbb{R}^n$, $a \in \mathbb{R}$ with $|a| < 1$, and let the scaled Liouville operator with symbol $f : \mathbb{R}^n \to \mathbb{R}^n$,

$$A_{f,a} : \mathcal{D}(A_{f,a}) \to H,$$

be given as $A_{f,a}g(x) = a\nabla g(ax)f(x)$ for all $x \in \mathbb{R}^n$ and

$$g \in \mathcal{D}(A_{f,a}) = \{h \in H : a\nabla h(ax)f(x) \in H\}.$$

From the definition of scaled Liouville operators, if $\gamma : [0, T] \to \mathbb{R}^n$ is a trajectory satisfying $\dot{\gamma} = f(\gamma)$, then

$$\int_0^T A_{f,a}g(\gamma(t))dt = \int_0^T a\nabla g(a\gamma(t))f(\gamma(t))dt = \langle A_{f,a}g, \Gamma_\gamma \rangle_H.$$

The following proposition then follows from arguments similar to the proof of Lemma 2.

Proposition 1 *For $\gamma : [0, T] \to \mathbb{R}^n$, such that $\dot{\gamma} = f(\gamma)$, $\Gamma_\gamma \in \mathcal{D}(A_{f,a}^*)$ and*

$$A_{f,a}^* \Gamma_\gamma = K(\cdot, a\gamma(T)) - K(\cdot, a\gamma(0)).$$

Theorem 1 and Corollary 1 demonstrate that for the Bargmann-Fock space, a large class of dynamics correspond to *compact* scaled Liouville operators.

Theorem 1 *Let $F^2(\mathbb{R}^n)$ be the Bargmann-Fock space of real valued functions, which is the native space for the exponential dot product kernel, $K(x, y) = \exp(x^T y)$, $a \in \mathbb{R}$ with $|a| < 1$, and let $A_{f,a}$ be the scaled Liouville operator with symbol $f : \mathbb{R}^n \to \mathbb{R}^n$.*

There exists a collection of coefficients, $\{C_\alpha\}_\alpha$, indexed by the multi-index α, such that if f is representable by a multi-variate power series, $f(x) = \sum_\alpha f_\alpha x^\alpha$, satisfying

$$\sum_\alpha |f_\alpha| C_\alpha < \infty,$$

then $A_{f,a}$ is bounded and compact over $F^2(\mathbb{R}^n)$.

Proof The proof has been relegated to the appendix to ease exposition. □

Corollary 1 *If f is a multi-variate polynomial, then $A_{f,a}$ is bounded and compact over $F^2(\mathbb{R}^n)$ for all $|a| < 1$.*

The compactness of scaled Liouville operators (over the Bargmann-Fock space) is critical for norm convergence of DMD methods. For bounded Koopman operators, results such as Korda and Mezić (2018) obtain convergence in the strong operator topology (SOT) of the DMD operator to the Koopman operator. SOT convergence is only pointwise convergence over a Hilbert space, and does not provide any generalization guarantees in the learning sense. Norm convergence on the other hand gives a uniform bound on the error estimates for all functions in the Hilbert space. Specifically, in this paper, the data-driven finite rank representation of the scaled Liouville operator, given in Sect. 4, is shown to converge, in norm, to the scaled Liouville operator.

While scaled Liouville operators are not identical to the Liouville operator, the selection of the parameter a close to 1 can be used to limit the difference between their finite rank representations to be within machine precision. Hence, the decomposition of scaled Liouville operators is computationally indistinguishable from that of the Liouville operator for a sufficiently close to 1.

4 Occupation Kernel Dynamic Mode Decomposition

4.1 Finite Rank Representation of the Liouville Operator

With the relevant theoretical background presented, this section develops the Occupation Kernel-based DMD method for continuous time systems. Let K be the kernel function for a RKHS, H, over $\mathbb{R}^n$ consisting of continuously differentiable functions. Let $\dot{x} = f(x)$ be a dynamical system corresponding to a densely defined Liouville operator, A_f, over H. Suppose that $\{\gamma_i : [0, T_i] \to X\}_{i=1}^M$ is a collection of trajectories satisfying $\dot{\gamma}_i = f(\gamma_i)$. There is a corresponding collection of *occupation kernels*, $\alpha := \{\Gamma_{\gamma_i}\}_{i=1}^M \subset H$, given as $\Gamma_{\gamma_i}(x) := \int_0^{T_i} K(x, \gamma_i(t))dt$. For each γ_i the action of A_f^* on the corresponding occupation kernel is given by $A_f^* \Gamma_{\gamma_i} = K(\cdot, \gamma_i(T_i)) - K(\cdot, \gamma_i(0))$.

Thus, when α is selected as an ordered basis for a vector space, the action of A_f^* is known on $\text{span}(\alpha)$. The objective of the DMD procedure is to express a matrix representation of the operator A_f^* on the finite dimensional vector space spanned by α followed by projection onto $\text{span}(\alpha)$.

Let $w_1, \cdots, w_M$ be the coefficients for the projection of a function $g \in H$ onto span(α) $\subset H$, written as $P_\alpha g = \sum_{i=1}^{M} w_i \Gamma_{\gamma_i}$. Using the fact that

$$\langle g, \Gamma_{\gamma_j} \rangle_H = \langle P_\alpha g, \Gamma_{\gamma_j} \rangle_H = \left(\langle \Gamma_{\gamma_1}, \Gamma_{\gamma_j} \rangle_H \cdots \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_j} \rangle_H \right) \begin{pmatrix} w_1 \\ \vdots \\ w_M \end{pmatrix},$$

for all $j = 1, \cdots, M$, the coefficients $w_1, \cdots, w_M$ may be obtained through the solution of the following linear system:

$$\begin{pmatrix} \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_M \end{pmatrix} = \begin{pmatrix} \langle g, \Gamma_{\gamma_1} \rangle_H \\ \vdots \\ \langle g, \Gamma_{\gamma_M} \rangle_H \end{pmatrix}, \tag{1}$$

where each of the inner products may be expressed as either single or double integrals as

$$\langle \Gamma_{\gamma_j}, \Gamma_{\gamma_i} \rangle_H = \int_0^{T_i} \int_0^{T_j} K(\gamma_i(\tau), \gamma_j(t)) dt d\tau, \text{ and } \langle g, \Gamma_{\gamma_i} \rangle_H = \int_0^{T_i} g(\gamma_i(t)) dt. \tag{2}$$

Furthermore, if $h = \sum_{i=1}^{M} v_i \Gamma_{\gamma_i} \in$ span(α) for some coefficients $\{v_i\}_{i=1}^{M} \subset \mathbb{R}$, then $A_f^* h \in H$, and it follows that

$$\langle A_f^* h, \Gamma_{\gamma_j} \rangle = \left\langle \sum_{i=1}^{M} v_i A_f^* \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \right\rangle_H$$

$$= \left(\left\langle A_f^* \Gamma_{\gamma_1}, \Gamma_{\gamma_j} \right\rangle_H, \cdots, \left\langle A_f^* \Gamma_{\gamma_M}, \Gamma_{\gamma_j} \right\rangle_H \right) \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}, \tag{3}$$

for all $j = 1, \cdots, M$. Using (1) and (3), the coefficients $\{w_i\}_{i=1}^{M}$ in the projection of $A_f^* h$ onto span(α) can be expressed as

$$\begin{pmatrix} w_1 \\ \vdots \\ w_M \end{pmatrix} = \begin{pmatrix} \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \langle A_f^* \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle A_f^* \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle A_f^* \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H & \cdots & \langle A_f^* \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}.$$

Lemma 2 then yields the finite rank representation for $P_\alpha A_f^*$, restricted to the occupation kernel basis, span(α), as

$$[P_\alpha A_f^*]_\alpha^\alpha = G^{-1}\mathcal{I}, \tag{4}$$

where

$$G := \begin{pmatrix} \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \tag{5}$$

is the Gram matrix of occupation kernels and

$$\mathcal{I} := \begin{pmatrix} \langle K(\cdot, \gamma_1(T_1)) - K(\cdot, \gamma_1(0)), \Gamma_{\gamma_1} \rangle_H & \cdots & \langle K(\cdot, \gamma_M(T_M)) - K(\cdot, \gamma_M(0)), \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle K(\cdot, \gamma_1(T_1)) - K(\cdot, \gamma_1(0)), \Gamma_{\gamma_M} \rangle_H & \cdots & \langle K(\cdot, \gamma_M(T_M)) - K(\cdot, \gamma_M(0)), \Gamma_{\gamma_M} \rangle_H \end{pmatrix}. \tag{6}$$

is the interaction matrix.

DMD requires a finite-rank representation of $P_\alpha A_f$, instead of $P_\alpha A_f^*$. Similar to the development above, Lemma 2 can be used to generate a finite rank representation of $P_\alpha A_f$ under the following additional assumption.

