Convergent Dynamic Mode Decomposition

Joel A. Rosenfeld and Rushikesh Kamalapurkar

Abstract— **This manuscript addresses convergence of dynamic mode decomposition (DMD) algorithms and the existence of associated Koopman modes. Convergence relies on reformulation of dynamic mode decomposition in terms of newly defined compact operators defined with pairs of Hilbert spaces selected separately as the domain and range of the operator. With the Hilbert spaces selected so that the domain is embedded in the range, an eigenfunction approach to DMD is developed by leveraging a finite rank representation. The finite rank representation is proven to converge, in norm, to the original operator with increasing rank. The manuscript concludes with the description of a DMD algorithm that converges when a dense collection of occupation kernels, arising from the data, are leveraged in the analysis.**

I. Introduction

Dynamic mode decomposition (DMD) methods are data analysis methods that aim to decompose a time series corresponding to a nonlinear dynamical system into a collection of dynamic modes [1]–[4]. The time series is then expressed as a linear combination of the dynamic modes. The coefficients in the linear combination are given by exponential functions of time. The dynamic modes and the growth rates of the exponential functions are derived from the spectrum of a finite rank representation of the Koopman operator (or in the continuous-time case, Koopman generator). Convergence of finite rank representations of the Koopman operator has only been established with respect to the strong operator topology (SOT) [4], which does not guarantee the convergence of the spectrum [5], and hence, the corresponding DMD algorithms are not guaranteed to converge.

Koopman operators analyze continuous time dynamics through a discrete time proxy obtained by fixing a time-step for a continuous time system [6]. However, only a small subset of continuous time dynamics satisfy the forward complete property necessary to obtain a discretization [7]. Moreover, to establish convergence guarantees for DMD routines, additional structure is required of Koopman operators, where a sequence of finite rank operators converge to Koopman operators in norm only if the Koopman operator is compact [5]. Since compactness is rarely satisfied for Koopman operators [7], the study of alternative operator representations of nonlinear systems is well-motivated.

Another motivation for the use of Koopman operators in the study of continuous time dynamical systems is a heuristic that for small time steps the spectra and eigenfunctions of the resultant Koopman operator should be close to that of the Liouville operator representing the continuous time systems [8]. However, for two fixed time steps, the corresponding Koopman operators can have different collections of eigenfunctions and eigenvalues, and these are artifacts of the discretization itself [9]. Since in most cases Koopman operators are used for this analysis, it is not clear if there is a method for distinguishing which of these eigenfunctions and eigenvalues are a product of the discretization and which are fundamental to the dynamics themselves.

In the approach developed in this paper, the above limitations are addressed by removing Koopman operators from the analysis in favor of Liouville operators (known as Koopman generators in special cases), and these operators are shown to be compact for certain pairs of Hilbert spaces selected separately as the domain and range of the operator. Assuming that the domain is embedded in the range of the operator, eigenfunctions of the Liouville operator can be approximated through finite rank approximations that converge to the Liouville operator. The result is a norm convergent DMD method which significantly improves upon the SOT convergent results previously established in the field [4].

It should be noted that there have been several attempts at providing compact operators for the study of DMD. The approaches [10] and [7] find compact operators through the multiplication of auxiliary operator against Koopman and Liouville operators respectively. However, the resultant operators are not the operators that truly correspond to the dynamics in question, and as such, the decomposition of those operators can only achieve heuristic results. The approach taken here generates compact Liouville operators directly connected with the continuous time dynamics.

The main results of this paper are discussed within a larger context of singular dynamic mode decomposition in [11].

II. Reproducing Kernel Hilbert Spaces

A. Occupation Kernels

A reproducing kernel Hilbert space (RKHS), H, over a compact set X, is a space of functions from X to $\mathbb{R}$ such that the functional of evaluation, $E_x g := g(x)$ is bounded for every $x \in X$. By the Riesz theorem, this means for each $x \in X$ there exists a function $K_x \in H$ such that $\langle f, K_x \rangle_H = f(x)$ for all f. The function K_x is called the kernel function

This research was supported by the Air Force Office of Scientific Research (AFOSR) under contract numbers FA9550-20-1-0127 and FA9550-21-1-0134, and the National Science Foundation (NSF) under awards 2027976 and 2027999. 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.

J. A. Rosenfeld is with the Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA, email: rosenfeldj@usf.edu

R. Kamalapurkar is with the School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK, USA, email: rushikesh.kamalapurkar@okstate.edu

centered at X, and the function $K(x,y) := \langle K_y, K_x \rangle_H$ is called the kernel function corresponding to H. Note that $K_y(x) = K(x,y)$. Classical examples of kernel functions in data science are the Gaussian radial basis function for $\mu > 0$, $K(x,y) = \exp(-\frac{1}{\mu}\|x-y\|^2)$, and the exponential dot product kernel, $\exp(\frac{1}{\mu}x^\top y)$ [12].

The function $K(x,y)$ is a positive definite kernel function, which means that for every finite collection of points, $\{x_1, \ldots, x_M\} \subset X$, the Gram matrix $(K(x_i,x_j))_{i,j=1}^M$ is positive definite. For each positive definite kernel function, there exists a unique RKHS for which K is the kernel function for that space by the Aronszajn-Moore theorem in [13].

