On Convergent Dynamic Mode Decomposition and its Equivalence with Occupation Kernel Regression [*]

Moad Abudia [*] **Joel Rosenfeld** [**] **Rushikesh Kamalapurkar** [*]

[*] *Oklahoma State University, Stillwater, OK, USA (e-mail:* {*abudia,rushikesh.kamalapurkar*}*@okstate.edu).*
[**] *University of South Florida, Tampa, FL, USA (e-mail:* *rosenfeldj@usf.edu)*

Abstract: This paper presents a new technique for norm-convergent dynamic mode decomposition of deterministic systems. The developed method utilizes recent results on singular dynamic mode decomposition where it is shown that by appropriate selection of domain and range Hilbert spaces, the Liouville operator (also known as the Koopman generator) can be made to be compact. In this paper, it is shown that by selecting appropriate collections of finite basis functions in the domain and the range, a novel finite-rank representation of the Liouville operator may be obtained. It is also shown that the model resulting from dynamic mode decomposition of the finite-rank representation is closely related to regularized regression using the so-called occupation kernels as basis functions.

Keywords: Operator Theoretic Methods in Systems Theory; Machine Learning and Control; System Identification

1. INTRODUCTION

Dynamic mode decomposition (DMD) methods are data analysis methods that aim to generate a finite-rank representation of a transfer operator corresponding to a nonlinear dynamical system using time series measurements (Kutz et al., 2016; Budišić et al., 2012; Mezić, 2005; Korda and Mezić, 2018). The convergence of the finite-rank representations to the true transfer operator (the Koopman operator) has been established in results such as Korda and Mezić (2018), but only with respect to the strong operator topology (SOT). Convergence in SOT does not guarantee the convergence of the spectrum (Pedersen, 2012), and therefore the corresponding dynamic mode decomposition (DMD) algorithms, which rely on spectrum of the operator, are not guaranteed to converge.

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). These operators are shown to be compact provided their domains and ranges are selected appropriately. The result is a norm convergent finite-rank representation which significantly improves upon the aforementioned SOT convergent results.

There have been several attempts to provide compact operators for DMD. The approaches in Das et al. (2021)

and Rosenfeld et al. (2022) find compact operators through the multiplication of an auxiliary operator against Koopman and Liouville operators, respectively. However, the resultant operators only approximately correspond to the dynamics in question, and as such, the resulting DMD methods, while useful, are heuristic.

The approach in this paper generates compact Liouville operators that truly correspond to the given continuous-time dynamics. Such operators were shown to exist for a large class of nonlinear systems in Rosenfeld and Kamalapurkar (2023), where norm-convergent finite-rank representations were also derived via the adjoint of the operators. In this paper, we develop a finite-rank representation that approximates the Liouville operator directly, rather than through its adjoint. The direct approximation makes the resulting DMD algorithm numerically efficient and conceptually easier. Interestingly, the resulting model is equivalent to the occupation kernel regression (OKR) model developed in Li and Rosenfeld (2020).

OKR is a generalization of kernel ridge regression Zhdanov and Kalnishkan (2013) where the loss function is defined in terms of inner products of components of the vector field that models the dynamics of the system and trajectory-dependent functions in a reproducing kernel Hilbert space (RKHS) called occupation kernels that represent integration along the trajectory. A Representer theorem is used to construct an approximation of the vector field as a linear combination of occupation kernels. Since the DMD model developed in this paper is seen to be identical to the OKR model without regularization, convergence results derived in this paper are also applicable to OKR, provided the regularization parameter is set to zero. To facilitate the

discussion, the following section recalls a few important characteristics of DMD methods and RKHSs.

2. REPRODUCING KERNEL HILBERT SPACES AND DYNAMIC MODE DECOMPOSITION

An RKHS, H, over a compact set X, is a space of functions from X to $\mathbb{R}$ such that the evaluation functional $E_x : H \to \mathbb{R}$, defined as $E_x g := g(x)$, is bounded for every $x \in X$. By the Riesz representation theorem (Roman, 2008, Theorem 13.32), 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 $f \in H$. The function K_x is called the kernel function centered at x, and the function $K(x, y) := \langle K_y, K_x \rangle_H$ is called the reproducing kernel of H (Steinwart and Christmann, 2008, Chapter 4).

