

# Dynamic Mode Decomposition for Continuous Time Systems with the Liouville Operator

Joel A. Rosenfeld · Rushikesh Kamalapurkar ·  
L. Forest Gruss · Taylor T. Johnson

Received: date / Accepted: date

**Abstract** Dynamic Mode Decomposition (DMD) has become synonymous with the Koopman operator, where continuous time dynamics are examined through a discrete time proxy determined by a fixed timestep 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, which can include Koopman generators. 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 these operators over several reproducing kernel Hilbert spaces (RKHSs). While Liouville operators are modally unbounded, this manuscript introduces the concept of a scaled Liouville operator, which for many dynamical systems yields a compact operator over the exponential dot product kernel’s native space. Hence, norm convergence of the DMD procedure is established when using scaled Liouville operators, which is a decided advantage over Koopman based DMD methods.

---

A subset of the results in this manuscript was presented at the 2021 American Control Conference and will be published in the proceedings [38]. A YouTube Playlist supporting the content of this manuscript (including MATLAB code) may be found here: [https://youtube.com/playlist?list=PLldiDnQu2phuIdps0DcIQJ\\_gF0YIb-g6y](https://youtube.com/playlist?list=PLldiDnQu2phuIdps0DcIQJ_gF0YIb-g6y)

---

Joel A. Rosenfeld  
Department of Mathematics and Statistics  
University of South Florida E-mail: rosenfeldj@usf.edu

Rushikesh Kamalapurkar  
Department of Mechanical and Aerospace Engineering  
Oklahoma State University E-mail: rushikesh.kamalapurkar@okstate.edu

L. Forest Gruss  
College of Nursing  
University of South Florida E-mail: lgrussrosenfeld@usf.edu

Taylor T. Johnson  
Institute for Software Integrated Systems (ISIS)  
Department of Electrical Engineering and Computer Science  
E-mail: taylor.johnson@vanderbilt.edu

**Keywords** Dynamic Mode Decomposition · Densely Defined Operators · Liouville Operator · Reproducing Kernel Hilbert Spaces

## 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. [5, 10, 26, 28, 29, 46, 47]). 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 [26]. The discretization process constrains Koopman based DMD methods to systems that are forward complete [4]. 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 to enable 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.

**Existence of Koopman operators in continuous time systems:** Consider the continuous time dynamical system given as  $\dot{x} = 1 + x^2$ . The discretization with time step 1 yields the following 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 [23], mass-action kinetics in thermodynamics [19, Section 6.3], chemical reactions [43, Section 8.4], and species populations [20, 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 [8, Chapter 1]) the resultant Koopman operator cannot be expected to be well-defined. This observation is validated by [4], 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. [25]). Boundedness of composition operators, like the Koop-

man 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  $|a| < 1$ . This is a very small collection, and reveals how unlikely it is to have a bounded Koopman operator arising from the discretization of continuous time nonlinear systems.

**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 [25], since SOT convergence is the topology of pointwise convergence [31], 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. [31, pg. 172]). While [25] 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.

The above 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. These limitations necessitate a new direction in the study of operators that avoid the limiting processes that lean on the assumption of bounded or even *extant* Koopman operators. In this paper, we develop a new DMD technique that directly utilizes Liouville operators (which include Koopman generators).

A subset of Liouville operators that require the assumption of forward completeness on the dynamical system, called Koopman generators, have been studied as limits of Koopman operators in works such as [11, 12, 14, 15, 16, 17]. 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. 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. 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.

The direct involvement of Liouville operators in a DMD routine allows for the study of dynamics that are locally rather than globally Lipschitz, since Liouville operators do not impose an *a priori* discretization. For the adjoint of a Liouville operator to be well defined, the operator must be densely defined over the underlying RKHS [40, 39]. 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 polynomials, exponentials, 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 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 [40, 39], 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. The results of this manuscript allow for the natural separation of DMD methods for discrete time systems using the Koopman operator and continuous time systems using the Liouville operator, where previously continuous time systems were artificially discretized to fit within the Koopman framework.

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 [47]. 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 [26, Example 2.3]. Such systems include mechanical systems with high sampling frequencies [7, 44], and neurobiological systems recorded via electroencephalography (EEG) where the typical sampling frequencies are of the order of 500 Hz [18]. In the development of DMD for continuous time systems the methods in the present manuscript also

replace snapshots with trajectories or segments of trajectories of the system. The emphasis on trajectories over individual snapshot reduces the dimensionality of systems with an insurmountable number of snapshot without discarding any data.

The presented algorithm obviates the truncated SVD that is utilized throughout DMD methods. For example, in [47] 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 [37], the presented algorithm is given purely with respect to the occupation kernels, and the resultant methods are considerably simpler than what is seen in [47].

**A comparison with similar literature:** Liouville operators are studied in the context of DMD procedures using limiting definitions in works such as [24]. The manuscript [24], 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 [24] 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, such as occupation measures and time averaging functionals. These other quantities lack some of the essential characteristics that accompany occupation kernels. 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 on occupation kernels, and this fact is leveraged in Section 4 where finite rank representations of the Liouville operators arise from the matrix representation of the projection operator. Occupation kernels are 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 per the discussion above, this limits the applicability of time average functionals to globally Lipschitz dynamics, and therefore their utilization is more constrained from that of occupation kernels, whose definition is independent of Koopman and Liouville operators.

The relevant preliminary concepts for the theoretical underpinnings of the approach taken in the present manuscript are reviewed in Section 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  $\{c_1, c_2, \dots, c_M\} \subset X$ , the corresponding Gram matrix,

$$\begin{pmatrix} K(c_1, c_1) & \cdots & K(c_1, c_M) \\ \vdots & \ddots & \vdots \\ K(c_M, c_1) & \cdots & K(c_M, c_M) \end{pmatrix}$$

is positive semi-definite. The Gram matrix arises in many contexts in machine learning, such as in support vector machines (cf. [21]), and particular to the subject matter of this manuscript, it plays a pivotal role in the construction of the kernel-based extended DMD method of [47] 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 [1]. 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 theorem. When the RKHS is obtained from the kernel function, it is frequently referred to as the native space of that kernel function [45].