Given a RKHS, H, over $X \in \mathbb{R}^n$ consisting of continuous functions and given a continuous signal, $\theta : [0,T] \to X$, the linear functional $g \mapsto \int_0^T g(\theta(t))dt$ is bounded. Hence, there exist a function, $\Gamma_\theta \in H$, such that $\langle g, \Gamma_\theta \rangle_H = \int_0^T g(\theta(t))dt$ for all $g \in H$. The function Γ_θ is called the occupation kernel in H corresponding to θ. These occupation kernels were first introduced in [14], [15].

B. Dynamic Mode Decomposition from the Perspective of RKHSs

The motivation in DMD is to compute an invariant subspace of a transfer operator that models the evolution of test functions, known as observables, along the trajectories of a dynamical system. When the full state observable (a vector-valued function that maps x to itself) is projected onto this subspace, a model is obtained in the form of a linear combination of a basis of the subspace. The subspace is typically constructed as the span of eigenfunctions of the transfer operator. While transfer operators over RKHSs may not admit a complete eigendecomposition, DMD methods aim to construct a finite rank representation of the transfer operator and to leverage the eigenfunctions of the approximating operator for modeling.

The objective is to find functions ϕ in the domain of the Liouville operator with symbol f, defined as $A_f g := \nabla g \cdot f$, that satisfy

$$|A_f\phi(x) - \lambda\phi(x)| < \epsilon \tag{1}$$

for some λ and some small positive ϵ and all x within some workspace. Once accomplished, given a trajectory, $\dot{x} = f(x)$, the eigenfunction behaves approximately as $\phi(x(t)) = x(0)e^{\lambda t}$. The significance of RKHSs is that norm convergence implies pointwise convergence, which has been leveraged in approximation and machine learning frameworks since the 1990s [12], [16]. Since norm convergence in a RKHS of continuous functions yields uniform convergence over compact sets, kernel methods allow for a relaxation of (1), where it is sufficient to satisfy $\|A_f\phi - \lambda\phi\|_H < \epsilon$. In turn, if a finite rank approximation of A_f, call it $\tilde{A}_f$, is close enough, it is sufficient to satisfy $\|A_f - \tilde{A}_f\| < \epsilon$, and the rest follows as

$$|A_f\phi(x) - \lambda\phi(x)| < C\|A_f\phi - \lambda\phi\|_H$$
$$C\|A_f\phi - \tilde{A}_f\phi\|_H < C\|A_f - \tilde{A}_f\|_H < C\epsilon,$$

where C is a positive constant that depends on the workspace and the kernel function, and the function ϕ is assumed to be normalized. In the case where the kernel is the Gaussian RBF, C may be taken to be 1.

It is important to note that finite rank operators themselves are almost always diagonalizable, and determining a collection of approximate eigenfunctions from the eigenfunctions of the finite rank operators is certainly well defined. However, to obtain a close approximation of a dynamic operator using a finite rank operator requires compactness, which motivates the investigation of the present manuscript.

III. Compact Liouville Operators

This section demonstrates the existence of compact Liouville operators where compactness is achieved through the consideration of differing spaces for the domain and range of the operator. Section III-A builds on a classical result where differentiation between differing weighted Hardy spaces can be readily shown to be compact. Following a similar argument, Section III-B presents several examples of compact Liouville operators over spaces of functions of several variables. We would like to emphasize that the collections of compact Liouville operators are not restricted to these particular pairs of functions spaces, but rather this section provides several examples demonstrating the existence of such operators, thereby validating the approach in the sequel.

A. Inspirations from Classical Function Theory

Consider the weighted Hardy spaces (cf. [17]), H^2_ω, where $\omega = \{\omega_m\}_{m=0}^\infty$ is a sequence of positive real numbers such that $|\omega_{m+1}/\omega_m| \to 1$, and $g(z) = \sum_{m=0}^\infty a_m z^m$ is a function in H^2_ω if the coefficients of g satisfy $\|g\|^2_{H^2_\omega} := \sum_{m=0}^\infty \omega_m |a_m|^2 < \infty$. Each weighted Hardy space is a RKHS over the complex unit disc $\mathbb{D} = \{z \in \mathbb{C} : |z| = 1\}$ with kernel function given as $K_\omega(z,w) = \sum_{m=0}^\infty \omega_m z^m \bar{w}^m$, and the monomials $\left\{\frac{z^m}{\sqrt{\omega_m}}\right\}_{m=0}^\infty$ form an orthonormal basis for each space.

The weighted Hardy space corresponding to the sequence $\omega_{(0)} := \{1,1,\ldots\}$ is the classical Hardy space, H^2, that was introduced by Riesz in 1923 [18]. The Dirichlet space corresponds to the weight sequence $\omega_{(1)} = \{(m+1)\}_{m=0}^\infty$, and the Bergman space corresponds to $\omega_{(-1)} = \{(m+1)^{-1}\}_{m=0}^\infty$. Of interest here is the weighted Hardy space corresponding to $\omega_{(3)} := \{(m+1))^3\}_{m=0}^\infty$, which will be denoted as H^2_3 for convenience.

Proposition 1: The operator $\frac{d}{dz} : H^2_3 \to H^2$ is compact. Moreover, if f is a bounded analytic function from the closure of $\mathbb{D}$ (denoted by $\overline{\mathbb{D}}$) to $\mathbb{D}$, corresponding to a bounded multiplication operator $M_f g := g(x)f(x)$ over the Hardy space, then the Liouville operator $A_f = M_f \frac{d}{dz}$ from H^2_3 to H^2 is compact.