Assumption 1 The occupation kernels are in the domain of the Liouville operator, i.e., $\alpha \subset \mathcal{D}(A_f)$.

Given $h = \sum_{i=1}^M v_i \Gamma_{\gamma_i} \in$ span(α) for some coefficients $\{v_i\}_{i=1}^M \subset \mathbb{R}$, Assumption 1 implies that $A_f h \in H$ and

$$\langle A_f h, \Gamma_{\gamma_j} \rangle_H = \sum_{i=1}^M v_i \langle A_f \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_H = \sum_{i=1}^M v_i \langle \Gamma_{\gamma_i}, A_f^* \Gamma_{\gamma_j} \rangle_H$$

$$= \left(\langle \Gamma_{\gamma_1}, A_f^* \Gamma_{\gamma_j} \rangle_H, \ldots, \langle \Gamma_{\gamma_M}, A_f^* \Gamma_{\gamma_j} \rangle_H \right) \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}. \tag{7}$$

Lemma 2 then yields a finite rank representation of $P_\alpha A_f$, restricted to span(α) as

$$[P_\alpha A_f]_\alpha^\alpha = G^{-1}\mathcal{I}^T. \tag{8}$$

4.2 Dynamic Mode Decomposition

Suppose that λ_i is the eigenvalue for the eigenvector $v_i := (v_{i1}, v_{i2}, \ldots, v_{iM})^T$, $i = 1, \ldots, M$, of $[P_\alpha A_f]_\alpha^\alpha$. The eigenvector v_i can be used to construct a normalized eigenfunction of $P_\alpha A_f$ restricted to span(α), given as $\varphi_i = \frac{1}{N_i} \sum_{j=1}^M v_{ij} \Gamma_{\gamma_j}$, where

$N_i := \sqrt{v_i^\dagger G v_i}$, and $(\cdot)^\dagger$ denotes the conjugate transpose. Let V be the matrix of coefficients of the normalized eigenfunctions, arranged so that each column corresponds to an eigenfunction.

The DMD procedure begins by expressing the identity function, also known as the full state observable, $g_{id}(x) := x \in \mathbb{R}^n$ as a linear combination of the eigenfunctions of A_f, i.e., $g_{id}(x) = \lim_{M\to\infty} \sum_{i=1}^{M} \xi_{i,M}\varphi_i(x)$. For a fixed M, the identity function can be approximated using the *Liouville modes* $\xi_i \in \mathbb{R}^n$ as $g_{id}(x) \approx \sum_{i=1}^{M} \xi_i \varphi_i(x)$. The j-th row of the matrix $\xi = (\xi_1 \cdots \xi_M)$ is obtained as

$$
\left((\xi_1)_j \; \cdots \; (\xi_M)_j\right) = \left(\begin{pmatrix} \langle\varphi_1,\varphi_1\rangle_H & \cdots & \langle\varphi_1,\varphi_M\rangle_H \\ \vdots & \ddots & \vdots \\ \langle\varphi_M,\varphi_1\rangle_H & \cdots & \langle\varphi_M,\varphi_M\rangle_H \end{pmatrix}^{-1} \begin{pmatrix} \langle(x)_j,\varphi_1\rangle_H \\ \vdots \\ \langle(x)_j,\varphi_M\rangle_H \end{pmatrix}\right)^T,
$$

where $(x)_j$ is viewed as the functional mapping $x \in \mathbb{R}^n$ to its j-th coordinate. By examining the inner products $\langle g_{id}, \Gamma_{\gamma_i}\rangle_H$, for $i = 1, \ldots, M$, the matrix ξ may be expressed as

$$
\xi = \left(\int_0^{T_1} \gamma_1(t)dt \; \cdots \; \int_0^{T_M} \gamma_M(t)dt\right)\left(V^T G\right)^{-1} \tag{9}
$$

Given a trajectory $x(\cdot)$ satisfying $\dot{x} = f(x)$, each eigenfunction of A_f satisfies $\dot{\varphi}_i(x(t)) = \lambda_i \varphi_i(x(t))$ and hence, $\varphi_i(x(t)) = \varphi_i(x(0))e^{\lambda_i t}$, and the following data driven model is obtained:

$$
x(t) \approx \sum_{i=1}^{M} \xi_i \varphi_i(x(0))e^{\lambda_i t}, \tag{10}
$$

where

$$
\varphi_i(x(0)) = \frac{1}{N_i}\sum_{j=1}^{M} v_{ij}\Gamma_{\gamma_j}(x(0)) = \frac{1}{N_i}\sum_{j=1}^{M} v_{ij}\int_0^{T_j} K\left(x(0),\gamma_j(t)\right)dt. \tag{11}
$$

The resultant DMD procedure is summarized in Algorithm 1.

4.3 Modifications for the Scaled Liouville Operator DMD Method

Since Liouville operators are not generally compact, convergence, as $M \to \infty$, of the finite rank representation $P_\alpha A_f$ to the Liouville operator A_f cannot be guaranteed. Convergence of the finite rank representation can be established in the case of the scaled Liouville operators and the approximations obtained via DMD, under Assumption 1, are provably cogent. We call the approach taken here Scaled Liouville DMD (SL-DMD).

By Theorem 2, for an infinite collection of trajectories $\{\gamma_i\}_{i=1}^\infty$ with a dense collection of corresponding occupation kernels, $\{\Gamma_{\gamma_i}\}_{i=1}^\infty \subset H$, the resultant sequence of

Algorithm 1 Pseudocode for the dynamic mode decomposition routine of Section 4. Once the Liouville modes, the normalized eigenvectors, and the eigenvalues are returned, (10) and (11) can be used along with a numerical integration routine to reconstruct trajectories of the system starting from any given initial condition $x(0)$. The choice of numerical integration routine can have a significant impact on the overall results, and it is advised that a high accuracy method is leveraged in practice.

Require: Sampled trajectories $\{\gamma_j : [0, T_j] \to \mathbb{R}^n\}_{j=1}^M$
Require: Kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ of an RKHS
Require: A numerical integration routine
1: Compute the Gram matrix G in (5) using (2) and a numerical integration routine
2: Compute the interaction matrix $\mathcal{I}$ in (6) using (2) and a numerical integration routine
3: Compute eigenvalues, λ_i, and eigenvectors, v_i, of $G^{-1}\mathcal{I}^T$
4: Use (9) and a numerical integration routine to compute the matrix ξ of Liouville modes
5: **return** Liouville modes, ξ_i, normalized eigenvectors, $\frac{v_i}{N_i}$, and eigenvalues λ_i for $i = 1, \cdots, M$

---

finite rank operators $P_{\alpha_M} A_{f,a} P_{\alpha_M}$ converges to $A_{f,a}$, where $\alpha_M := \{\Gamma_{\gamma_1}, \ldots, \Gamma_{\gamma_M}\}$. Consequently, the spectrum of $[P_{\alpha_M} A_{f,a}]_{\alpha_M}^{\alpha_M}$, the finite rank representation of $P_{\alpha_M} A_{f,a}$, restricted to span(α), converges to that of $A_{f,a}$.[1]

Furthermore, when a is sufficiently close to 1 and the observed trajectories contained in a compact set are perturbed to within machine precision, the finite rank representations of $A_{f,a}$ and A_f are computationally indistinguishable.

DMD using scaled Liouville operators is similar to the unscaled case. In particular, recall that for $|a| < 1$ and f as above, $A_{f,a}^* \Gamma_{\gamma_i} = K(\cdot, a\gamma_i(T_i)) - K(\cdot, a\gamma_i(0))$. Hence, a finite rank representation of $A_{f,a}$, obtained from restricting and projecting to span(α), is given as

$$[P_\alpha A_{f,a}]_\alpha^\alpha = G^{-1}\mathcal{I}_a^T,$$

where

$$\mathcal{I}_a := \begin{pmatrix} \langle K(\cdot, a\gamma_1(T_1)) - K(\cdot, a\gamma_1(0)), \Gamma_{\gamma_1} \rangle_H & \cdots & \langle K(\cdot, a\gamma_M(T_M)) - K(\cdot, a\gamma_M(0)), \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle K(\cdot, a\gamma_1(T_1)) - K(\cdot, a\gamma_1(0)), \Gamma_{\gamma_M} \rangle_H & \cdots & \langle K(\cdot, a\gamma_M(T_M)) - K(\cdot, a\gamma_M(0)), \Gamma_{\gamma_M} \rangle_H \end{pmatrix}.$$

The approximate normalized eigenfunctions, $\{\varphi_{i,a}\}_{i=1}^M$, for $A_{f,a}$ may then be obtained in an identical fashion as for the Liouville operator.