RKHSs interact with function theoretic operators, such as Koopman (composition) operators [22, 27, 47], multiplication operators [34, 35], and Toeplitz operators [36], in many nontrivial ways. For example, the kernel functions themselves play the role of eigenfunctions for the adjoints of multiplication operators [42], 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 [9]. 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 Section 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 [41, 21]. Specifically, a feature map corresponding to a RKHS is given as the mapping  $x \mapsto \Psi(x) := (\Psi_1(x), \Psi_2(x), \dots)^T \in \ell^2(\mathbb{N})$  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  $\mathbf{g} = (g_1, g_2, \dots)^T \in \ell^2(\mathbb{N})$  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 [47] 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 that of the Gaussian radial basis function's native space. The Gaussian radial basis

function 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 kernel 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 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 operator theory, the theory concerning bounded operators is most complete (cf. [31, 13]). 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 minimum  $C$  that holds 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) [13, 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 [31, 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  $\{g_m\}_{m=1}^\infty \in \mathcal{D}(W)$ , and both  $\{g_m\}_{m=1}^\infty$  and  $\{Wg_m\}_{m=1}^\infty$  are convergent sequences where  $g_m \rightarrow g \in H$  and  $Wg_m \rightarrow h \in H$ , then  $g \in \mathcal{D}(W)$  and  $Wg = h$  [31, 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 [39].  $\square$

The closedness of Koopman operators is well known in the study of RKHS, where they are more commonly known as composition operators (cf. [22, 27]). 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 [31, Chapter 5].

**Definition 2** For an operator  $W$  let

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

be dense in  $H$ . For each  $h \in \mathcal{D}(W^*)$  the Riesz theorem guarantees a function  $W^*h \in H$  such that  $\langle Wg, h \rangle_H = \langle g, W^*h \rangle_H$ . The adjoint of the operator  $W$  is thus given as  $W^* : \mathcal{D}(W^*) \rightarrow H$  via the assignment  $h \mapsto W^*h$ .

For a closed operator over a Hilbert space, the adjoint is densely defined [31]. Hence, Liouville operators with domains given as in Lemma 1, their adjoints 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. [39]).

**Definition 3** Let  $X$  be a metric space,  $\gamma : [0, T] \rightarrow X$  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 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  $\Gamma_\gamma$  is the *occupation kernel* corresponding to  $\gamma$  in  $H$ .

**Lemma 2** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the dynamics for a dynamical system, and suppose that  $\gamma : [0, T] \rightarrow \mathbb{R}^n$  is a trajectory satisfying  $\dot{\gamma} = f(\gamma(t))$  in the Caratheodory sense. In this setting,  $\Gamma_\gamma \in \mathcal{D}(A_f^*)$ . Moreover,  $A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$ .

*Proof* This lemma was established in [39].  $\square$

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. Moreover, for other RKHSs, different classes of dynamics will correspond to densely defined operators, and this requires 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  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,

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

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  $\gamma : [0, T] \rightarrow \mathbb{R}^n$  is a trajectory satisfying  $\dot{\gamma} = f(\gamma)$ , then

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

**Proposition 1** For  $\gamma : [0, T] \rightarrow \mathbb{R}^n$ , 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  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . There exists a collection of coefficients  $\{C_\alpha\}_\alpha$  indexed by the multi-index  $\alpha$  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.  $\square$

**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) position them as an answer for the desire of norm convergence for DMD type methods. For bounded Koopman operators, work such as [25] obtain convergence in the strong operator topology (SOT) of the DMD method 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, the norm convergence of the DMD procedure given in 4 will be demonstrated 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 limits the adjustments to the time series data through the operator to within machine precision, and 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. This method differs from the kernel-based extended DMD method of [47], where the kernel functions for the inputs are now replaced by occupation kernels, and the output is now a difference of kernel functions. This formulation allows for the snapshots of typical DMD methods to be strung together as trajectories. The occupation kernel-based DMD method then allows for the incorporation of all the snapshots of a given system to be incorporated into the DMD analysis in a way that reduces the dimensionality of the resultant problem to be less than the number of snapshots, while simultaneously allowing for the direct treatment of continuous time dynamical systems. If the rank of the resulting matrices needs to be increased, the trajectories may be segmented up to the number of snapshots.

It should also be noted that this method differs from [47] in that it avoids direct evaluations of the feature space. Thus, the succeeding method keeps with the spirit of the “kernel trick,” where the feature space is only accessed through evaluation of the kernel functions [41, pg. 19].

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  $\{\gamma_i : [0, T_i] \rightarrow X\}_{i=1}^M$  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  $\gamma_i$  the action of  $A_f^*$  on the corresponding occupation kernel is  $A_f^* \Gamma_{\gamma_i} = K(\cdot, \gamma_i(T_i)) - K(\cdot, \gamma_i(0))$ .

Thus, when  $\alpha$  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  $\alpha$  followed by projection onto  $\text{span}(\alpha)$ .

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

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

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

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

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

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

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

$$\langle A_f^* h, \Gamma_{\gamma_j} \rangle = \left\langle \sum_{i=1}^M v_i A_f^* \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \right\rangle_H = (\langle A_f^* \Gamma_{\gamma_1}, \Gamma_{\gamma_j} \rangle_H, \dots, \langle A_f^* \Gamma_{\gamma_M}, \Gamma_{\gamma_j} \rangle_H) \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}, \quad (3)$$

for all  $j = 1, \dots, M$ . Using (1) and (3), the coefficients  $\{w_i\}_{i=1}^M$  in the projection of  $A_f^* h$  onto  $\text{span}(\alpha)$  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,  $\text{span}(\alpha)$ , as  $[P_\alpha A_f^*]_\alpha^\alpha = G^{-1} \mathcal{I}$  where

$$G := \begin{pmatrix} \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_1}, \Gamma_{\gamma_M} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_1} \rangle_H & \cdots & \langle \Gamma_{\gamma_M}, \Gamma_{\gamma_M} \rangle_H \end{pmatrix} \quad (4)$$

is the Gram matrix of occupation kernels and

$$\mathcal{I} := \begin{pmatrix} \langle K(\cdot, \gamma_1(T_1)) - K(\cdot, \gamma_1(0)), \Gamma_{\gamma_1} \rangle_H & \cdots & \langle K(\cdot, \gamma_M(T_M)) - K(\cdot, \gamma_M(0)), \Gamma_{\gamma_1} \rangle_H \\ \vdots & \ddots & \vdots \\ \langle K(\cdot, \gamma_1(T_1)) - K(\cdot, \gamma_1(0)), \Gamma_{\gamma_M} \rangle_H & \cdots & \langle K(\cdot, \gamma_M(T_M)) - K(\cdot, \gamma_M(0)), \Gamma_{\gamma_M} \rangle_H \end{pmatrix}. \quad (5)$$