Proof: To see that differentiation is a compact operator from the H^2_3 to the Hardy space, we may select a sequence of finite rank operators that converge in norm to differentiation. In particular, note that the monomials form an orthonormal basis of the Hardy space as is evident from the given norm. Let $\alpha_M := \{1, z, \ldots, z^M\}$ be the first M monomials in

z, and let P_{α_M} be the projection onto the span of these monomials. The operator $P_{\alpha_M}\frac{d}{dz}$ is a finite rank operator, where the image of this operator is a polynomial of degree up to M.

To demonstrate that this sequence of finite rank operators converges to differentiation in the operator norm it must be shown that the difference under the operator norm,

$$\left\| P_{\alpha_M}\frac{d}{dz} - \frac{d}{dz} \right\|_{H^2_3}^{H^2} := \sup_{g\in H^2_3} \frac{\|P_{\alpha_M}\frac{d}{dz}g - \frac{d}{dz}g\|_{H^2}}{\|g\|_{H^2_3}},$$

goes to zero. Note that

$$\begin{aligned}\|P_{\alpha_M}\frac{d}{dz}g - \frac{d}{dz}g\|_{H^2}^2 &= \sum_{m=M+1}^{\infty}(m+1)^2|a_{m+1}|^2 \\ &= \sum_{m=M+1}^{\infty}\frac{1}{m+1}(m+1)^3|a_{m+1}|^2 \\ \leq \frac{1}{M+1}\sum_{m=M+1}^{\infty}(m+1)^3|a_{m+1}|^2 &\leq \frac{1}{M+1}\|g\|_{H^2_3}^2.\end{aligned}$$

Hence $\left\|P_{\alpha_M}\frac{d}{dz} - \frac{d}{dz}\right\|_{H^2_3}^{H^2} \leq \frac{1}{M+1} \to 0$. This proves that differentiation is a compact operator from H^2_3 to H^2.

If a function, f, is a bounded analytic function on the closed unit disc, then it is the symbol for a bounded multiplier over H^2. Hence, the $M_f\frac{d}{dz}$ is a compact operator from H^2_3 to H^2. To be explicit, since $P_{\alpha_M}\frac{d}{dz}$ has finite rank, $M_f\left(P_{\alpha_M}\frac{d}{dz}\right)$ also has finite rank. Moreover, $\left\|M_f P_{\alpha_M}\frac{d}{dz} - M_f\frac{d}{dz}\right\|_{H^2_3}^{H^2} = \left\|M_f\left(P_{\alpha_M}\frac{d}{dz} - \frac{d}{dz}\right)\right\|_{H^2_3}^{H^2} \leq \|M_f\|_{H^2}^{H^2}\left\|P_{\alpha_M}\frac{d}{dz} - \frac{d}{dz}\right\|_{H^2_3}^{H^2} \to 0$. Hence, $M_f\frac{d}{dz}$ is an operator norm limit of finite rank operators, and is compact. Finally, it can be seen that $M_f\frac{d}{dz}g(z) = g'(z)f(z) = A_f g(z)$, and A_f is a compact Liouville operator from H^2_3 to H^2. ∎

B. Compact Liouville Operators of Several Variables

The exponential dot product kernel, with parameter $\mu > 0$, is given as $K(x,y) = \exp\left(\mu x^T y\right)$. In the single variable case, the native space for this kernel may be expressed as $F^2_\mu(\mathbb{R}) = \left\{f(x) = \sum_{m=0}^{\infty} a_m x^m : \sum_{m=0}^{\infty}|a_m|^2\frac{m!}{\mu^m} < \infty\right\}$. This definition can be readily extended to higher dimensions, where collection of monomials, $x^\alpha\frac{\mu^{|\alpha|}}{\sqrt{\alpha!}}$, with multi-indices $\alpha \in \mathbb{N}^n$ form an orthonormal basis. The norm of functions in $F^2_\mu(\mathbb{R}^n)$ will be denoted by $\|g\|_\mu$.

In this setting, compactness of partial differentiation with respect to each variable can be established under the following hypothesis.

Lemma 1: If $\eta < \mu$, then the operators $\frac{\partial}{\partial x_i} : F^2_\eta(\mathbb{R}^n) \to F^2_\mu(\mathbb{R}^n)$ are compact for all $i = 1,\ldots,n$.

Proof: Follows from arguments similar to Proposition 1. ∎

Since the Liouville operator involves composition of partial differentiation operators and multiplication operators, and since multiplication operators are unbounded from $F^2_\mu(\mathbb{R}^n)$ to itself for every $\mu > 0$, another step is necessary to ensure compactness of the Liouville operator.

Lemma 2: Suppose that $\eta < \mu$, then given any polynomial of several variables, f, the multiplication operator $M_f : F^2_\eta(\mathbb{R}^n) \to F^2_\mu(\mathbb{R}^n)$ is bounded.

Proof: See [11, Lemma 3.2]. ∎

Remark 1: The authors emphasize that the collection of bounded multiplication operators between these spaces is strictly larger than the those with polynomial symbols. The purpose of this lemma is to simply support the existence of compact Liouville operators, rather than to provide a complete classification.

Theorem 1: Let $\mu_3 > \mu_1$, and suppose that f is a vector valued function over several variables, where each entry is a polynomial. Then the Liouville operator $A_f : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_3}(\mathbb{R}^n)$ given by $A_f g = \nabla g \cdot f$ is a compact operator.