Thus, the expression of the full state observable, g_{id}, in terms of the eigenfunctions yields $g_{id}(x) \approx \sum_{i=1}^M \xi_{i,a} \varphi_{i,a}(x)$ with (scaled) Liouville modes $\xi_{i,a}$.

As the eigenfunctions satisfy

$$\dot{\varphi}_{i,a}(ax(t)) = a\nabla\varphi_i(ax(t))f(x(t)) = A_{f,a}\varphi_{i,a}(x(t)) = \lambda_{i,a}\varphi_{i,a}(x(t)),$$

---

[1] It should be noted that the operator $P_{\alpha_M} A_{f,a} P_{\alpha_M}$ is simply $P_{\alpha_M} A_{f,a}$ when restricted to span(α_M) as $P_{\alpha_M} g = g$ for all $g \in$ span(α_M).

it can be seen that $\varphi_{i,a}(x(t)) \neq e^{t\lambda_{i,a}}\varphi_{i,a}(x(0))$. When a is close to 1, it can be demonstrated that $\varphi_{i,a}(x(t))$ is very nearly equal to $e^{t\lambda_{i,a}}\varphi_{i,a}(x(0))$, and the error can be controlled when $x(t)$ remains in a compact domain or workspace.

Proposition 2 *Let H be a RKHS of twice continuously differentiable functions over $\mathbb{R}^n$, f be Lipschitz continuous, and suppose that $\varphi_{i,a}$ is an eigenfunction of $A_{f,a}$ with eigenvalue $\lambda_{i,a}$. Let D be a compact subset of $\mathbb{R}^n$ that contains $x(t)$ for all $0 < t < T$. In this setting, if $\lambda_{i,a} \to \lambda_{i,1}$ and $\varphi_{i,a}(x(0)) \to \varphi_{i,1}(x(0))$ as $a \to 1^-$, then*

$$\sup_{0 \le t \le T} \|\varphi_{i,a}(x(t)) - e^{\lambda_{i,a}t}\varphi_{i,a}(x(0))\|_2 \to 0.$$

Proof The proof has been relegated to the appendix to ease exposition. □

Thus, under the hypothesis of Proposition 2, for a sufficiently close to 1, a data-driven model for a trajectory x satisfying $\dot{x} = f(x)$ is established as

$$x(t) \approx \sum_{i=1}^{M} \xi_{i,a}\varphi_{i,a}(x(0))e^{\lambda_{i,a}t}. \tag{12}$$

The principle advantage of using scaled Liouville operators is that these operators are compact over the Bargmann-Fock space for a large collection of nonlinear dynamics. Moreover, the sequence finite rank operators obtained through the DMD procedure achieves norm convergence when the sequence of recorded trajectories corresponds to a collection of occupation kernels that are dense in the Hilbert space.

Theorem 2 *Let $|a| < 1$. Suppose that $\{\gamma_i : [0, T_i] \to \mathbb{R}^n\}_{i=1}^{\infty}$ is a sequence of trajectories satisfying $\dot{\gamma} = f(\gamma)$ for a dynamical system f corresponding to a compact scaled Liouville operator, $A_{f,a}$. If the collection of functions, $\{\Gamma_{\gamma_i}\}_{i=1}^{\infty}$ is dense in the Bargmann-Fock space, then the sequence of operators $\{P_{\alpha_M}A_{f,a}P_{\alpha_M}\}_{M=1}^{\infty}$ converges to $A_{f,a}$ in the norm topology, where $\alpha_M = \{\Gamma_{\gamma_1}, \ldots, \Gamma_{\gamma_M}\}$.*

Proof The proof has been relegated to the appendix to ease exposition. □

5 Numerical Experiments

This section includes two collections of numerical experiments solved using the methods of the paper. The first surround the problem of flow across a cylinder, which has become a classic example for DMD. This provides a benchmark for comparison of the present method with kernel-based extended DMD. There it will be demonstrated that scaled Liouville modes and Liouville modes are very similar.

The second experiment performs a decomposition using electroencephalography (EEG) data, which has been sampled at 250 Hz over a period of 8 seconds. The high sampling frequency gives a large number of snapshots, which then leads to a high-dimensional learning problem when using the snapshots alone. The purpose of this experiment is to demonstrate how the Liouville operator based DMD can incorporate the large number of snapshots to generate Liouville modes without discarding data.

5.1 Flow Across a Cylinder

This experiment utilizes the data set from Kutz et al. (2016), which includes snapshots of flow velocity and flow vorticity generated from a computational fluid dynamics simulation. The data correspond to the wake behind a circular cylinder, and the Reynolds number for this flow is 100. The simulation was generated with time steps of $\Delta t = 0.02$ second and ultimately sampled every $10\Delta t$ seconds yielding 151 snapshots. Each snapshot of the system is a vector of dimension 89, 351. More details may be found in Kutz et al. (2016, Chapter 2).

Figure 1 presents the Liouville modes obtained from the cylinder vorticity data set where the collection 151 snapshots was subdivided into 147 trajectories, each of length 5. This figure should be compared with Fig. 2, which presents the scaled Liouville modes, with parameter $a = 0.99$, corresponding to the same data set. The modes were generated using the Gaussian kernel with $\mu = 500$. Figure 3 compares snapshots of the true vorticity against vorticity reconstructed using the unscaled and scaled Liouville DMD models in (10) and (12), respectively.

5.2 SsVEP Dataset

This experiment uses data from Gruss and Keil (2019). The data for this experiment was taken from an electroencephalography (EEG) recording of the visual cortex of one human participant during the active viewing of flickering images (Gruss and Keil 2019). By modulating luminance or contrast of an image at a constant rate (e.g. 12Hz), image flickering reliably evokes the steady state visually evoked potential (SsVEP) in early visual cortex (Regan 1989; Petro et al. 2017), reflecting entrainment of neuronal oscillations at the same driving frequency. SsVEP in the current data was evoked by pattern-reversal Gabor patch flickering at 12Hz (i.e. contrast-modulated) for a trial length of 7 seconds, with greatest signal strength originating from the occipital pole (Oz) of a 129-electrode cap. Data was sampled at 500Hz, band-pass filtered online from 0.5 - 48Hz, offline from 3 - 40Hz, with 53 trials retained for this individual after artifact rejection. Of these trials, the first 40 trials were used in the continuous time DMD method and each trial was subdivided into 50 trajectories. SsVEP data have the advantage of having an exceedingly high signal-to-noise ratio and high phase coherence due to the oscillatory nature of the signal, ideally suited for signal detection algorithms (such as brain-computer interfaces (Bakardjian et al. 2010; Bin et al. 2009; Middendorf et al. 2000)).