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

$$\begin{aligned} \langle A_f h, \Gamma_{\gamma_j} \rangle_H &= \sum_{i=1}^M v_i \langle A_f \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_H = \sum_{i=1}^M v_i \langle \Gamma_{\gamma_i}, A_f^* \Gamma_{\gamma_j} \rangle_H \\ &= (\langle \Gamma_{\gamma_1}, A_f^* \Gamma_{\gamma_j} \rangle_H, \dots, \langle \Gamma_{\gamma_M}, A_f^* \Gamma_{\gamma_j} \rangle_H) \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}. \end{aligned} \quad (6)$$

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

$$[P_\alpha A_f]_\alpha^\alpha = G^{-1} \mathcal{I}^T. \quad (7)$$

#### 4.2 Dynamic mode decomposition

Suppose that  $\lambda_i$  is the eigenvalue corresponding to the eigenvector  $v_i := (v_{i1}, v_{i2}, \dots, v_{iM})^T$ ,  $i = 1, \dots, M$ , 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  $\text{span}(\alpha)$ , 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 combination of the approximate eigenfunctions of  $A_f$  and *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

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

and  $(x)_j$  is viewed here 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  $i = 1, \dots, M$ , the matrix  $\xi$  may be expressed as

$$\xi = \left( (V^T G V)^{-1} V^T \begin{pmatrix} \int_0^T \gamma_1(t)^T dt \\ \vdots \\ \int_0^T \gamma_M(t)^T dt \end{pmatrix} \right)^T. \quad (8)$$

Given a trajectory  $x$  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:

$$x(t) \approx \sum_{i=1}^M \xi_i \varphi_i(x(0)) e^{\lambda_i t}, \quad (9)$$

where

$$\varphi_i(x(0)) = \frac{1}{N_i} \sum_{j=1}^M v_{ij} \Gamma_{\gamma_j}(x(0)) = \frac{1}{N_i} \sum_{j=1}^M v_{ij} \int_0^{T_j} K(x(0), \gamma_j(t)) dt. \quad (10)$$

---

**Algorithm 1** Pseudocode for the dynamic mode decomposition routine of Section 4. Once the Liouville modes, the normalized eigenvectors, and the eigenvalues are returned, (9) and (10) 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  $\{\gamma_j : [0, T] \rightarrow \mathbb{R}^n\}_{j=1}^M$

**Require:** Kernel function  $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  of an RKHS

**Require:** A numerical integration routine

- 1: Compute the Gram matrix  $G$  in (4) using (2) and a numerical integration routine
  - 2: Compute the interaction matrix  $\mathcal{I}$  in (5) using (2) and a numerical integration routine
  - 3: Compute eigenvalues,  $\lambda_i$ , and eigenvectors,  $v_i$ , of  $G^{-1}\mathcal{I}^T$
  - 4: Use (8) and a numerical integration routine to compute the matrix  $\xi$  of Liouville modes
  - 5: **return** Liouville modes,  $\xi_i$ , normalized eigenvectors,  $\frac{v_i}{N_i}$ , and eigenvalues  $\lambda_i$  for  $i = 1, \dots, M$
- 

#### 4.3 Modifications for the Scaled Liouville Operator DMD Method

Since Liouville operators are not generally compact, convergence of the finite rank representation  $P_\alpha A_f$ , restricted to  $\text{span } \alpha$ , to  $A_f$  cannot be guaranteed as  $M \rightarrow \infty$ . 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,  $\{T_{\gamma_i}\}_{i=1}^\infty \subset H$ , the resultant sequence of finite rank operators  $P_{\alpha_M} A_{f,a} P_{\alpha_M}$  converges to  $A_{f,a}$ , where  $\alpha_M := \{\gamma_1, \dots, \gamma_M\}$ . 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  $\text{span}(\alpha)$ , converges to that of  $A_{f,a}$ .<sup>1</sup>

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,  $A_{f,a}^* T_{\gamma_i} = K(\cdot, a\gamma_i(T_i)) - K(\cdot, a\gamma_i(0))$ . Hence, a finite rank representation of  $A_f$ , obtained from restricting and projecting to  $\text{span}(\alpha)$ , is given as

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

where

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

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.

---

<sup>1</sup> 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  $\text{span}(\alpha_M)$  as  $P_{\alpha_M} g = g$  for all  $g \in \text{span}(\alpha_M)$ .

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

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

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  $\lambda_{i,a} \rightarrow \lambda_{i,1}$  and  $\varphi_{i,a}(x(0)) \rightarrow \varphi_{i,1}(x(0))$  as  $a \rightarrow 1^-$ , then*

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

*Proof* The proof has been relegated to the appendix to ease exposition.  $\square$

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

$$x(t) \approx \sum_{i=1}^M \xi_{i,a} \varphi_{i,a}(x(0)) e^{\lambda_{i,a}t}. \quad (11)$$

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 achieve norm convergence when a sequence of trajectories corresponds to a collection of occupation kernels that are dense in the Hilbert space.

**Theorem 2** *Let  $|a| < 1$ . Suppose that  $\{\gamma_i : [0, T_i] \rightarrow \mathbb{R}^n\}_{i=1}^\infty$  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  $\alpha_M = \{\Gamma_{\gamma_1}, \dots, \Gamma_{\gamma_M}\}$ .*

*Proof* The proof has been relegated to the appendix to ease exposition.  $\square$

## 5 Numerical Experiments

This section gives the results two collections of numerical experiments 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 [26], 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 [26, Chapter 2].