A symmetric function $K : X \times X \to \mathbb{R}$ is called a positive semidefinite kernel if for every integer $M > 0$ and 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 semidefinite. By the Aronszajn-Moore theorem (Aronszajn, 1950), given any positive semidefinite kernel K, there exists a unique RKHS H such that K is the reproducing kernel of H.

The motivation in DMD is to compute an invariant subspace of a transfer operator A_f that models the evolution of test functions along the trajectories of a dynamical system $\dot{x} = f(x)$. The transfer operator maps a test function g to its time derivative $\nabla g \cdot f$. The subspace is typically constructed as the span of eigenfunctions of the operator. While such transfer operators over RKHSs may not admit point spectra, DMD methods aim to construct a finite-rank representation of the transfer operator and to leverage the spectrum of the approximating operator for modeling.

The objective is to find functions for which
$$|A_f \phi(x) - \lambda \phi(x)| < \epsilon \tag{1}$$
for some λ and some small positive ϵ and all x within a domain of interest. Since norm convergence in a RKHS of continuous functions yields uniform convergence over compact sets (Steinwart and Christmann, 2008; Wendland, 2004), 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 domain of interest and the kernel function, and the function ϕ is assumed to be normalized.

A convergent approximation of the spectrum of the transfer operator using the spectrum of finite-rank operators requires compactness and convergence in norm, which motivates the following section.

3. COMPACT LIOUVILLE OPERATORS AND OCCUPATION KERNELS

This section, included here for completeness, recalls the results from Rosenfeld and Kamalapurkar (2023) pertaining to the existence of compact Liouville operators, given

formally as $A_f g(x) = \nabla g(x) f(x)$. Compactness is achieved through the consideration of differing spaces for the domain and the range of the operator. We emphasize that compact Liouville operators are not restricted to these particular pairs of functions spaces. This section only provides examples demonstrating the existence of such operators, thereby validating the approach in the sequel.

3.1 Liouville Operators on Real Bergmann-Fock Spaces

The exponential dot product kernel, with parameter $\mu > 0$, is given as $K(x, y) = \exp\left(\frac{x^\top y}{\mu}\right)$. In the single variable case, the native space for this kernel is the restriction of the Bergmann-Fock space to real numbers, denoted by $F_\mu^2(\mathbb{R})$. This space consists of the set of polynomials of the form $h(x) = \sum_{k=0}^\infty a_k x^k$, where the coefficients satisfy $\sum_{k=0}^\infty |a_k|^2 \mu^k k! < \infty$, and the norm is given by $\|h\|_\mu^2 = \sum_{k=0}^\infty |a_k|^2 \mu^k k!$. Extension of this definition to the multivariable case yields the space $F_\mu^2(\mathbb{R}^n)$ where the collection of monomials, $x^\alpha \frac{\mu^{|\alpha|}}{\sqrt{\alpha!}}$, with multi-indices $\alpha \in \mathbb{N}^n$ forms an orthonormal basis, where for $\alpha \in \mathbb{N}^n$, $\alpha! = \prod_{i=1}^n \alpha_i!$, $|\alpha| = \sum_{i=1}^n \alpha_i$, and $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$. In this setting, differential operators from $F_{\mu_1}^2(\mathbb{R}^n)$ to $F_{\mu_2}^2(\mathbb{R}^n)$ can be shown to be compact provided $\mu_2 < \mu_1$.

Proposition 1. (Rosenfeld and Kamalapurkar (2023)). If $\mu_2 < \mu_1$, then the differential operators $\frac{\partial}{\partial x_i} : F_{\mu_1}^2(\mathbb{R}^n) \to F_{\mu_2}^2(\mathbb{R}^n)$, are compact for $i = 1, \ldots, n$.

Multiplication operators can be shown to be bounded provided that their symbols are polynomial.

Proposition 2. (Rosenfeld and Kamalapurkar (2023)). If $\mu_2 < \mu_1$, then for any polynomial function $p : \mathbb{R}^n \to \mathbb{R}$, the multiplication operator $M_p : F_{\mu_1}^2(\mathbb{R}^n) \to F_{\mu_2}^2(\mathbb{R}^n)$, defined as $[M_p h](x) = p(x)h(x)$, is bounded.