In this setting each independent trial can be used as a trajectory for a single occupation kernel. This differs from the implementation of Koopman-based DMD, where most often each snapshot corresponds to a single trajectory. The continuous time DMD method was performed using the Gaussian kernel function with $\mu = 50$.
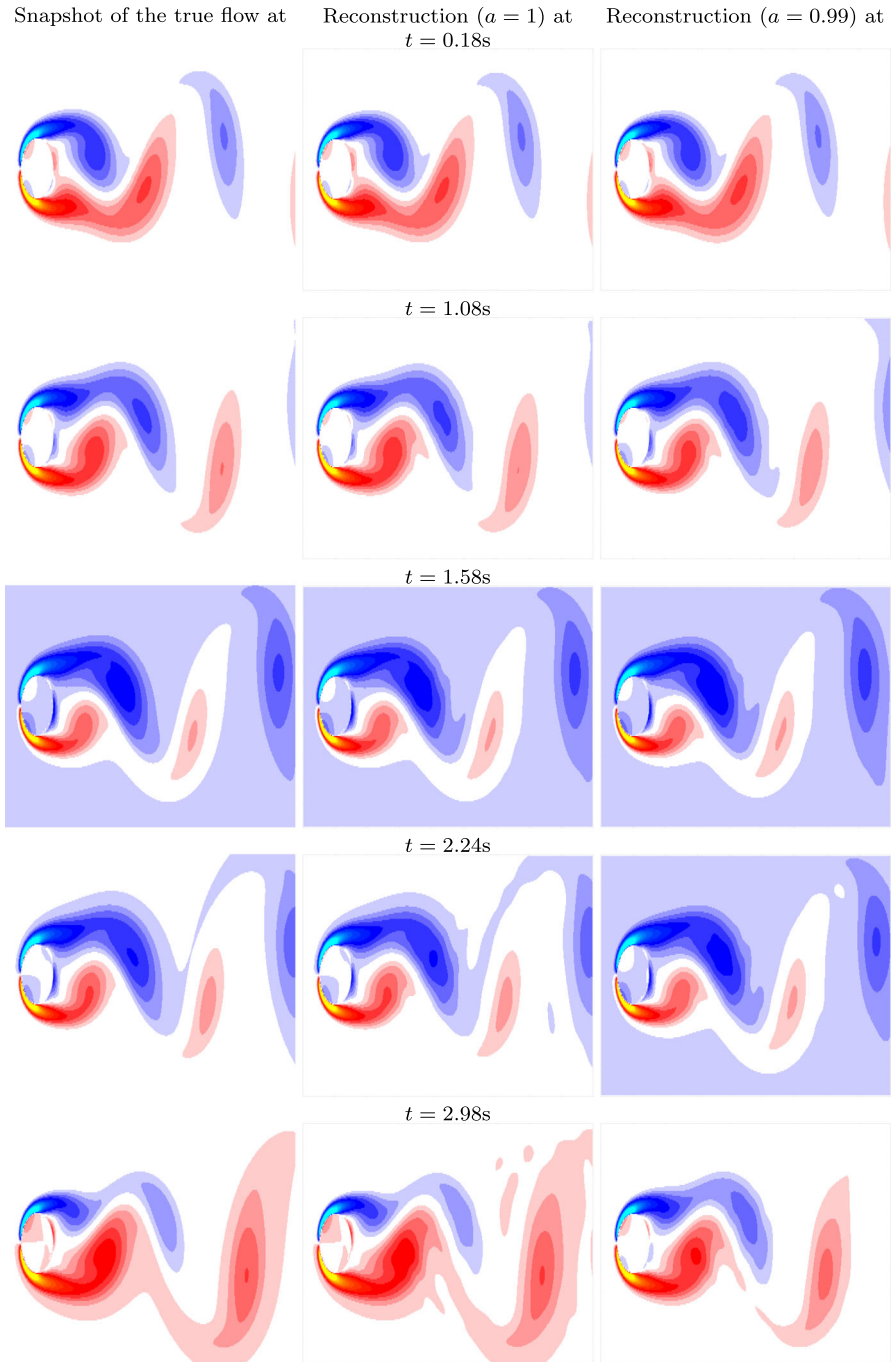
Figure 4 presents the obtained eigenvalues, and Fig. 5 gives log scaled spectrum obtained from the eigenvectors. It can be seen that the spectrum has strong peaks near the 12 Hz range, which suggests that the continuous time DMD procedure using occupation kernels can extract frequency information without using shifted copies of the trajectories as in Kutz et al. (2016).
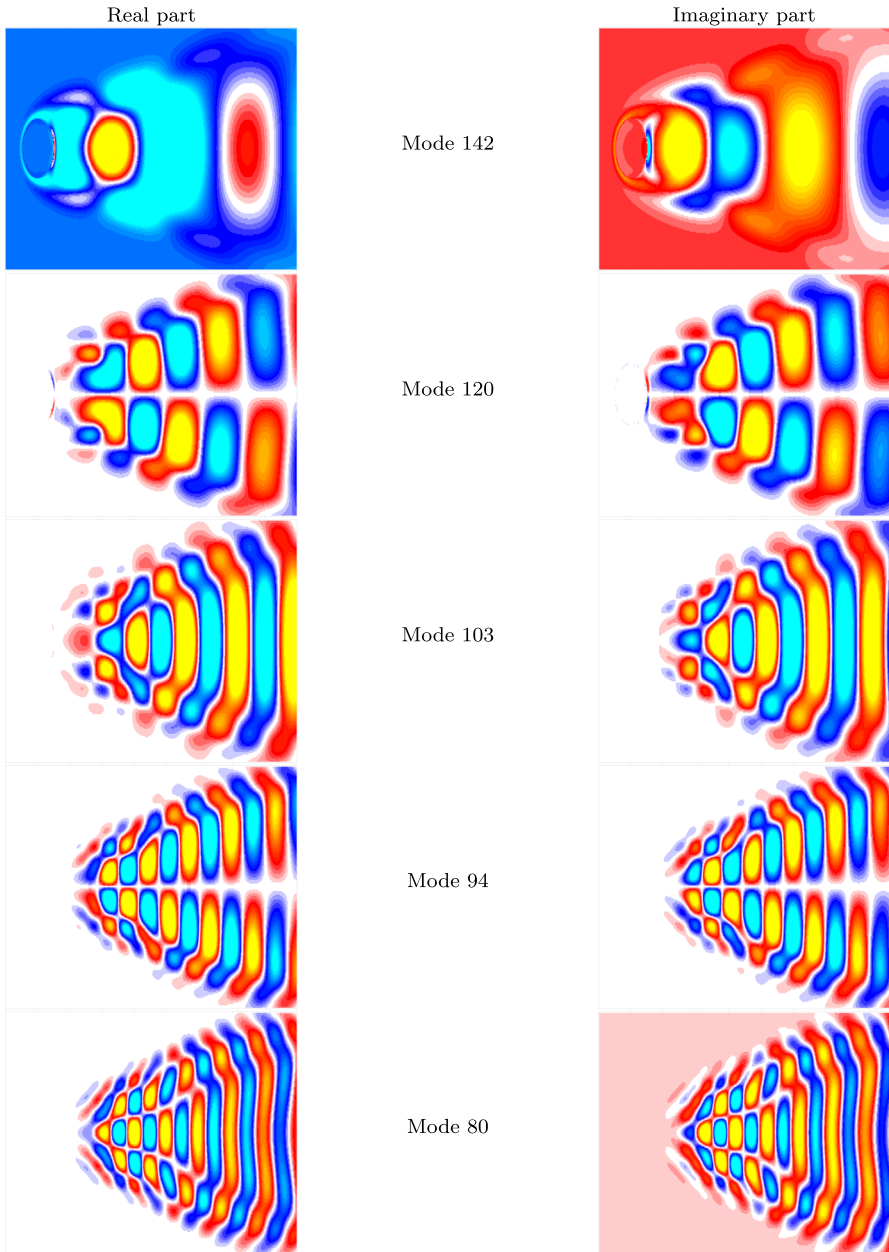
Fig. 1 This figure presents the real and imaginary parts of a selection of ten Liouville modes determined by the continuous time DMD method given in the present manuscript corresponding to the vorticity of a flow across a cylinder (data available in Kutz et al. 2016)
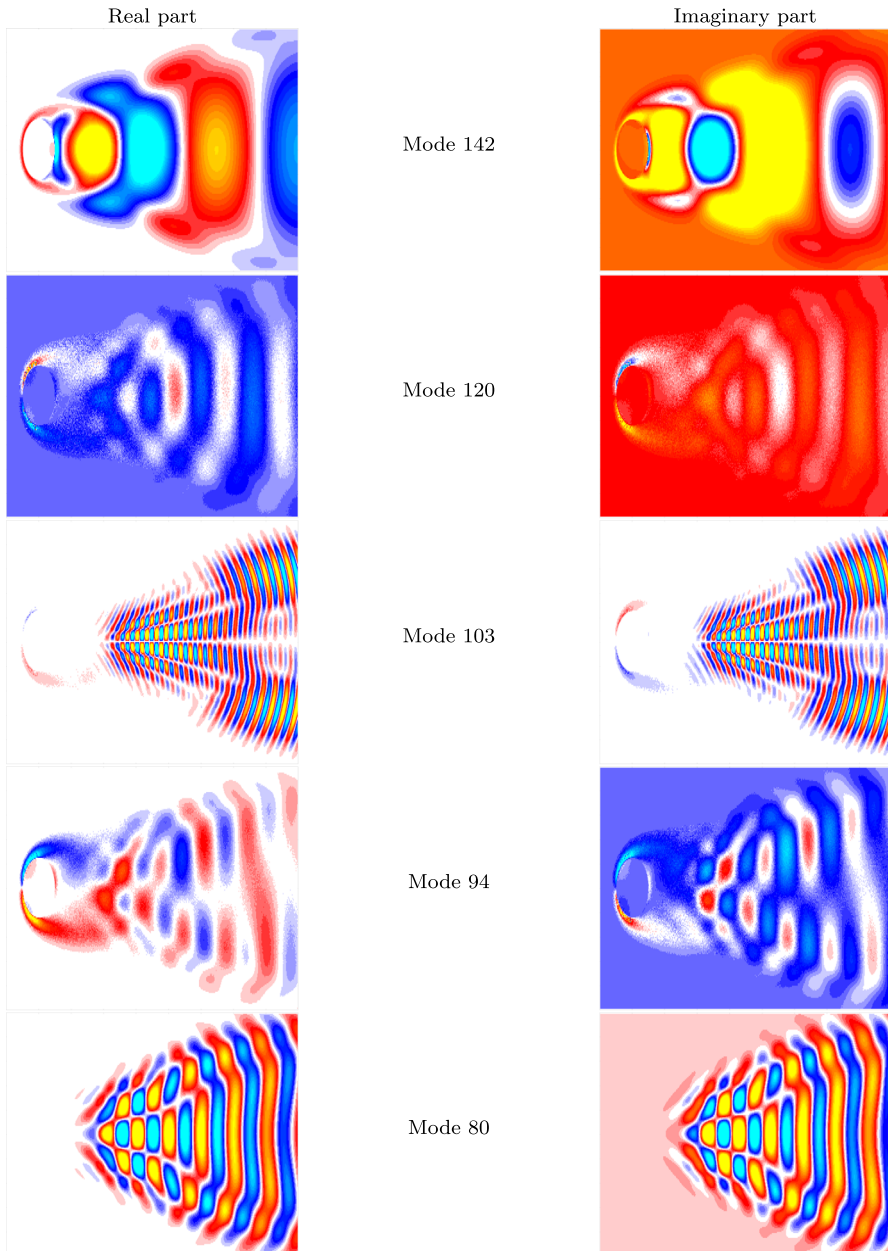
Fig. 2 This figure presents the real and imaginary parts of a selection of five scaled Liouville DMD modes for the cylinder wake vorticity data in Kutz et al. (2016). The difference between these modes and the modes in Fig. 1 was anticipated for several reasons; the selection of $a = 0.99$ is expected to result in slightly different modes, and there is no consistent method of ordering the Liouville modes as the significance of each mode depends not only on its magnitude, but also the associated eigenvector and initial value