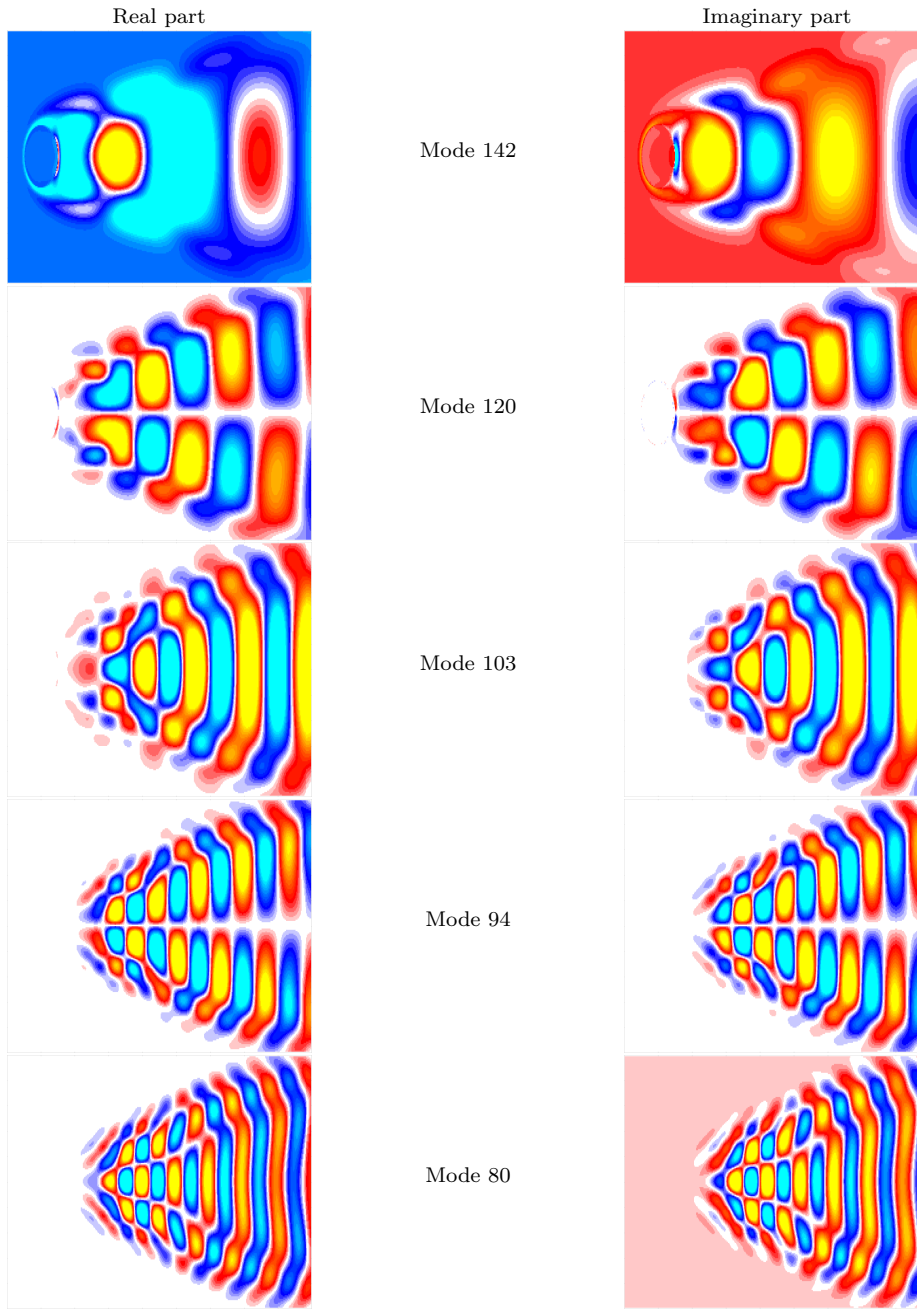
Figure 1 presents the Liouville modes obtained from the cylinder vorticity data set where the collection 151 snapshots was subdivided into 147 trajectories, each of length 5. This figure should be compared with Figure 2, which presents the scaled Liouville modes, with parameter  $a = 0.99$ , corresponding 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 (9) and (11), respectively.

### 5.2 SsVEP Dataset

This experiment uses data from [18]. 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 [18]. 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 [33, 32], 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 [2, 3, 30]).

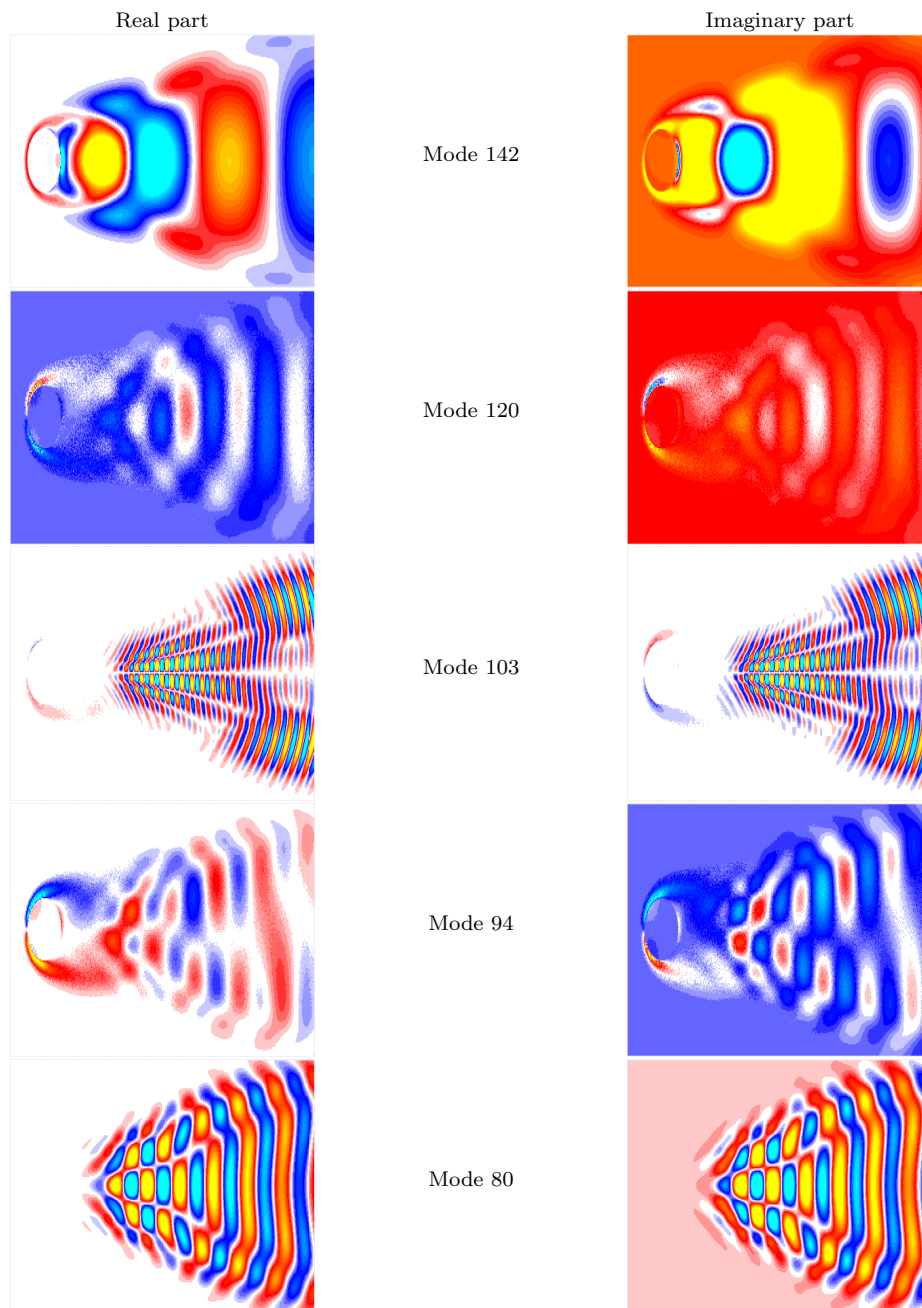
In this setting each independent trial can be used as a trajectory for a single occupation kernel. This differs from the implementation of Koopman based DMD, where most often each snapshot corresponds to a single trajectory. The continuous time DMD method was performed using the Gaussian kernel function with  $\mu = 50$ .