Boundedness of Liouville operators with polynomial symbols then follows trivially from the above two results.

Theorem 1. (Rosenfeld and Kamalapurkar (2023)). If $\mu_2 < \mu_1$ and if $f : \mathbb{R}^n \to \mathbb{R}^n$ is a component-wise polynomial function, then the Liouville operator $A_f : F_{\mu_1}^2(\mathbb{R}^n) \to F_{\mu_2}^2(\mathbb{R}^n)$ defined as $A_f g = \nabla g \cdot f$ is a compact operator.

3.2 Occupation Kernels and Liouville Operators

Let H be an RKHS over a compact set $X \in \mathbb{R}^n$ consisting of continuous functions and let K be the reproducing kernel of H. Given a continuous signal $\gamma : [0, T] \to X$, the linear functional $g \mapsto \int_0^T g(\gamma(t)) \mathrm{d}t$ is bounded. Hence, by the Riesz representation theorem (Roman, 2008, Theorem 13.32), there exists a function $\Gamma_\gamma \in H$ such that $\langle g, \Gamma_\gamma \rangle_H = \int_0^T g(\gamma(t)) \mathrm{d}t$ for all $g \in H$. The function Γ_γ is called the occupation kernel corresponding to γ in H. These occupation kernels were first introduced in Rosenfeld et al. (2019b, 2024).

Occupation kernels corresponding to trajectories of a dynamical system have a useful relationship with adjoints of Liouville operators and the reproducing kernels of the underlying RKHSs.

Proposition 3. (Rosenfeld et al. (2022)). If $\gamma : [0, T] \to X$ is a solution of $\dot{x} = f(x)$ for some locally Lipschitz continuous function $f : X \to \mathbb{R}^n$, then $A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$.

Proposition 4. (Rosenfeld et al. (2019a)). Given any continuous function $\gamma : [0, T] \to X$, the occupation kernel corresponding to γ in H may be expressed as $\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t)) \mathrm{d}t$. As a consequence, given two functions $\gamma_i : [0, T_i] \to X$ and $\gamma_j : [0, T_j] \to X$

$$\left\langle \Gamma_{\gamma_j}, \Gamma_{\gamma_i} \right\rangle_H = \int_0^{T_i} \int_0^{T_j} K\left(\gamma_i(\tau), \gamma_j(t)\right) \mathrm{d}t \mathrm{d}\tau. \quad (2)$$

4. FINITE-RANK REPRESENTATION OF THE LIOUVILLE OPERATOR

This section provides a novel finite-rank representation of the Liouville operator A_f and subsequently, a novel approach to obtain the dynamic modes of the underlying dynamical system. Let H_d and H_r be RKHSs with reproducing kernels K_d and K_r, respectively. The RKHS H_d is used as the domain of the Liouville operator and the RKHS H_r is used as the range. In what follows, finite collections of vectors $d^M \subset H_d$ and $r^M \subset H_r$ are selected to establish the needed finite-rank representation of the Liouville operator.

Since the adjoint of A_f maps occupation kernels to kernel differences (Proposition 3), the span of the collection of kernel differences

$$d^M = \{K_d(\cdot, \gamma_i(T_i)) - K_d(\cdot, \gamma_i(0))\}_{i=1}^M \subset H_d \quad (3)$$

are selected to be the domain of A_f. The corresponding Gram matrix is denoted by $G_{d^M} = \left(\langle d_i, d_j \rangle_{H_d}\right)_{i,j=1}^M$. The result of A_f operating on functions in span d^M is projected onto the span of the collection of occupation kernels

$$\text{span } r^M = \text{span} \{\Gamma_{\gamma_i}\}_{i=1}^M \subset H_r. \quad (4)$$

The corresponding Gram matrix is denoted by $G_{r^M} = \left(\langle r_i, r_j \rangle_{H_r}\right)_{i,j=1}^M$. A rank-$M$ representation of the operator A_f is then given by $P_{r^M} A_f P_{d^M} : H_d \to \text{span } r^M$, where P_{r^M} and P_{d^M} denote projection operators onto span r^M and span d^M, respectively. Under the compactness assumptions and given rich enough data so that the spans of $\{d_i\}_{i=1}^\infty$ and $\{r_i\}_{i=1}^\infty$ are dense in H_d and H_r, respectively, the sequence of finite-rank operators $\{P_{r^M} A_f P_{d^M}\}_{M=1}^\infty$ can be shown to converge, in the norm topology, to A_f. To facilitate the proof of convergence, we recall the following result from Rosenfeld et al. (2022).