Proof: Let $f = (f_1, f_2, \ldots, f_n)^T$, and select μ_2 such that $\mu_1 < \mu_2 < \mu_3$. For each $i = 1,\ldots,n$, the operator of partial differentiation $\frac{\partial}{\partial x_i} : F^2_{\mu_1}(\mathbb{R}^n) \to F^2_{\mu_2}(\mathbb{R}^n)$ is a compact operator, and the multiplication operator $M_{f_i} : F^2_{\mu_2}(\mathbb{R}^n) \to F^2_{\mu_3}(\mathbb{R}^n)$ is bounded. Hence, the operator $M_{f_i}\frac{\partial}{\partial x_i}$ is compact. As $A_f = M_{f_1}\frac{\partial}{\partial x_1} + \cdots + M_{f_n}\frac{\partial}{\partial x_n}$, it follows that A_f is a compact operator from $F^2_{\mu_1}(\mathbb{R}^n)$ to $F^2_{\mu_3}(\mathbb{R}^n)$. ∎

This section has thus established the existence of compact Liouville operators between various pairs of spaces. It is emphasized that these are not the only pairs for which a compact Liouville may be determined.

IV. Eigenfunction Approach to Convergent DMD

While the majority of this manuscript is aimed at the singular DMD, where the domain and range are different for the compact Liouville operator, there is still a possibility of obtaining an eigendecomposition in special cases. In particular, for many of the examples shown above, the domain and range spaces have similar structure and the range space has less stringent requirement for the functions it contains. This means that the domain itself may be embedded in the range space, and if there is a complete set of eigenfunctions in this embedded space, then the operator may still be diagonalized.

Note that the operator is still mapping between two different Hilbert spaces, which means that the inner product on the embedding is different than the inner product on the domain. This difference will appear in the numerical methods given in subsequent sections.

The following is a well known result (cf. [19]), and is included here for completeness.

Proposition 2: If $\mu_1 < \mu_2$, then $F^2_{\mu_1}(\mathbb{R}^n) \subset F^2_{\mu_2}(\mathbb{R}^n)$.

A simple example demonstrating that an eigenbasis may be found between the two spaces arises in the study of $A_x : F^2_{\mu_1}(\mathbb{R}) \to F^2_{\mu_2}(\mathbb{R})$ for $\mu_1 < \mu_2$.

Example 1: Given $A_x : F^2_{\mu_1}(\mathbb{R}) \to F^2_{\mu_2}(\mathbb{R})$ for $\mu_1 < \mu_2$, an eigenfunction, φ, for A_x must reside in $F^2_{\mu_1}(\mathbb{R}) \cap F^2_{\mu_2}(\mathbb{R}) = F^2_{\mu_1}(\mathbb{R})$, and satisfy $\varphi'(x)x = \lambda\varphi(x)$. Consequently, takes the form $\varphi(x) = x^\lambda$, and is in $F^2_{\mu_1}(\mathbb{R})$ only for $\lambda = 0, 1, 2, \ldots$. Hence, the eigenfunctions of A_x are the monomials. Monomials are contained in $F^2_{\mu_1}(\mathbb{R})$ and form a

complete eigenbasis for both spaces. Note that the norm of x^m is $\sqrt{\frac{m!}{\mu_1^m}}$ in $F^2_{\mu_1}(\mathbb{R})$ and $\sqrt{\frac{m!}{\mu_2^m}}$ in $F^2_{\mu_2}(\mathbb{R})$.

The following proposition is obtained in the same manner as in the classical case.

Proposition 3: Suppose that H and $\tilde{H}$ are two RKHSs over $\mathbb{R}^n$, and that $H \subset \tilde{H}$. If $\varphi \in H$ is an eigenfunction for A_f as $A_f\varphi = \lambda\varphi$, then given a trajectory $x : [0,T] \to \mathbb{R}^n$ satisfying $\dot{x} = f(x)$ the following holds $\varphi(x(t)) = e^{\lambda t}\varphi(x(0))$.

Proof: Since $A_f\varphi = \nabla\varphi f$, it follows that

$$\frac{d}{dt}\varphi(x(t)) = \nabla\varphi(x(t))f(x(t)) = A_f\varphi(x(t)) = \lambda\varphi(x(t)).$$

That is, $\frac{d}{dt}\varphi(x(t)) = \lambda\varphi(x(t))$. Thus, the conclusion follows. ∎

Suppose that $A_f : H \to \tilde{H}$ has a complete eigenbasis in the sense that the span of the eigenfunctions, $\{\varphi_m\}_{m=1}^{\infty}$, are dense in H. If g_{id}, is the full state observable, then each entry of g_{id}, $(x)_i$ for $i = 1, \ldots, n$, may be expressed as $(x)_i = \lim_{M\to\infty}\sum_{m=1}^{M}(\xi_{m,M})_i\varphi_m(x)$, where $(\xi_{m,M})_i$ is the m-th coefficient obtained from projecting $(x)_i$ onto the span of the first M eigenfunctions. If the eigenfunctions are orthogonal, then the dependence on M may be removed from $\xi_{m,M}$. Hence, the full state observable is obtained from $g_{id}(x) = \lim_{M\to\infty}\sum_{m=1}^{M}\xi_{m,M}\varphi_m(x)$, with $\xi_{m,M}$ being the vector obtained by stacking $(\xi_{m,M})_i$. Finally, by substituting $x(t)$ into this representation (where $\dot{x} = f(x)$), $x(t) = g_{id}(x(t)) = \lim_{M\to\infty}\sum_{m=1}^{M}\xi_{m,M}e^{\lambda t}\varphi_m(x(0))$. Hence, this methodology yields a DMD routine, where the finite rank representations will converge to the compact Liouville operators, following the proof given in the Appendix of [7].