Figure 4 presents the obtained eigenvalues, and Figure 5 gives log scaled spectrum obtained from the eigenvectors. It can be seen that the spectrum has strong

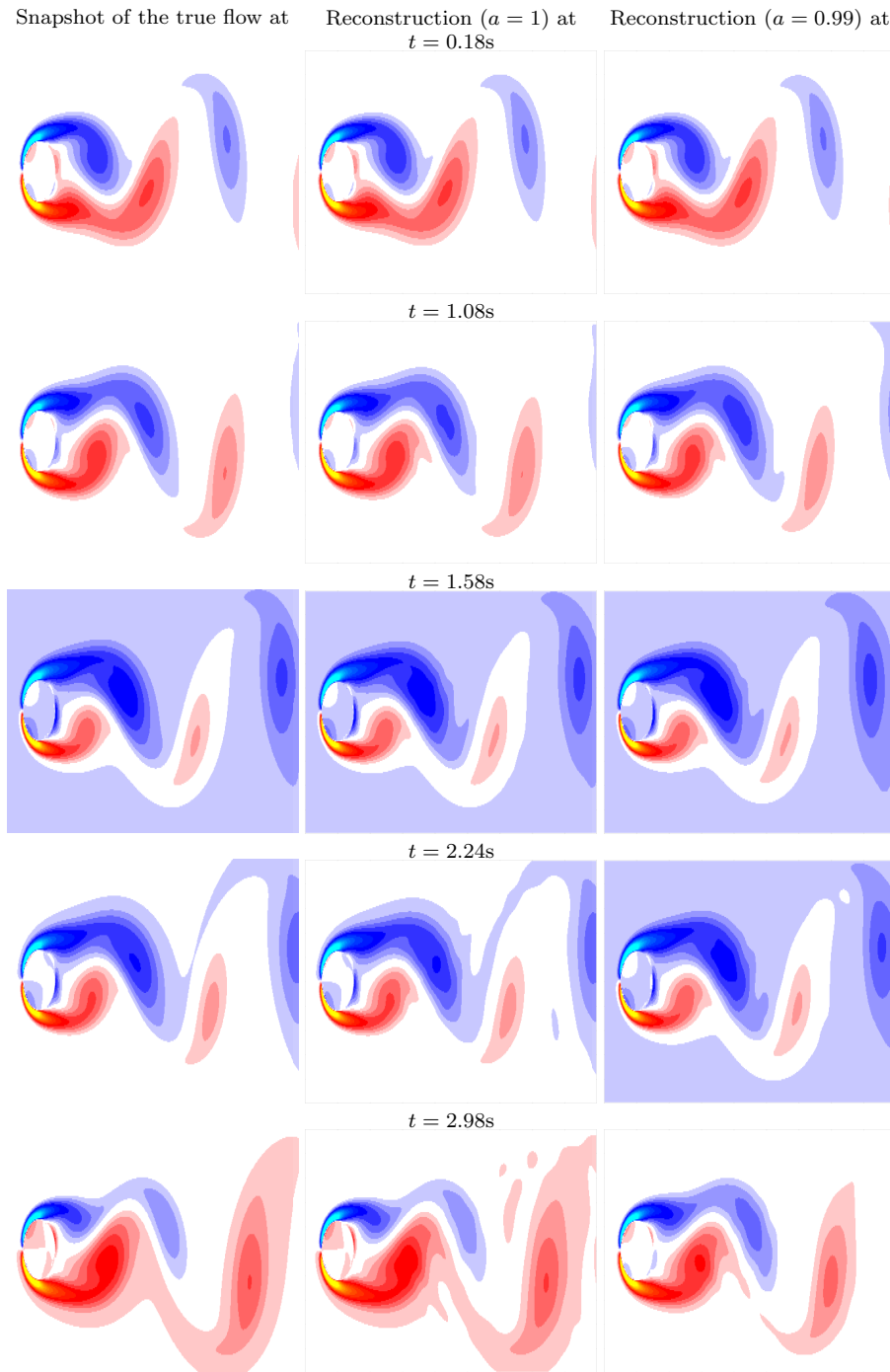


**Fig. 1** This figure presents the real and imaginary parts of a selection of ten Liouville modes determined by the continuous time DMD method given in the present manuscript corresponding to the vorticity of a flow across a cylinder (data available in [26]).