Snapshot of the true flow at	Reconstruction ($a = 1$) at	Reconstruction ($a = 0.99$) at

$t = 0.18$s

$t = 1.08$s

$t = 1.58$s

$t = 2.24$s

$t = 2.98$s

Fig. 3 Snapshots of the true flow compared with reconstruction via the Liouville DMD model in (10) and the scaled Liouville DMD model in (12) with $a = 0.99$
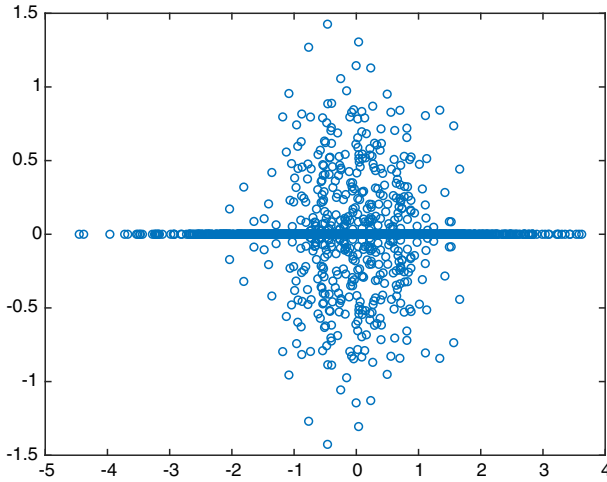
Fig. 4 Eigenvalues corresponding to the SsVEP dataset from Gruss and Keil (2019). This plot is on the complex plane, where the vertical axis indicates the imaginary part of the eigenvalue, and the horizontal axis indicates the real part
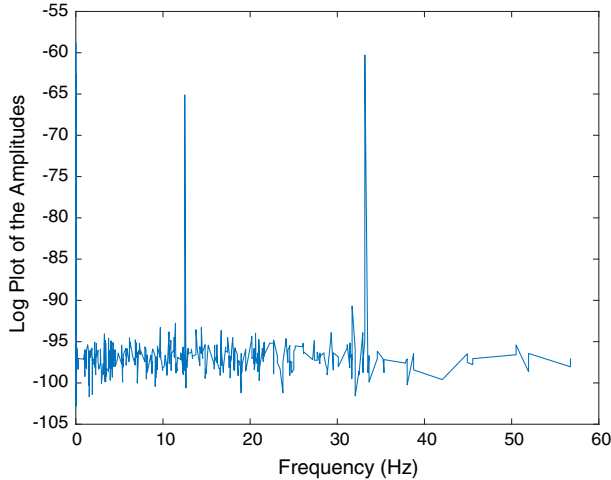
Fig. 5 Rescaled spectrum obtained from the SsVEP dataset. This doesn't quite correspond to the spectrum that would be computed through the Fourier transform. However, note the significant peak around 12 Hz, which corresponds to the SsVEP

For this example, the resultant dimensionality of Koopman-based DMD makes the analysis of this data set intractable without discarding a significant number of samples.

6 Discussion

6.1 Unboundedness of Liouville and Koopman Operators

Traditional DMD approaches aim to estimate a continuous nonlinear dynamical system by first selecting a fixed time-step and then investigate the induced discretized dynamics through the Koopman operator. The algorithm developed in this manuscript estimates the continuous nonlinear dynamics directly by employing occupation kernels, which represent trajectories via an integration functional that interfaces with the Liouville operator. That is, the principle advantage realized through DMD using Liouville operators and occupation kernels over that of kernel-based DMD and the Koopman operator is that the resulting finite-rank representation corresponds to a continuous time system rather than a discrete time proxy. This is significant, since not all continuous time systems can be discretized for use with the Koopman operator framework. Moreover, through employment of scaled Liouville operators, many dynamical systems yield a compact operator over the Bargmann-Fock space, which allows for norm convergence of DMD procedures.

Liouville operators are unbounded in most cases due to the inclusion of the gradient in their definition. Koopman operators are also unbounded in all but a few cases. In the specific instance where the selected kernel function is the exponential dot product kernel, Koopman operators are only bounded if the dynamics are affine (cf. Carswell et al. 2003). In contrast, large classes of both Liouville and Koopman operators are densely defined and closed operators over RKHSs. Thus, connections between DMD and Koopman/Liouville operators need to generally rely on the theory of unbounded operators.

6.2 Finite Rank Representations

Since Liouville operators are generally unbounded, convergence of the finite rank representation (in the norm topology) of the method in Sect. 4 cannot be established for most selections of f. Moreover, the selection of observables on which the operator is applied must come from the functions that reside in the domain of the Liouville operator, $\mathcal{D}(A_f)$. As bounded Koopman operators are rare as well, the need for care in the selection of observables is shared by both operators. In the design of the algorithm of this manuscript, an additional assumption was made where the domain of the Liouville operator was required to contain the occupation kernels corresponding to the observed trajectories. It should be noted that even if the occupation kernels are not in the domain of the Liouville operator, they are always in the domain of the *adjoint* of the Liouville operator, as long as the Liouville operator is closed and densely defined. As a result, an alternative DMD algorithm may be designed using the action of the adjoint on the occupation kernels. Interestingly, as evidenced by (4) and (8), the only adjustment to the algorithm in this setting is transposition of the matrix $\mathcal{I}$.

6.3 Approximating the Full State Observable

The decomposition of the full state observable relies strongly on selection of the RKHS. In the case of the Bargmann-Fock space, $x \mapsto (x)_i$ is a function in the space for each $i = 1, \ldots, n$. However, this is not the case for the native space of the Gaussian radial basis function kernel, which does not contain any polynomials. In both cases, the spaces are universal, which means that any continuous function may be arbitrarily well estimated by a function in the space with respect to the supremum norm over a compact subset. Thus, it is not expected that a good approximation of the full state observable will hold over all of $\mathbb{R}^n$, but a sufficiently small estimation error is possible over a compact workspace.

6.4 Scaled Liouville Operators

One advantage of the Liouville approach to DMD is that the Liouville operators may be readily modified to generate a compact operator through the so-called scaled Liouville operator. A large class of dynamics correspond to compact scaled Liouville operators, while Koopman operators cannot be modified in a similar fashion. Allowing this compact modification indicates that on an operator theoretic level, the study of nonlinear dynamical systems through Liouville operators allows for more flexibility.

The experiments presented in Sect. 5 demonstrate that the Liouville modes obtained with the continuous time DMD procedure using Liouville operators and occupation kernels are similar in form to the Koopman modes obtained using kernel-based extended DMD (Williams et al. 2015b). Moreover, occupation kernels allow for trajectories to be utilized as a fundamental unit of data, which can reduce the dimensionality of the learning problem while retaining some fidelity that would be otherwise lost through discarding data.

6.5 Time Varying Systems

The present framework can be adapted to handle time varying systems of the form $\dot{x} = f(t, x)$ with little adjustment. In particular, an analysis of this system may be achieved through state augmentation where time is included as a state variable as $z = [t, x^T]^T$, which leads to an adjusted dynamical system given as $\dot{z} = [1, f(t, x)^T]^T$. Hence, the analysis of time varying dynamics are included in the present approach.

6.6 Strong Operator Topology Convergence Versus Norm Convergence

One of the major contributions of this manuscript is the definition of the scaled Liouville operators, which for certain selections of a, these operators are compact over the exponential dot product kernel's space. This compactness enables the norm convergence of DMD routines, where the finite rank operators made for DMD are essentially operator level interpolants.