V. The Eigenfunction based DMD Algorithm

In this section it will be assumed that $A_f : H \to \tilde{H}$ is a compact operator, and that $H \subset \tilde{H}$. Since A_f is compact, it is bounded, which means that unlike [9] and [7], no additional assumptions are needed concerning the domain of this operator.

A. Derivation of the Eigenfunction Method

Let $A_f : H \to \tilde{H}$ be the Liouville operator with symbol f and let $H \subset \tilde{H}$. For a collection of trajectories, $\{\gamma_1, \ldots, \gamma_M\}$ let $\alpha = \{\Gamma_{\gamma_1}, \ldots, \Gamma_{\gamma_M}\}$ be the corresponding occupation kernels in H and $\beta = \{\tilde{\Gamma}_{\gamma_1}, \ldots, \tilde{\Gamma}_{\gamma_M}\}$ be the corresponding occupation kernels in $\tilde{H}$. Let $P_\alpha : H \to H$ be the projection onto the span of α viewed as a subspace of H, $\tilde{P}_\alpha : H \to \tilde{H}$ be the projection onto the span of α viewed as a subspace of $\tilde{H}$, and $\tilde{P}_\beta : H \to \tilde{H}$ be projection onto the span of β. The following theorem constructs a matrix representation for the operator $\tilde{P}_\alpha\tilde{P}_\beta A_f$ restricted to $\operatorname{span}\alpha \subset H$, denoted by $\tilde{P}_\alpha\tilde{P}_\beta A_f|_\alpha$. In particular, the matrix, $[\tilde{P}_\alpha\tilde{P}_\beta A_f]^\alpha_\alpha$, represents the operator $\tilde{P}_\alpha\tilde{P}_\beta A_f|_\alpha$.

Theorem 2: If $h = \sum_{i=1}^{M} a_i\Gamma_{\gamma_i} \in \operatorname{span}\alpha \subset H$ and $g = \sum_{i=1}^{M} b_i\Gamma_{\gamma_i} \in \operatorname{span}\alpha \subset \tilde{H}$ are functions such that $g = \tilde{P}_\alpha\tilde{P}_\beta A_f h$, then

$$b = G^{-1}_{\alpha,\tilde{H}}G_{\alpha,\beta,\tilde{H}}G^{-1}_{\beta,\tilde{H}}D_\alpha a,$$

where the matrix $G_{\beta,\tilde{H}} := \left[\langle\tilde{\Gamma}_{\gamma_i}, \tilde{\Gamma}_{\gamma_j}\rangle_{\tilde{H}}\right]_{i,j=1}^{M}$ is the Gram matrix for the basis β in the space $\tilde{H}$, the matrix $G_{\alpha,\tilde{H}} := \left[\langle\Gamma_{\gamma_i}, \Gamma_{\gamma_j}\rangle_{\tilde{H}}\right]_{i,j=1}^{M}$ is the Gram matrix for the basis α in the space $\tilde{H}$, $G_{\alpha,\beta,\tilde{H}} = \left[\langle\Gamma_{\gamma_i}, \tilde{\Gamma}_{\gamma_j}\rangle_{\tilde{H}}\right]_{i,j=1}^{M}$ is the interaction matrix, and $D_\alpha = \left[\Gamma_{\gamma_j}(\gamma_i(T_i)) - \Gamma_{\gamma_j}(\gamma_i(0))\right]_{i,j=1}^{M}$ is the matrix of occupation kernel differences. That is, the matrix

$$[\tilde{P}_\alpha\tilde{P}_\beta A_f]^\alpha_\alpha := G^{-1}_{\alpha,\tilde{H}}G_{\alpha,\beta,\tilde{H}}G^{-1}_{\beta,\tilde{H}}D_\alpha \tag{2}$$

represents the operator $\tilde{P}_\alpha\tilde{P}_\beta A_f|_\alpha$.

Proof: See [11, Section 7]. ∎

Note that when $H = \tilde{H}$ and the occupation kernels are assumed to be in the domain of the Liouville operator, the product $G^{-1}_{\alpha,\tilde{H}}G_{\alpha,\beta,\tilde{H}}$ reduces to the identity matrix and the representation reduces to that of [7].

Algorithm 1 Pseudocode for the eigenfunction based DMD routine of Section V. Once the singular DMD modes, the normalized eigenfunctions, and the eigenvalues are returned, (5) can be used along with a numerical integration routine to reconstruct trajectories of the system starting from any given initial condition $x(0)$. Evaluation of the eigenfunctions at $x(0)$ requires the integral representation $\Gamma_{\gamma_j}(\cdot) = \int_0^{T_j} K(\cdot, \gamma_j(t))\,dt$. 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. If the matrices in steps 1, 2, and 3 are close to singular, they can be regularized by adding $\epsilon \mathrm{I}_{M\times M}$ where $\epsilon > 0$ is a small constant.