Proposition 5. (Rosenfeld et al. (2022)). Let H_d and H_r be RKHSs defined on $X \subset \mathbb{R}^n$ and let $A_N : H_d \to H_r$ be a finite-rank operator with rank N. If span$\{d_i\}_{i=1}^\infty \subset H_d$ is dense in H_d and span$\{r_i\}_{i=1}^\infty \subset H_r$ is dense in H_r, then for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$ and $h \in H_d$, $\|A_N h - A_N P_{d^i} h\|_{H_r} \leq \epsilon \|h\|_{H_d}$ and $\|A_N h - P_{r^i} A_N h\|_{H_r} \leq \epsilon \|h\|_{H_d}$.

The convergence result for Liouville operators on Bergmann-Fock spaces restricted to the set of real numbers follows from the following more general result.

Proposition 6. If $A : H_d \to H_r$ is a compact operator, and spans of the collections $\{d_i\}_{i=1}^\infty$ and $\{r_i\}_{i=1}^\infty$ are dense in H_d

and H_r, respectively, then $\lim_{M \to \infty} \|A - P_{r^M} A P_{d^M}\|_{H_d}^{H_r} = 0$, where $\|\cdot\|_{H_d}^{H_r}$ denotes the operator norm of operators from H_d to H_r.

Proof. Let $\{A_N\}_{N=1}^\infty$ be a sequence of rank-N operators converging, in norm, to A. For an arbitrary $h \in H_d$,

$$\|Ah - P_{r^M} A P_{d^M} h\|_{H_r} \leq \|Ah - A_N h\|_{H_r}$$
$$+ \|A_N h - A_N P_{d^M} h\|_{H_r} + \|A_N P_{d^M} h - P_{r^M} A_N P_{d^M} h\|_{H_r}$$
$$+ \|P_{r^M} A_N P_{d^M} h - P_{r^M} A P_{d^M} h\|_{H_r}.$$

Using the fact that A_N, $P_{r^M} A_N$, and $A_N P_{d^M}$ are all finite-rank operators and the fact that the projection operator is bounded with norm bound 1, Proposition 5, can be used to conclude that for all $\epsilon > 0$, there exists $M(N) \in \mathbb{N}$ such that for all $i \geq M(N)$

$$\|Ah - P_{r^i} A P_{d^i} h\|_{H_r} \leq \|(A - A_N)h\|_{H_r} + 2\epsilon \|h\|_{H_d}$$
$$+ \|(A_N - A) P_{d^i} h\|_{H_r}.$$

Since A_N converges to A in norm, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $j \geq N$, and $g \in H_d$ $\|Ag - A_j g\|_H \leq \epsilon \|g\|_{H_d}$. Thus, for all $j \geq N$ and $i \geq M(j)$, $\|Ah - P_{r^i} A P_{d^i} h\|_{H_r} \leq 4\epsilon \|h\|_{H_d}$. $\square$

4.1 Matrix Representation of the Finite-rank Operator

To formulate a matrix representation of the finite-rank operator $P_{r^M} A_f P_{d^M}$, the operator is restricted to span d^M to yield the finite-rank operator $P_{r^M} A_f|_{d^M} : \text{span } d^M \to \text{span } r^M$. The resulting matrix representation is denoted by $[A_f]_d^r$. For brevity of exposition, the superscript M is suppressed hereafter and d and r are interpreted as $M-$dimensional vectors.

Theorem 2. Let $g = a^\top r \in \text{span } r$ and $h = \delta^\top d \in \text{span } d$ be functions with coefficients $a \in \mathbb{R}^M$ and $\delta \in \mathbb{R}^M$, respectively. If $g = P_r A_f|_d h$, then $a = G_r^+ G_d \delta$. That is, $[A_f]_d^r := G_r^+ G_d$, where G_r^+ denotes the Moore-Penrose pseudoinverse of G_r, is a matrix representation of $P_r A_f|_d$.