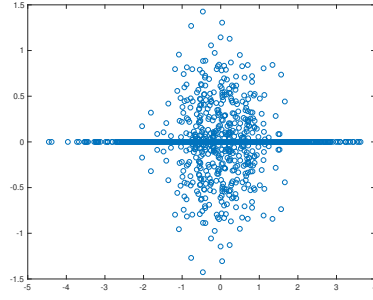




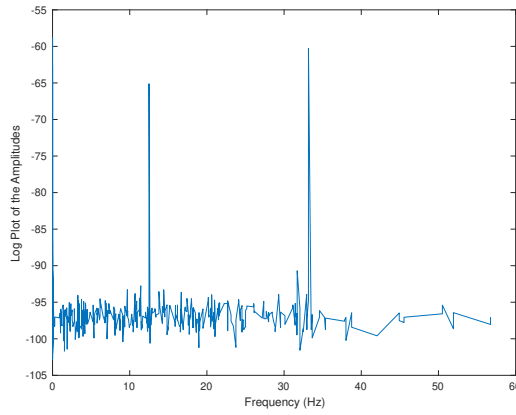
**Fig. 2** This figure presents the real and imaginary parts of a selection of five scaled Liouville DMD modes for the cylinder wake vorticity data in [26]. The difference between these modes and the modes in Figure 1 was anticipated for several reasons; the selection of  $a = 0.99$  is expected to result 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.



**Fig. 3** Snapshots of the true flow compared with reconstruction via the Liouville DMD model in (9) and the scaled Liouville DMD model in (11) with  $a = 0.99$ .



**Fig. 4** Eigenvalues corresponding to the ssVEP dataset from [18]. 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.

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 [26].

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, the employment of scaled Liouville operators, many dynamics yield a compact operator over the Bargman-Fock space, which allows for norm convergence of DMD procedures.

Liouville operators are unbounded in most cases, which owes strongly to the inclusion of the derivative or gradient in its definition. However, Koopman operators are also unbounded operators in all but a few cases. In the specific instance where the selected kernel function is the exponential dot product kernel, the Koopman operators are only bounded for linear dynamics. Thus, connections between DMD and the Koopman operator necessarily invoke the theory of unbounded operators. However, large classes of both Liouville and Koopman operators can be realized where they are densely defined and closed operators over RKHSs.

## 6.2 Finite Rank Representations

Since a Liouville operator is generally unbounded, convergence of the finite rank representation (in the norm topology) of Section 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, where for the exponential dot product kernel only Koopman operators corresponding to affine dynamics are bounded (cf. [6]), this 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 if this is not satisfied for the Liouville operator directly, that the occupation kernels are always in the domain of a closed and densely defined Liouville operator's adjoint. So an alternative algorithm may be designed using the adjoint's actions on the occupation kernels instead. Interestingly, the only adjustment the the algorithm in this setting would be a transposition of the matrix  $\mathcal{I}$ .

## 6.3 Approximating the Full State Observable

The establishment the decomposition of the full state observable relies very strongly on the selection of RKHS. In the case of the Bargmann-Fock space,  $x \mapsto (x)_i$  is a function in the space for each  $i = 1, \dots, n$ . However, this is not the case for the native space of the Gaussian RBF, which does not contain any polynomials in its native space. In both cases, these 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 a compact operator in this scale Liouville operator case, while Koopman operators cannot be modified in the same 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 in a certain sense.

The experiments presented in Section 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 [47]. 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 to this operator, 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 [25]. 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  $T_m g \rightarrow Tg$  for all  $g \in H$ . This mode of convergence is extremely limited, and not entirely appropriate for the spectral methods of 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, any infinitely many  $T_m$  from the sequence of operators converging in SOT to  $T$  may have dramatically different spectra from  $T$ , since those operators may not be part of the subsequence for convergence. This was remarked in [25]. Moreover,

the convergence result for DMD to the Koopman operator is a special case of a much more general theorem, where it is known that finite rank operators are dense in the collection of bounded operators with respect to SOT [31].

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 where convergence would be achievable are compact operators. Hence, this allows for the norm convergence of the finite rank approximations generated for DMD to scaled Liouville operators. More information concerning spectral theory and operator theory in general can be found in [31].

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

## 7 Conclusions

By targeting the DMD decomposition on Liouville operators, which includes Koopman generators as a proper subset, a decomposition of a continuous time dynamical system can be performed directly rather than that of a discrete time proxy for the dynamical system with the Koopman operator. Moreover, by obviating the limiting process using the Koopman operators in the definition of Liouville operators, a broader class of dynamics is accessible through this method, since the requirement of forward completeness may be relaxed. The notion of occupation kernels were leveraged to enable a DMD analysis of the Liouville operator, and a scaled Liouville operator was introduced to provide a collection of compact operators which allows for norm convergence of the DMD procedure. Two examples were presented, one from fluid dynamics and another EEG dataset, which allowed for the comparison of Liouville and scaled Liouville modes. The method presented here provides a new approach to DMD, which impacts the fundamental operator theory underlying traditional DMD with the Koopman operator.

## 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  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . There exists a collection of coefficients  $\{C_\alpha\}_\alpha$  indexed by the multi-index  $\alpha$  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  $A_{x^m,a} = A_{x,\sqrt[m]{a}}^m$  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  $A_{f,a} = \lim_{m \rightarrow \infty} \sum_{m=0}^M f_m A_{x^m,a}$ , 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  $\|g\|_{F^2(\mathbb{R})}^2 = \sum_{m=0}^{\infty} |g_m|^2 < \infty$ . 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,  $\|A_{x,a}g\|_{F(\mathbb{R})}^2 = |a|^{2m} m^2 |g_m|^2 < \infty$  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  $|a|^m m^2 \rightarrow 0$ , there is an  $M$  such that for all  $m > M$ ,  $|a|^m m^2 < 1$ . Let  $P_M$  be the projection onto  $\text{span}\{1, x, x^2, \dots, x^M\}$ . Now consider

$$\begin{aligned} \|(A_{x,a} - A_{x,a}P_M)g\|^2 &= \sum_{m=M+1}^{\infty} |g_m|^2 |a|^{2m} m^2 \\ &\leq \sum_{m=M+1}^{\infty} |g_m|^2 |a|^m \\ &\leq |a|^M \sum_{m=M+1}^{\infty} |g_m|^2 |a|^{m-M} \\ &\leq |a|^M \sum_{m=M+1}^{\infty} |g_m|^2 \leq |a|^M \|g\|_{F^2(\mathbb{R})}^2. \end{aligned}$$