Require: Sampled trajectories $\{\gamma_j : [0,T] \to \mathbb{R}^n\}_{j=1}^{M}$
Require: Kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ of the domain RKHS H
Require: Kernel function $\tilde{K} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ of the range RKHS $\tilde{H}$
Require: A numerical integration routine

1: Construct the matrix $G_{\beta,\tilde{H}}$ using $\langle\tilde{\Gamma}_{\gamma_j}, \tilde{\Gamma}_{\gamma_i}\rangle_{\tilde{H}} = \int_0^{T_i}\int_0^{T_j}\tilde{K}(\gamma_i(\tau), \gamma_j(t))dtd\tau$ and a numerical integration routine (cf. [7])
2: Construct the matrix $G_{\alpha,\tilde{H}}$ using (8) and a numerical integration routine
3: Construct the matrix $G_{\alpha,\beta,\tilde{H}}$ using (7) and a numerical integration routine
4: Construct the matrix D_α using the integral representation $\Gamma_{\gamma_j}(\cdot) = \int_0^{T_j} K(\cdot, \gamma_j(t))\,dt$ and a numerical integration routine
5: Construct the matrix $[\tilde{P}_\alpha\tilde{P}_\beta A_f]^\alpha_\alpha$ using (2) and compute its eigenvalues, λ_i, and eigenvectors, V_i
6: Use (3) and a numerical integration routine to compute the eigenfunctions $\hat{\varphi}_j$
7: Use (4) and a numerical integration routine to compute the singular DMD modes $\hat{\xi}_j$
8: **return** Singular DMD modes, ξ_j, eigenfunctions, $\hat{\varphi}_j$, and eigenvalues λ_j for $j = 1, \cdots, M$

Under the assumption of diagonalizability for (2), which

holds for almost all matrices, an eigendecomposition for (2) may be determined as

$$[\tilde{P}_\alpha \tilde{P}_\beta A_f]_\alpha^\alpha = V \Lambda V^{-1},$$

where each column, V_j, of the matrix V is an eigenvector of $[\tilde{P}_\alpha \tilde{P}_\beta A_f]_\alpha^\alpha$ with eigenvalue λ_j, the j-th diagonal element of the diagonal matrix Λ. The corresponding normalized eigenfunction is given as

$$\hat{\varphi}_j(x) = \frac{1}{\sqrt{V_j^\top G_{\alpha,H} V_j}} V_j^\top \begin{pmatrix} \Gamma_{\gamma_1} \\ \vdots \\ \Gamma_{\gamma_M} \end{pmatrix}, \tag{3}$$

where the normalization is performed using $G_{\alpha,H} := \left[\langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_H\right]_{i,j=1}^M$, the Gram matrix for the basis α in the Hilbert space H. Set $\bar{V}_j := \frac{1}{\sqrt{V_j^\top G_{\alpha,H} V_j}} V_j$, and let $\bar{V} := \begin{pmatrix} \bar{V}_1 \cdots \bar{V}_M \end{pmatrix}$.

The Gram matrix for the normalized eigenbasis may be quickly computed as $\bar{V}^\top G_{\alpha,H} \bar{V}$, and the weights for the projection of the full state observable onto this eigenbasis may be written as

$$\begin{pmatrix} -\hat{\xi}_1^\top - \\ \vdots \\ -\hat{\xi}_M^\top - \end{pmatrix} = (\bar{V}^\top G_{\alpha,H} \bar{V})^{-1} \bar{V}^\top \begin{pmatrix} \int_0^{T_1} \gamma_1(t)^\top dt \\ \vdots \\ \int_0^{T_1} \gamma_M(t)^\top dt \end{pmatrix} \tag{4}$$

and thus, $g_{id}(x) \approx \sum_{m=1}^M \hat{\xi}_m \hat{\varphi}_m(x)$. The approximation error (with respect to the norm of the RKHS) approaches zero if the number of trajectories increases and the corresponding collection of occupation kernels forms a dense set. Convergence in the norm of the RKHS implies uniform convergence on compact subsets of the domain.

Consequently, a trajectory $x : [0, T] \to \mathbb{R}^n$ satisfying $\dot{x} = f(x)$ may be approximately expressed as

$$x(t) = g_{id}(x(t)) \approx \sum_{m=1}^M \hat{\xi}_m e^{\lambda_m t} \hat{\varphi}_m(x(0)), \tag{5}$$

where the eigenfunctions for the finite rank approximation of A_f play the role of eigenfunctions for the original operator, A_f. Furthermore, the vector field f may be approximated as

$$f(x) \approx \hat{f}(x) := \sum_{m=1}^M \lambda_m \hat{\xi}_m \hat{\varphi}_m(x). \tag{6}$$

B. Computation of Inner Products

Some entries for the matrices in the above computations require a bit more analysis. Namely, this includes the inner products, $\langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_{\tilde{H}}$ and $\langle \Gamma_{\gamma_i}, \tilde{\Gamma}_{\gamma_j} \rangle_{\tilde{H}}$. All the other quantities have been discussed at length in [7], [14], [15].