Presently, the best convergence results for DMD methods are SOT convergence results (Korda and Mezić 2018). SOT convergence yields pointwise convergence of operators in that a sequence of bounded operators, T_m, converges to T in SOT if and only if $T_m g \to T g$ for all $g \in H$. This mode of convergence is limited, and not entirely appropriate for spectral methods like DMD, where the only guarantees provided are that the spectrum of the limiting operator may be obtained as a subsequence of the members of the spectrum of the sequence of operators under consideration. Thus, as observed by the authors of Korda and Mezić (2018), infinitely many operators T_m, from the sequence of operators converging in SOT to T, may not be part of the subsequence for convergence, and as a result, may have dramatically different spectra from T. Moreover, the convergence result for Koopman-based DMD is a special case of a more general theorem, which implies that finite rank operators are dense in the collection of bounded operators with respect to the SOT (Pedersen 2012).

In contrast, norm convergence of operators is much stronger, where if two operators are close in norm, then their spectra are also close. Hence, convergence in norm of a sequence of bounded operators, T_m, to an operator, T guarantees the convergence of the spectra. Since DMD is a method where finite rank operators are designed to represent an unknown operator, the only operators amenable for norm convergence are compact operators. Compactness of the scaled Liouville operators thus allows for norm convergence of the finite rank approximations, generated for DMD, to the respective scaled Liouville operators. As a result, convergence of the estimated spectra to the spectra of the scaled Liouville operators is also established. More information concerning spectral theory and operator theory in general can be found in Pedersen (2012).

Scaled Liouville operators are compact over the native space of the exponential dot product kernel for a wide range of dynamical systems, including all polynomial dynamical systems. In contrast, every Koopman operator corresponding to a discretization of the trivial dynamics $\dot{x} = 0$ is the identity operator (for any selection of underlying function space), which is not compact when the underlying function space is an infinite dimensional Hilbert space.

7 Conclusions

In this paper, the notion of occupation kernels is leveraged to enable spectral analysis of the Liouville operator via DMD. A family of scaled Liouville operators is introduced and shown to be compact, which allows for norm convergence of the DMD procedure. Two examples are presented, one from fluid dynamics and another from EEG, which demonstrate reconstruction of trajectories, approximation of the spectrum, and a comparison of Liouville and scaled Liouville DMD.

The method presented here provides a new approach to DMD and builds the operator theoretic foundations for spectral decomposition of continuous time dynamical systems. By targeting the DMD procedure towards Liouville operators, which include Koopman generators as a proper subset, continuous time dynamical system are modeled directly, without discretization. Moreover, by obviating the limiting process used in the definition of Koopman generators, in favor of direct formulation via Liouville

operators, the requirement of forward completeness is relaxed and the resulting methods are applicable to a much broader class of dynamical systems.

A Proofs of Theorem 1 and Proposition 2

Theorem 1 restated: *Let $F^2(\mathbb{R}^n)$ be the Bargmann-Fock space of real valued functions, which is the native space for the exponential dot product kernel, $K(x, y) = \exp(x^T y)$, $a \in \mathbb{R}$ with $|a| < 1$, and let $A_{f,a}$ be the scaled Liouville operator with symbol $f : \mathbb{R}^n \to \mathbb{R}^n$. There exists a collection of coefficients, $\{C_\alpha\}_\alpha$, indexed by the multi-index α, such that if f is representable by a multi-variate power series, $f(x) = \sum_\alpha f_\alpha x^\alpha$, satisfying*

$$\sum_\alpha |f_\alpha| C_\alpha < \infty,$$

then $A_{f,a}$ is bounded and compact over $F^2(\mathbb{R}^n)$.

Proof The proof for the case $n = 1$ is presented to simplify the exposition. The case for $n > 1$ follows with some additional bookkeeping of the multi-index.

If $A_{x,a}$ is compact for all $|a| < 1$, then $A_{x^m,a} = A^m_{x, \sqrt[m]{a}}$ is compact since products of compact operators are compact. If $f(x) = \sum_{m=0}^\infty f_m x^m$ is such that $\sum_{m=0}^\infty |f_m| \|A_{x^m,a}\| < \infty$, then $A_{f,a} = \lim_{m \to \infty} \sum_{m=0}^M f_m A_{x^m,a}$, with respect the operator norm via the triangle inequality, and $A_{f,a}$ is compact since it is the limit of compact operators. Thus, it is sufficient to demonstrate that $A_{x,a}$ is compact to prove the theorem.

Let $g \in F^2(\mathbb{R})$, then $g(x) = \sum_{m=0}^\infty g_m \frac{x^m}{\sqrt{m!}}$ with norm $\|g\|^2_{F^2(\mathbb{R})} = \sum_{m=0}^\infty |g_m|^2 < \infty$. Applying the scaled Liouville operator, $A_{x,a}$, yields

$$A_{x,a} g(x) = axg'(ax) = \sum_{m=0}^\infty g_m a^m m \frac{x^m}{\sqrt{m!}}.$$

Hence, $\|A_{x,a} g\|^2_{F(\mathbb{R})} = |a|^{2m} m^2 |g_m|^2 < \infty$ as for large enough m, $|a|^{2m} m^2 < 1$. Hence, $A_{x,a}$ is everywhere defined and by the closed graph theorem $A_{x,a}$ is bounded.

As $|a|^m m^2 \to 0$, there is an M such that for all $m > M$, $|a|^m m^2 < 1$. Let P_M be the projection onto $\mathrm{span}\{1, x, x^2, \ldots, x^M\}$. Now consider

$$\|(A_{x,a} - A_{x,a} P_M)g\|^2 = \sum_{m=M+1}^\infty |g_m|^2 |a|^{2m} m^2$$

$$\leq \sum_{m=M+1}^{\infty} |g_m|^2 |a|^m$$

$$\leq |a|^M \sum_{m=M+1}^{\infty} |g_m|^2 |a|^{m-M}$$

$$\leq |a|^M \sum_{m=M+1}^{\infty} |g_m|^2 \leq |a|^M \|g\|^2_{F^2(\mathbb{R})}.$$

Hence, the operator norm of $(A_{x,a} - A_{x,a} P_M)$ is bounded by $|a|^{M/2}$, and as $|a| < 1$, $A_{x,a} P_m \to A_{x,a}$ in the operator norm. P_m is finite rank and therefore compact. It follows that $A_{x,a} P_m$ is compact, since compact operators form an ideal in the ring of bounded operators. Thus, $A_{x,a}$ is compact as it is the limit of compact operators. $\qquad\square$

Proposition 2 restated: *Let H be a RKHS of twice continuously differentiable functions over $\mathbb{R}^n$, f be Lipschitz continuous, and suppose that $\varphi_{i,a}$ is an eigenfunction of $A_{f,a}$ with eigenvalue $\lambda_{i,a}$. Let D be a compact subset of $\mathbb{R}^n$ that contains $x(t)$ for all $0 < t < T$. In this setting, if $\lambda_{i,a} \to \lambda_{i,1}$ and $\varphi_{i,a}(x(0)) \to \varphi_{i,1}(x(0))$ as $a \to 1^-$, then*

$$\sup_{0 \leq t \leq T} \|\varphi_{i,a}(x(t)) - e^{\lambda_{i,a}t} \varphi_{i,a}(x(0))\|_2 \to 0.$$

Proof Suppose that $x(t)$ remains in a compact set $D \subset \mathbb{R}^n$. Since $\phi_{m,a} \in H$ and H consists of twice continuously differentiable functions, there exists $M_1, M_2, F > 0$ such that

$$\sup_{x \in D} \|f(x)\| < F \quad \sup_{x \in D}, \|\nabla \phi_{m,a}(x)\| < M_{1,a}, \text{ and } \sup_{x \in D} \|\nabla^2 \phi_{m,a}(x)\| < M_{2,a}.$$

First, it is necessary to demonstrate that $M_{1,a}$ and $M_{2,a}$ may be bounded independent of a. For each $i, j = 1, \ldots, n$ and $y \in \mathbb{R}^n$, the functionals $g \mapsto \frac{\partial}{\partial x_i} g(y)$ and $g \mapsto \frac{\partial^2}{\partial x_i \partial x_j} g(y)$ are bounded (cf. Steinwart and Christmann 2008). Setting, $k_y = K(\cdot, y)$, it can be seen that the functions $\frac{\partial}{\partial x_i} k_y$ and $\frac{\partial^2}{\partial x_i \partial x_j} k_y$ are the unique functions that represent these functionals through the inner product of the RKHS (cf. Steinwart and Christmann 2008). As $\phi_{m,a}$ is a normal vector, $\|\phi_{m,a}\|_H = 1$, and by Cauchy-Schwarz

$$\|\nabla \phi_{m,a}(y)\|_2 = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial}{\partial x_i} \phi_{m,a}(y) \right)^2}$$

$$= \sqrt{\sum_{i=1}^{n} \left(\left\langle \phi_{m,a}, \frac{\partial}{\partial x_i} k_y \right\rangle_H \right)^2}$$

$$\leq \sqrt{\sum_{i=1}^{n} \left\| \phi_{m,a} \right\|_H^2 \left\| \frac{\partial}{\partial x_i} k_y \right\|_H^2}$$

$$= \sqrt{\sum_{i=1}^{n} \left\| \frac{\partial}{\partial x_i} k_y \right\|_H^2}. \tag{13}$$

(13) is bounded over D as $x \mapsto \frac{\partial}{\partial x_i} k_y(x)$ is continuous. Thus, $M_{1,a}$ is bounded independent of a. A similar argument may be carried out for $M_{2,a}$. Let M_1 and M_2 be the respective bounding constants.