Proof. Since $g = P_r A_f|_d h = a^\top r$, the coefficients a solve

$$G_r a = \begin{bmatrix} \langle A_f \delta^\top d, r_1 \rangle_{H_r} \\ \vdots \\ \langle A_f \delta^\top d, r_M \rangle_{H_r} \end{bmatrix} = \begin{bmatrix} \langle \delta^\top d, A_f^* r_1 \rangle_{H_d} \\ \vdots \\ \langle \delta^\top d, A_f^* r_M \rangle_{H_d} \end{bmatrix}$$

Proposition 3 implies that $A_f^* r_i = d_i$ for all $i = 1, \cdots, M$, and as such, $G_r a = G_d \delta$. Since $a = G_r^+ G_d \delta$ is a solution of $G_r a = G_d \delta$, we have, $a = G_r^+ G_d \delta$, and as a result, $G_r^+ G_d$ is a matrix representation of $P_r A_f|_d$. $\square$

In the following section, the matrix representation $[A_f]_d^r$ is used to construct a data-driven representation of the singular values and the left and right singular functions of the operator $P_r A_f|_d$.

4.2 Singular Functions of the Finite-rank Operator

The tuples $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^M$, with $\sigma_i \in \mathbb{R}$, $\phi_i \in H_d$, and $\psi_i \in H_r$, are singular values, left singular vectors, and right singular vectors of $P_r A_f|_d$, respectively, if $\forall h \in \text{span } d$, $P_r A_f h = \sum_{i=1}^M \sigma_i \psi_i \langle h, \phi_i \rangle_{H_d}$. The following proposition states that the SVD of $P_r A_f|_d$ can be computed using

matrices in the matrix representation $[A_f]_d^r$ developed in the previous section.

Theorem 3. If (W, Σ, V) is the SVD of G_r^+ with $W = [w_1, \ldots, w_M]$, $V = [v_1, \ldots, v_M]$, and $\Sigma = \mathrm{diag}([\sigma_1, \ldots, \sigma_M])$, then for all $i = 1, \ldots, M$, σ_i are singular values of of $P_r A_f|_d$ with left singular functions $\phi_i := v_i^\top d$ and right singular functions $\psi_i := w_i^\top r$, respectively.

Proof. Let $\phi_i = v_i^\top d$ and $\psi_i = w_i^\top r$ and $h = \delta^\top d$. Then,

$$P_r A_f h = \sum_{i=1}^M \sigma_i \psi_i \langle h, \phi_i \rangle_{H_d} \iff$$

$$P_r A_f \delta^\top d = \sum_{i=1}^M \sigma_i w_i^\top r \langle \delta^\top d, v_i^\top d \rangle_{H_d}$$

Using the finite-rank representation, the collection $\{(\sigma_i, \phi_i, \psi_i)\}_{i=1}^M$, is a SVD of $P_r A_f|_d$, if for all $\delta \in \mathbb{R}^M$,

$$\left(G_r^+ G_d \delta \right)^\top r = \left(\sum_{i=1}^M \sigma_i \langle \delta^\top d, v_i^\top d \rangle_{H_d} w_i^\top \right) r. \quad (5)$$

Simple matrix manipulations yield the chain of implications

$$(5) \impliedby \forall \delta \in \mathbb{R}^M, G_r^+ G_d \delta = \sum_{i=1}^M \sigma_i \langle \delta^\top d, v_i^\top d \rangle_{H_d} w_i$$

$$\iff \forall \delta \in \mathbb{R}^M, G_r^+ G_d \delta = \sum_{i=1}^M \sigma_i \left(w_i v_i^\top G_d \right) \delta$$

$$\impliedby G_r^+ G_d = \sum_{i=1}^M \sigma_i \left(w_i v_i^\top \right) G_d$$

$$\impliedby G_r^+ = \sum_{i=1}^M \sigma_i w_i v_i^\top = W \Sigma V^\top,$$

which proves the proposition. $\qquad\square$

In the following section, the singular values and the left and right singular vectors are used to generate a data-driven model via a method termed singular Liouville dynamic mode decomposition (SLDMD).

5. THE SLDMD ALGORITHM

Let the identity observable $(h_{\mathrm{id}})_j$ be defined as $(h_{\mathrm{id}})_j(x) = x_j$, where x_j is the j−th component of x.