The second quantity simply utilizes the functional definition of the function $\tilde{\Gamma}_{\gamma_j}$ as a function in $\tilde{H}$, $\langle \Gamma_{\gamma_i}, \tilde{\Gamma}_{\gamma_j} \rangle_{\tilde{H}} = \int_0^{T_j} \Gamma_{\gamma_i}(\gamma_j(t)) dt = \int_0^{T_j} \int_0^{T_i} K(\gamma_j(t), \gamma_i(\tau)) d\tau dt$, where K is the kernel function for H. Note that this means

$$\langle \Gamma_{\gamma_i}, \tilde{\Gamma}_{\gamma_j} \rangle_{\tilde{H}} = \langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_H. \tag{7}$$

However, the first quantity is more complicated and is context dependent. In particular, Γ_{γ_i} is not the occupation kernel corresponding to $\tilde{H}$, so it's functional relationship cannot be exploited in the same manner. On the other hand, $\Gamma_{\gamma_i}(x) = \int_0^{T_i} K(x, \gamma_i(t))$. To compute the inner product in $\tilde{H}$, a specific selection of spaces must be considered.

In the particular setting where $H = F_{\mu_1}^2(\mathbb{R}^n)$ and $\tilde{H} = F_{\mu_2}^2(\mathbb{R}^n)$, with $\mu_1 < \mu_2$, it follows that $\Gamma_{\gamma_i}(x) = \int_0^\top e^{\mu_1 x^\top \gamma_i(t)} dt$. Moreover, $K(x, \gamma_i(t)) = e^{\mu_1 x^\top \gamma_i(t)} = e^{\mu_2 x^\top \left(\frac{\mu_1}{\mu_2} \gamma(t)\right)} = \tilde{K}(x, (\mu_1/\mu_2)\gamma_i(t))$. Hence, $\Gamma_{\gamma_i}(x) = \tilde{\Gamma}_{(\mu_1/\mu_2)\gamma_i}(x)$, and

$$\begin{aligned} \langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle_{\tilde{H}} &= \langle \tilde{\Gamma}_{(\mu_1/\mu_2)\gamma_i}, \tilde{\Gamma}_{(\mu_1/\mu_2)\gamma_j} \rangle_{\tilde{H}} \\ &= \int_0^{T_i} \int_0^{T_j} \tilde{K}((\mu_1/\mu_2)\gamma_i(t), (\mu_1/\mu_2)\gamma_j(\tau)) d\tau dt. \end{aligned} \tag{8}$$

The eigenfunction approach to convergent DMD is summarized in Algorithm 1.

VI. Numerical Experiment

This section presents the results obtained through implementation of the developed method in with the domain viewed as embedded in the range of the operator. To demonstrate the effectiveness of the developed algorithm, 100 trajectories of the Duffing oscillator $\dot{x} = \begin{pmatrix} x_2 \\ -\delta x_2 - \beta x_1 - \alpha x_1^3 \end{pmatrix}$ with $\delta = 0.011$, $\alpha = 1$, and $\beta = -1$ are generated, starting from a 10×10 grid of initial conditions in the domain $D := [-5, 5] \times [-5, 5]$. Each trajectory is 10 seconds long and sampled every 0.05 seconds. The convergent DMD algorithm is implemented using exponential dot product kernels with $\mu_1 = 300$ and $\mu_2 = 298$. The Gram matrices are regularized using $\epsilon = 10^{-8}$.

To verify the model generated by convergent DMD, a trajectory $\hat{x}(\cdot)$ starting from $\hat{x}(0) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ is estimated by solving the differential equation $\dot{\hat{x}} = \hat{f}(\hat{x})$, where $\hat{f}$ is the vector field defined in (6). The estimated trajectory is plotted against the true trajectory $x(\cdot)$ in Figure 1. Figure 2 shows the pointwise 2-norm of the trajectory estimation error, $\|x(t) - \hat{x}(t)\|$, plotted as a function of time. Figure 3 shows the pointwise 2-norm of the relative vector field estimation error, $\frac{\|f(x) - \hat{f}(x)\|}{\max_{x \in D} \|f(x)\|}$, as a function of x.

VII. Results and Discussion

Simulation results demonstrate the efficacy of the developed method. As shown in Figure 1, the developed model can generate trajectories of the Duffing oscillator starting from initial conditions that are within the domain spanned by the data. A 20 second long trajectory is reproduced from a dataset that uses 100 trajectories, each 10 seconds long. As demonstrated by Figure 3, the Duffing vector field is well-approximated on the same domain by the convergent DMD-based model. While the approximation results in Figures 1 and 3 indicate divergence of the reproduced trajectory and vector field from the true trajectory and vector field for large t and x near the boundary of the domain spanned by the

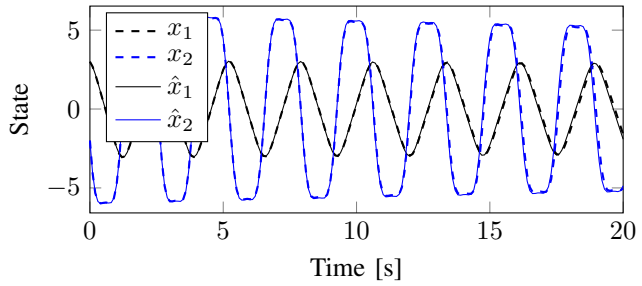

Fig. 1. True (dashed) and reconstructed (solid) trajectories of the Duffing oscillator. Reconstruction is done by numerically solving (using Matlab® `ode45` function) $\dot{x} = \hat{f}(x)$ where $\hat{f}$ is the estimated vector field in (6).