Note that

$$\frac{\partial}{\partial t} \phi_{m,a}(ax(t)) = a\nabla\phi_{m,a}(ax(t)) f(x(t)) = A_{f,a}\phi_{m,a}(x(t)) = \mu_{m,a}\phi_{m,a}(x(t)).$$

Then by the mean value inequality, Cauchy-Schwarz, and the bounds given above,

$$\left| \frac{\partial}{\partial t} \phi_{m,a}(ax(t)) - \frac{\partial}{\partial t} \phi_{m,a}(x(t)) \right|$$
$$= \left| a\nabla\phi_{m,a}(ax(t)) f(x(t)) - \nabla\phi_{m,a}(x(t)) f(x(t)) \right|$$
$$\leq F \left\| a\nabla\phi_{m,a}(ax(t)) - a\nabla\phi_{m,a}(x(t)) + a\nabla\phi_{m,a}(x(t)) - \nabla\phi_{m,a}(x(t)) \right\|_2$$
$$\leq |a|F \left\| \nabla\phi_{m,a}(ax(t)) - \nabla\phi_{m,a}(x(t)) \right\|_2 + F|a-1|M_1 \|x(t)\|_2$$
$$\leq |a||a-1|M_2 F \|x(t)\|_2 + |a-1|M_1 F \|x(t)\|_2 = O(|a-1|).$$

Setting $\epsilon_a(t) := \frac{\partial}{\partial t} \phi_{m,a}(ax(t)) - \frac{\partial}{\partial t} \phi_{m,a}(x(t))$, it follows that $\sup_{0 \leq t \leq T} \|\epsilon_a(t)\|_2 = O(|a-1|)$. Hence,

$$\mu_{m,a}\phi_{m,a}(x(t)) = \frac{\partial}{\partial t} \phi_{m,a}(ax(t))$$
$$= \frac{\partial}{\partial t} \phi_{m,a}(x(t)) + \epsilon(t),$$

and

$$\phi_{m,a}(x(t)) = e^{\mu_{m,a}t}\phi_{m,a}(x(0)) - e^{\mu_{m,a}t} \int_0^t e^{-\mu_{m,a}\tau}\epsilon(\tau)d\tau.$$

As the time interval is fixed to $[0, T]$, $e^{\mu_{m,a}t} \int_0^t e^{-\mu_{m,a}\tau}\epsilon(\tau)d\tau = O(|a-1|)$, since $\mu_{m,a}$ is bounded with respect to a.   □

Theorem 2 restated: *Let $|a| < 1$. Suppose that $\{\gamma_i : [0, T_i] \to \mathbb{R}^n\}_{i=1}^{\infty}$ is a sequence of trajectories satisfying $\dot\gamma = f(\gamma)$ for a dynamical system f corresponding to a compact scaled Liouville operator, $A_{f,a}$. If the collection of functions, $\{\Gamma_{\gamma_i}\}_{i=1}^{\infty}$ is dense in the Bargmann-Fock space, then the sequence of operators $\{P_{\alpha_M} A_{f,a} P_{\alpha_M}\}_{M=1}^{\infty}$ converges to $A_{f,a}$ in the norm topology, where $\alpha_M = \{\Gamma_{\gamma_1}, \ldots, \Gamma_{\gamma_M}\}$.*

Proof The following proof is more general than what is indicated in the theorem statement of Theorem 2. In fact, for any compact operator, T, and any set $\{g_i\}_{i=1}^{\infty}$ such that $\overline{\text{span}(\{g_i\}_{i=1}^{\infty})} = H$, the sequence of operators $P_{\alpha_M} T P_{\alpha_M} \to T$ in norm, where P_{α_M} is the projection onto $\text{span}(\{g_i\}_{i=1}^{M})$. Henceforth, it will be assumed that $\{g_i\}_{i=1}^{\infty}$ is an orthonormal basis for H, since given any complete basis in H, an orthonormal basis may be obtained via the Gram-Schmidt process.

First note that every compact operator has a representation as $T = \sum_{i=1}^{\infty} \lambda_i \langle \cdot, v_i \rangle_H u_i$, where $\{v_i\}$ and $\{u_i\}$ are orthonormal collections of vectors (functions) in H, and $\{\lambda_i\}_{i=1}^{\infty} \subset \mathbb{C}$ are the singular values of T. If $T_M := \sum_{i=1}^{M} \lambda_i \langle \cdot, v_i \rangle_H u_i$ then $T_M \to T$ as $M \to \infty$ in the operator norm.

Suppose that $\epsilon > 0$. Select M such that $\|T - T_M\| < \epsilon$, and select N such that for all $n > N$,

$$\|u_i - P_n u_i\|_H < \frac{\epsilon}{\sum_{i=1}^{M} |\lambda_i|^2} \text{ and } \|v_i - P_n v_i\|_H < \frac{\epsilon}{\sqrt{M} \left(\sum_{i=1}^{M} |\lambda_i|^2 \right)^{1/2}}$$

for all $i = 1, \ldots, M$. Let $g \in H$ be arbitrary.

Now consider,

$$\|Tg - P_n T P_n g\|_H = \|Tg - T_M g + T_M g - P_n T P_n f\|_H$$
$$\leq \|T - T_M\| \|g\|_H + \|T_M g - P_n T P_n g\|_H \leq \epsilon \|g\|_H + \|T_M g - P_n T P_n g\|_H.$$

The second term after the inequality may be expanded as

$$\|T_M g - P_n T P_n g\|_H \leq \|T_M g - P_n T_M g\|_H + \|P_n T_M g - P_n T P_n g\|_H$$
$$\|T_M g - P_n T_M g\|_H + \|T_M g - T P_n g\|_H$$
$$\leq \|T_M g - P_n T_M g\|_H + \|T_M g - T_M P_n g\|_H + \|T_M P_n g - T P_n g\|_H$$
$$\leq \|T_M g - P_n T_M g\|_H + \|T_M g - T_M P_n g\|_H + \epsilon \|g\|_H.$$

Now the objective is to demonstrate that both $\|T_M g - P_n T_M g\|_H$ and $\|T_M g - T_M P_n g\|_H$ are proportional to $\epsilon \|g\|_H$. Note that

$$\|T_M g - P_n T_M g\|_H = \left\| \sum_{i=1}^{M} \lambda_i \langle g, v_i \rangle_H (u_i - P_n u_i) \right\|_H$$

$$\leq \sum_{i=1}^{M} |\lambda_i| |\langle g, v_i \rangle_H| \|u_i - P_n u_i\|_H$$

$$\leq \sqrt{\sum_{i=1}^{M} |\langle g, v_i \rangle_H|^2} \left(\sum_{i=1}^{M} |\lambda_i|^2 \|u_i - P_n u_i\|_H^2 \right)^{1/2}$$

$$\leq \|g\|_H \left(\sum_{i=1}^{M} |\lambda_i|^2 \|u_i - P_n u_i\|_H^2 \right)^{1/2} \leq \epsilon \|g\|_H,$$

and

$$\|T_M(g - P_n g)\|_H = \left\| \sum_{i=1}^{M} \lambda_i \langle g - P_n g, v_i \rangle_H u_i \right\|$$

$$\leq \left\| \sum_{i=1}^{M} \lambda_i \left\langle \sum_{j=n+1}^{\infty} \langle g, g_j \rangle_H g_j, v_i \right\rangle_H u_i \right\|$$

$$\leq \sum_{j=n+1}^{\infty} |\langle g, g_j \rangle_H| \left(\sum_{i=1}^{M} |\lambda_i| |\langle g_j, v_i \rangle_H| \right)$$

$$\leq \left(\sum_{j=n+1}^{\infty} |\langle g, g_i \rangle_H|^2 \right)^{1/2} \left(\sum_{j=n+1}^{\infty} \left(\sum_{i=1}^{M} |\lambda_i| |\langle g_i, v_i \rangle_H| \right)^2 \right)^{1/2}$$

$$\leq \|g\|_H \left(\sum_{j=n+1}^{\infty} \left(\sum_{i=1}^{M} |\lambda_i|^2 \right) \left(\sum_{i=1}^{M} |\langle g_j, v_i \rangle_H|^2 \right) \right)^{1/2}$$

$$\leq \|g\|_H \left(\sum_{i=1}^{M} |\lambda_i|^2 \right)^{1/2} \left(\sum_{i=1}^{M} \sum_{j=n+1}^{\infty} |\langle g, v_i \rangle_H|^2 \right)^{1/2}$$

$$= \|g\|_H \left(\sum_{i=1}^{M} |\lambda_i|^2 \right)^{1/2} \left(\sum_{i=1}^{M} \|v_i - P_n v_i\|^2 \right)^{1/2} \leq \epsilon \|g\|_H.$$