Hence, the operator norm of  $(A_{x,a} - A_{x,a}P_M)$  is bounded by  $|a|^{M/2}$ , and as  $|a| < 1$ ,  $A_{x,a}P_M \rightarrow A_{x,a}$  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.  $\square$

**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  $\lambda_{i,a} \rightarrow \lambda_{i,1}$  and  $\varphi_{i,a}(x(0)) \rightarrow \varphi_{i,1}(x(0))$  as  $a \rightarrow 1^-$ , then*

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

*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

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

First, it is necessary to demonstrate that  $M_{1,a}$  and  $M_{2,a}$  may be bounded independent of  $a$ . For each  $i, j = 1, \dots, n$  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. [41]). 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. [41]). As  $\phi_{m,a}$  is a normal vector,  $\|\phi_{m,a}\|_H = 1$ , and by Cauchy-Schwarz

$$\begin{aligned} \|\nabla \phi_{m,a}(y)\|_2 &= \sqrt{\sum_{i=1}^n \left( \frac{\partial}{\partial x_i} \phi_{m,a}(y) \right)^2} \\ &= \sqrt{\sum_{i=1}^n \left( \left\langle \phi_{m,a}, \frac{\partial}{\partial x_i} k_y \right\rangle_H \right)^2} \\ &\leq \sqrt{\sum_{i=1}^n \|\phi_{m,a}\|_H^2 \left\| \frac{\partial}{\partial x_i} k_y \right\|_H^2} \\ &= \sqrt{\sum_{i=1}^n \left\| \frac{\partial}{\partial x_i} k_y \right\|_H^2}. \end{aligned} \quad (12)$$

(12) 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,

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

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,

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

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  $a$ .  $\square$

**Theorem 2 restated** Let  $|a| < 1$ . Suppose that  $\{\gamma_i : [0, T_i] \rightarrow \mathbb{R}^n\}_{i=1}^\infty$  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  $\alpha_M = \{\Gamma_{\gamma_1}, \dots, \Gamma_{\gamma_M}\}$ .

*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  $P_{\alpha_M} T P_{\alpha_M} \rightarrow T$  in norm, where  $P_{\alpha_M}$  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  $T_M \rightarrow T$  as  $M \rightarrow \infty$  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  $i = 1, \dots, M$ . Let  $g \in H$  be arbitrary.

Now consider,

$$\begin{aligned} \|Tg - P_n T P_n g\|_H &= \|Tg - T_M g + T_M g - P_n T P_n g\|_H \\ &\leq \|T - T_M\| \|g\|_H + \|T_M g - P_n T P_n g\|_H \leq \epsilon \|g\|_H + \|T_M g - P_n T P_n g\|_H. \end{aligned}$$

The second term after the inequality may be expanded as

$$\begin{aligned} \|T_M g - P_n T P_n g\|_H &\leq \|T_M g - P_n T_M g\|_H + \|P_n T_M g - P_n T P_n g\|_H \\ &\quad \|T_M g - P_n T_M g\|_H + \|T_M g - T P_n g\|_H \\ &\leq \|T_M g - P_n T_M g\|_H + \|T_M g - T_M P_n g\|_H + \|T_M P_n g - T P_n g\|_H \\ &\leq \|T_M g - P_n T_M g\|_H + \|T_M g - T_M P_n g\|_H + \epsilon \|g\|_H. \end{aligned}$$

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

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

and

$$\begin{aligned}
\|T_M(g - P_n g)\|_H &= \left\| \sum_{i=1}^M \lambda_i \langle g - P_n g, v_i \rangle_H u_i \right\| \\
&\leq \left\| \sum_{i=1}^M \lambda_i \left\langle \sum_{j=n+1}^{\infty} \langle g, g_j \rangle_H g_j, v_i \right\rangle_H u_i \right\| \\
&\leq \sum_{j=n+1}^{\infty} |\langle g, g_j \rangle_H| \left( \sum_{i=1}^M |\lambda_i| |\langle g_j, v_i \rangle_H| \right) \\
&\leq \left( \sum_{j=n+1}^{\infty} |\langle g, g_j \rangle_H|^2 \right)^{1/2} \left( \sum_{j=n+1}^{\infty} \left( \sum_{i=1}^M |\lambda_i| |\langle g_j, v_i \rangle_H| \right)^2 \right)^{1/2} \\
&\leq \|g\|_H \left( \sum_{j=n+1}^{\infty} \left( \sum_{i=1}^M |\lambda_i|^2 \right) \left( \sum_{i=1}^M |\langle g_j, v_i \rangle_H|^2 \right) \right)^{1/2} \\
&\leq \|g\|_H \left( \sum_{i=1}^M |\lambda_i|^2 \right)^{1/2} \left( \sum_{i=1}^M \sum_{j=n+1}^{\infty} |\langle g, v_i \rangle_H|^2 \right)^{1/2} \\
&= \|g\|_H \left( \sum_{i=1}^M |\lambda_i|^2 \right)^{1/2} \left( \sum_{i=1}^M \|v_i - P_n v_i\|^2 \right)^{1/2} \leq \epsilon \|g\|_H.
\end{aligned}$$

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  $n \rightarrow \infty$ ,  $P_n T P_n \rightarrow T$  in the operator norm.  $\square$

## Acknowledgments

This research was supported by the Air Force Office of Scientific Research (AFOSR) under contract numbers FA9550-20-1-0127, FA9550-15-1-0258, FA9550-16-1-0246, and FA9550-18-1-0122, the Air Force Research Laboratory (AFRL) under contract number FA8651-19-2-0009, and the Office of Naval Research (ONR) under contract N00014-18-1-2184. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agencies.