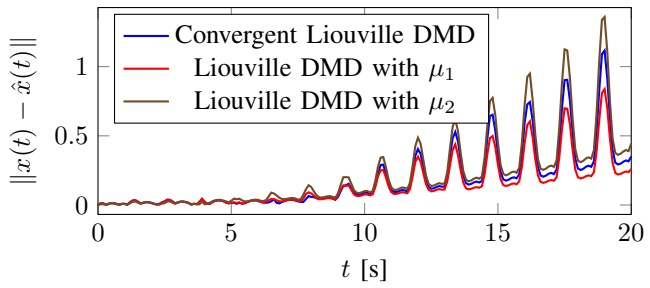

Fig. 2. Comparison of reconstruction errors for convergent Liouville DMD and the Liouville DMD method in [7].

data, respectively, such divergence is to be expected in purely data-driven methods.

One interesting result of the structure of the finite rank approximation given in Section V is that as $\mu_1 \to \mu_2$, the first two matrices cancel. The matrix computations then approach the computations in [7]. Hence, as seen in Figure 2, for close enough μ_1 and μ_2 the computations and the results of convergent DMD are very similar to those of [7] over a fixed compact set containing the trajectories.

VIII. Conclusion

In this paper, a theoretical and algorithmic framework that achieves many long standing goals of DMD is developed. To wit, it is shown that compact versions of Liouville operators (sometimes Koopman generators) can be obtained by selecting differing domains and ranges. This comes at the sacrifice of eigenfunctions when the domain is not embedded in the range of the operator. If the domain is in the range of the operator, then eigenfunctions can be appropriately defined and constructed. Reconstruction of system trajectories can then be achieved by solving initial value problems. The resulting algorithm, while comparable in performance with algorithms based on more traditional definitions of the Liouville operator, achieves theoretical convergence guarantees that are lacking in the traditional formulation.

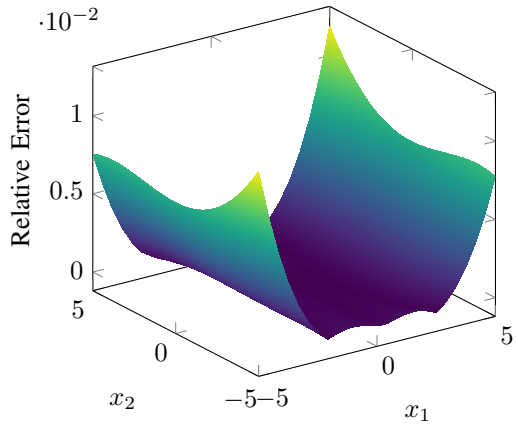

Fig. 3. Pointwise 2-norm of the relative vector field estimation error $\frac{\|f(x)-\hat{f}(x)\|}{\max_{x\in D}\|f(x)\|}$, as a function of x.

References

[1] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition - data-driven modeling of complex systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2016.

[2] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, p. 047510, 2012.

[3] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dyn.*, vol. 41, no. 1, pp. 309–325, 2005.

[4] M. Korda and I. Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," *J. Nonlinear Sci.*, vol. 28, no. 2, pp. 687–710, 2018.

[5] G. K. Pedersen, *Analysis now*. Springer Science & Business Media, 2012, vol. 118.

[6] A. Mauroy and I. Mezić, "Global stability analysis using the eigenfunctions of the Koopman operator," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3356–3369, 2016.

[7] J. A. Rosenfeld, R. Kamalapurkar, L. F. Gruss, and T. T. Johnson, "Dynamic mode decomposition for continuous time systems with the Liouville operator," *J. Nonlinear Sci.*, vol. 32, no. 1, pp. 1–30, 2022.

[8] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.

[9] E. Gonzalez, M. Abudia, M. Jury, R. Kamalapurkar, and J. A. Rosenfeld, "The kernel perspective on dynamic mode decompositions," arXiv:2106.00106.

[10] S. Das, D. Giannakis, and J. Slawinska, "Reproducing kernel Hilbert space compactification of unitary evolution groups," *Appl. Comput. Harmon. Anal.*, vol. 54, pp. 75–136, 2021.

[11] J. A. Rosenfeld and R. Kamalapurkar, "Singular dynamic mode decomposition," *SIAM J. Appl. Dyn. Syst.*, vol. 22, no. 3, pp. 2357–2381, 2023.

[12] I. Steinwart and A. Christmann, *Support vector machines*, ser. Information Science and Statistics. New York: Springer, 2008.

[13] N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[14] J. A. Rosenfeld, R. Kamalapurkar, B. Russo, and T. T. Johnson, "Occupation kernels and densely defined Liouville operators for system identification," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 6455–6460.

[15] J. A. Rosenfeld, B. Russo, R. Kamalapurkar, and T. Johnson, "The occupation kernel method for nonlinear system identification," arXiv:1909.11792, submitted to SIAM Journal on Control and Optimization.

[16] H. Wendland, *Scattered data approximation*. Cambridge University Press, 2004.

[17] C. Bénéteau, M. C. Fleeman, D. S. Khavinson, D. Seco, and A. A. Sola, "Remarks on inner functions and optimal approximants," *Can. Math. Bull.*, vol. 61, no. 4, pp. 704–716, 2018.

[18] F. Riesz, "Über die Randwerte einer analytischen Funktion," *Math. Z.*, vol. 18, no. 1, pp. 87–95, 1923.

[19] K. Zhu, *Analysis on Fock spaces*, ser. Graduate Texts in Mathematics. Springer, New York, 2012, vol. 263.