Thus, for every $\epsilon > 0$, there is an N such that for all $n > N$, $\|Tg - P_n T P_n g\|_H \leq 4\epsilon \|g\|_H$. Hence, it follows that $\|T - P_n T P_n\| \leq 4\epsilon$. Thus, as $n \to \infty$, $P_n T P_n \to T$ in the operator norm. $\square$

References

Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**(3), 337–404 (1950)

Bakardjian, H., Tanaka, T., Cichocki, A.: Optimization of SSVEP brain responses with application to eight-command brain-computer interface. Neurosci. Lett. **469**(1), 34–38 (2010)

Bin, G., Gao, X., Yan, Z., Hong, B., Gao, S.: An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. J. Neural Eng. **6**(4), 046002 (2009)

Bittracher, A., Koltai, P., Junge, O.: Pseudogenerators of spatial transfer operators. SIAM J. Appl. Dyn. Syst. **14**(3), 1478–1517 (2015)

Budišić, M., Mohr, R., Mezić, I.: Applied Koopmanism. Chaos Interdiscip. J. Nonlinear Sci. **22**(4), 047510 (2012)

Carswell, B., MacCluer, B.D., Schuster, A.: Composition operators on the Fock space. Acta Sci. Math. (Szeged) **69**(3–4), 871–887 (2003)

Cichella, V., Kaminer, I., Dobrokhodov, V., Xargay, E., Choe, R., Hovakimyan, N., Aguiar, A.P., Pascoal, A.M.: Cooperative path following of multiple multirotors over time-varying networks. IEEE Trans. Autom. Sci. Eng. **12**(3), 945–957 (2015)

Coddington, E.A., Levinson, N.: Theory of Ordinary Differential Equations. Tata McGraw-Hill Education, New York (1955)

Cowen, C.C., Jr., MacCluer, B.I.: Composition Operators on Spaces of Analytic Functions, vol. 20. CRC Press, Boca Raton (1995)

Črnjarić-Žic, N., Maćešić, S., Mezić, I.: Koopman operator spectrum for random dynamical systems. J. Nonlinear Sci. **30**, 2007–2056 (2020)

Cvitanovic, P., Artuso, R., Mainieri, R., Tanner, G., Vattay, G., Whelan, N., Wirzba, A.: Chaos: Classical and Quantum. ChaosBook. org. Niels Bohr Institute, Copenhagen (2005)

Das, S., Giannakis, D.: Koopman spectra in reproducing kernel Hilbert spaces. Appl. Comput. Harm. Anal. **49**(2), 573–607 (2020)

Folland, G.B.: Real Analysis: Modern Techniques and Their Applications. Wiley, Hoboken (2013)

Froyland, G., González-Tokman, C., Quas, A.: Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. J. Comput. Dyn. **1**(2), 249–278 (2014)

Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. Appl. Comput. Harm. Anal. **47**(2), 338–396 (2019)

Giannakis, D., Das, S.: Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis. Physica D **402**, 132211 (2020)

Giannakis, D., Kolchinskaya, A., Krasnov, D., Schumacher, J.: Koopman analysis of the long-term evolution in a turbulent convection cell. arXiv:1804.01944 (2018)

Gonzalez, E., Abudia, M., Jury, M., Kamalapurkar, R., Rosenfeld, J.A.: Anti-koopmanism. arXiv:2106.00106v2 (2021)

Gruss, L.F., Keil, A.: Sympathetic responding to unconditioned stimuli predicts subsequent threat expectancy, orienting, and visuocortical bias in human aversive Pavlovian conditioning. Biol. Psychol. **140**, 64–74 (2019)

Haddad, W.: A Dynamical Systems Theory of Thermodynamics. Princeton Series in Applied Mathematics. Princeton University Press, Princeton (2019)

Hallam, T.G., Levin, S.A.: Mathematical Ecology: An Introduction, vol. 17. Springer Science & Business Media, Berlin (2012)

Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. Math. Intell. **27**(2), 83–85 (2005)

Jury, M.T.: C*-algebras generated by groups of composition operators. Indiana Univ. Math. J. **56**(6), 3171–3192 (2007)

Khalil, H.K.: Nonlinear Systems, 3rd edn. Prentice Hall, Upper Saddle River (2002)

Klus, S., Nüske, F., Peitz, S., Niemann, J.H., Clementi, C., Schütte, C.: Data-driven approximation of the Koopman generator: model reduction, system identification, and control. Physica D **406**, 132416 (2020)

Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. J. Nonlinear Sci. **28**(2), 687–710 (2018)

Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. SIAM, Philadelphia (2016)

Luery, K.E.: Composition Operators on Hardy Spaces of the Disk and Half-Plane. University of Florida, Gainesville (2013)

Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**(1–3), 309–325 (2005)

Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. Annu. Rev. Fluid Mech. **45**, 357–378 (2013)

Middendorf, M., McMillan, G., Calhoun, G., Jones, K.S.: Brain-computer interfaces based on the steady-state visual-evoked response. IEEE Trans. Rehabil. Eng. **8**(2), 211–214 (2000)

Pedersen, G.K.: Analysis Now, Graduate Texts in Mathematics, vol. 118. Springer Science & Business Media, Berlin (2012)

Petro, N.M., Gruss, L.F., Yin, S., Huang, H., Miskovic, V., Ding, M., Keil, A.: Multimodal imaging evidence for a frontoparietal modulation of visual cortex during the selective processing of conditioned threat. J. Cogn. Neurosci. **29**(6), 953–967 (2017)

Regan, D.: Human brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. Elsevier, Amsterdam (1989)

Rosenfeld, J.A.: Densely defined multiplication on several Sobolev spaces of a single variable. Complex Anal. Oper. Theory **9**(6), 1303–1309 (2015a)

Rosenfeld, J.A.: Introducing the polylogarithmic hardy space. Integral Equ. Oper. Theory **83**(4), 589–600 (2015b)

Rosenfeld, J.A.: The Sarason sub-symbol and the recovery of the symbol of densely defined Toeplitz operators over the Hardy space. J. Math. Anal. Appl. **440**(2), 911–921 (2016)

Rosenfeld, J.A., Kamalapurkar, R.: Dynamic mode decomposition with control Liouville operators. In: IFAC-PapersOnLine, vol. 54, pp. 707–712 (2021)

Rosenfeld, J.A., Kamalapurkar, R., Gruss, L.F., Johnson, T.T.: On occupation kernels, Liouville operators, and dynamic mode decomposition. In: Proceedings of the American Control Conference, pp. 3957–3962. New Orleans, LA, USA (2021)

Rosenfeld, J.A., Kamalapurkar, R., Russo, B., Johnson, T.T.: Occupation kernels and densely defined Liouville operators for system identification. In: Szafraniec are Ramirez de Arellano, E. and Shapiro, M. V. and Tovar, L. M. and Vasilevski N. L. Proceedings of the IEEE Conference on Decision and Control, pp. 6455–6460 (2019a)

Rosenfeld, J.A., Russo, B., Kamalapurkar, R., Johnson, T.T.: The occupation kernel method for nonlinear system identification. arXiv:1909.11792 (2019b)

Steinwart, I., Christmann, A.: Support Vector Machines. Springer Science & Business Media, Berlin (2008)

Szafraniec, F.H.: The reproducing kernel Hilbert space and its multiplication operators. In: Ramirez de Arellano, E., Shapiro, M. V., Tovar, L. M., Vasilevski N. L. (eds.) Complex Analysis and Related Topics, pp. 253–263. Springer (2000)

Tóth, J., Nagy, A.L., Papp, D.: Reaction Kinetics: Exercises, Programs and Theorems. Springer, Berlin (2018)

Walters, P., Kamalapurkar, R., Voight, F., Schwartz, E.M., Dixon, W.E.: Online approximate optimal station keeping of a marine craft in the presence of an irrotational current. IEEE Trans. Robot. **34**(2), 486–496 (2018)

Wendland, H.: Scattered Data Approximation, Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge (2004)

Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. **25**(6), 1307–1346 (2015a)

Williams, M.O., Rowley, C.W., Kevrekidis, I.G.: A kernel-based method for data-driven Koopman spectral analysis. J. Comput. Dyn. **2**(2), 247–265 (2015b)