## References

1. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American mathematical society **68**(3), 337–404 (1950)
2. Bakardjian, H., Tanaka, T., Cichocki, A.: Optimization of SSVEP brain responses with application to eight-command brain–computer interface. Neuroscience letters **469**(1), 34–38 (2010)
3. 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. Journal of neural engineering **6**(4), 046002 (2009)
4. Bittracher, A., Koltai, P., Junge, O.: Pseudogenerators of spatial transfer operators. SIAM Journal on Applied Dynamical Systems **14**(3), 1478–1517 (2015)
5. Budišić, M., Mohr, R., Mezić, I.: Applied Koopmanism. Chaos: An Interdisciplinary Journal of Nonlinear Science **22**(4), 047510 (2012)
6. Carswell, B., MacCluer, B.D., Schuster, A.: Composition operators on the Fock space. Acta Sci. Math.(Szeged) **69**(3-4), 871–887 (2003)

7. 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 Transactions on Automation Science and Engineering* **12**(3), 945–957 (2015)
8. Coddington, E.A., Levinson, N.: *Theory of ordinary differential equations*. Tata McGraw-Hill Education (1955)
9. Cowen Jr, C.C., MacCluer, B.I.: *Composition operators on spaces of analytic functions*, vol. 20. CRC Press (1995)
10. Črnjarić-Žić, N., Maćešić, S., Mezić, I.: Koopman operator spectrum for random dynamical systems. *Journal of Nonlinear Science* **30**, 2007–2056 (2020)
11. 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) **69** (2005)
12. Das, S., Giannakis, D.: Koopman spectra in reproducing kernel Hilbert spaces. *Applied and Computational Harmonic Analysis* **49**(2), 573–607 (2020)
13. Folland, G.B.: *Real analysis: modern techniques and their applications*. John Wiley & Sons (2013)
14. Froyland, G., González-Tokman, C., Quas, A.: Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. *Journal of Computational Dynamics* **1**(2), 249–278 (2014)
15. Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis* **47**(2), 338–396 (2019)
16. Giannakis, D., Das, S.: Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis. *Physica D: Nonlinear Phenomena* **402**, 132211 (2020)
17. Giannakis, D., Kolchinskaya, A., Krasnov, D., Schumacher, J.: Koopman analysis of the long-term evolution in a turbulent convection cell. *arXiv:1804.01944* (2018)
18. Gruss, L.F., Keil, A.: Sympathetic responding to unconditioned stimuli predicts subsequent threat expectancy, orienting, and visuocortical bias in human aversive Pavlovian conditioning. *Biological psychology* **140**, 64–74 (2019)
19. Haddad, W.: *A Dynamical Systems Theory of Thermodynamics*. Princeton Series in Applied Mathematics. Princeton University Press (2019)
20. Hallam, T.G., Levin, S.A.: *Mathematical Ecology: An Introduction*, vol. 17. Springer Science & Business Media (2012)
21. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* **27**(2), 83–85 (2005)
22. Jury, M.T.:  $C^*$ -algebras generated by groups of composition operators. *Indiana University Mathematics Journal* pp. 3171–3192 (2007)
23. Khalil, H.K.: *Nonlinear Systems*, Third edn. Prentice hall Upper Saddle River, NJ (2002)
24. 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: Nonlinear Phenomena* **406**, 132416 (2020)
25. Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. *Journal of Nonlinear Science* **28**(2), 687–710 (2018)
26. Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM (2016)
27. Luery, K.E.: *Composition operators on Hardy spaces of the disk and half-plane*. University of Florida (2013)
28. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics* **41**(1-3), 309–325 (2005)
29. Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics* **45**, 357–378 (2013)
30. Middendorff, M., McMillan, G., Calhoun, G., Jones, K.S.: Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE transactions on rehabilitation engineering* **8**(2), 211–214 (2000)
31. Pedersen, G.K.: *Analysis Now, Graduate Texts in Mathematics*, vol. 118. Springer Science & Business Media (2012)
32. 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. *Journal of cognitive neuroscience* **29**(6), 953–967 (2017)

33. Regan, D.: Human brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. Elsevier (1989)
34. Rosenfeld, J.A.: Densely defined multiplication on several Sobolev spaces of a single variable. *Complex Analysis and Operator Theory* **9**(6), 1303–1309 (2015)
35. Rosenfeld, J.A.: Introducing the Polylogarithmic Hardy Space. *Integral Equations and Operator Theory* **83**(4), 589–600 (2015)
36. Rosenfeld, J.A.: The Sarason sub-symbol and the recovery of the symbol of densely defined Toeplitz operators over the Hardy space. *Journal of Mathematical Analysis and Applications* **440**(2), 911–921 (2016)
37. Rosenfeld, J.A., Kamalapurkar, R.: Dynamic mode decomposition with control Liouville operators. In: *International Symposium on Mathematical Theory of Networks and Systems* (2021). To appear, see [arXiv:2101.02620](https://arxiv.org/abs/2101.02620)
38. Rosenfeld, J.A., Kamalapurkar, R., Gruss, L.F., Johnson, T.T.: On occupation kernels, Liouville operators, and dynamic mode decomposition. In: *Proc. Am. Control Conf.* (2021). To appear, see <https://scc-lab.github.io/Preprints/SCC.Rosenfeld.Kamalapurkar.ea2021.pdf>
39. Rosenfeld, J.A., Kamalapurkar, R., Russo, B., Johnson, T.T.: Occupation kernels and densely defined Liouville operators for system identification. In: *IEEE Conference on Decision and Control*, pp. 6455–6460 (2019)
40. Rosenfeld, J.A., Russo, B., Kamalapurkar, R., Johnson, T.T.: The occupation kernel method for nonlinear system identification. [arXiv:1909.11792](https://arxiv.org/abs/1909.11792) (2019)
41. Steinwart, I., Christmann, A.: Support vector machines. Springer Science & Business Media (2008)
42. Szafraniec, F.H.: The reproducing kernel Hilbert space and its multiplication operators. In: *Complex Analysis and Related Topics*, pp. 253–263. Springer (2000)
43. Tóth, J., Nagy, A.L., Papp, D.: Reaction kinetics: exercises, programs and theorems. Springer (2018)
44. 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 Transactions on Robotics* **34**(2), 486–496 (2018)
45. Wendland, H.: Scattered data approximation, *Cambridge Monographs on Applied and Computational Mathematics*, vol. 17. Cambridge university press (2004)
46. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science* **25**(6), 1307–1346 (2015)
47. Williams, M.O., Rowley, C.W., Kevrekidis, I.G.: A kernel-based method for data-driven Koopman spectral analysis. *Journal of Computational Dynamics* **2**(2), 247–265 (2015)