Theorem 4. If $(h_{\mathrm{id}})_j \in H_d$ for $j = 1, \cdots, n$, then

$$\hat{f}_M(x) := \begin{bmatrix} [P_r A_f P_d (h_{\mathrm{id}})_1](x) \\ \vdots \\ [P_r A_f P_d (h_{\mathrm{id}})_n](x) \end{bmatrix} = DG_r^+ r(x) \quad (6)$$

where $D := \left((\gamma_j(T_j))_i - (\gamma_j(0))_i \right)_{i,j=1}^{n,M}$ and $(\gamma_j(\cdot))_i$ denotes the i−th component of $\gamma_j(\cdot)$.

Proof. Using Theorem 3 and the definition of singular values and singular functions of $P_r A_f|_d$,

$$\hat{f}_M(x) = \sum_{i=1}^M \sigma_i \xi_i w_i^\top r(x) = \xi \Sigma W^\top r(x), \quad (7)$$

where $\xi_i := \left[\langle P_d(h_{\mathrm{id}})_1, \phi_i \rangle_{H_d} \ \ldots, \langle P_d(h_{\mathrm{id}})_n, \phi_i \rangle_{H_d} \right]^\top$ and $\xi := [\xi_1 \ \cdots \ \xi_M]$.

The modes ξ can be computed using $\phi_i = v_i^\top d$ as

$$\xi = \begin{bmatrix} \langle P_d(h_{\mathrm{id}})_1, v_1^\top d \rangle_{H_d}, & \ldots, & \langle P_d(h_{\mathrm{id}})_1, v_M^\top d \rangle_{H_d} \\ \vdots & \ddots & \vdots \\ \langle P_d(h_{\mathrm{id}})_n, v_1^\top d \rangle_{H_d}, & \ldots, & \langle P_d(h_{\mathrm{id}})_n, v_M^\top d \rangle_{H_d} \end{bmatrix}$$

$$= \begin{bmatrix} \langle \delta_1^\top d, d_1 \rangle_{H_d}, & \ldots, & \langle \delta_1^\top d, d_M \rangle_{H_d} \\ \vdots & \ddots & \vdots \\ \langle \delta_n^\top d, d_1 \rangle_{H_d}, & \ldots, & \langle \delta_n^\top d, d_M \rangle_{H_d} \end{bmatrix} V = \delta^\top G_d V,$$

where $\delta := [\delta_1, \ldots, \delta_n]$. Using the reproducing property of the reproducing kernel of H_d, the coefficients δ_i in the projection of $(h_{\mathrm{id}})_i$ onto d solve the system of linear equations

$$G_d \delta_i = \begin{bmatrix} \langle (h_{\mathrm{id}})_i, d_1 \rangle_{H_d} \\ \vdots \\ \langle (h_{\mathrm{id}})_i, d_M \rangle_{H_d} \end{bmatrix} = \begin{bmatrix} (\gamma_1(T_1))_i - (\gamma_1(0))_i \\ \vdots \\ (\gamma_M(T_M))_i - (\gamma_M(0))_i \end{bmatrix}.$$

Letting $D := \left((\gamma_j(T_j))_i - (\gamma_j(0))_i \right)_{i,j=1}^{n,M}$ it can be concluded that $\delta^\top G_d = D$. Finally, the modes ξ are given by $\xi = DV$ and

$$\hat{f}_M(x) = DV\Sigma W^\top r(x) = DG_r^+ r(x). \qquad \square$$

Using $\hat{f}_M$ as a rank−M estimate of f, the estimated system model is of the form $\dot{x} \approx \hat{f}_M(x) = Ar(x)$, where $A \in \mathbb{R}^{n \times M}$ is a solution of

$$AG_r = D. \quad (8)$$

The use of $\hat{f}_M$ as an estimate of f is justified by the following result, which follows from the fact that $P_r A_f P_d$ converges to A_f in norm as $M \to \infty$.

Theorem 5. If $\mu_r < \mu_d$, $H_d = F_{\mu_d}^2(\mathbb{R}^n)$, $H_r = F_{\mu_r}^2(\mathbb{R}^n)$, $f : \mathbb{R}^n \to \mathbb{R}^n$ is a component-wise polynomial function, and the spans of the collections $\{d_i\}_{i=1}^\infty$ and $\{r_i\}_{i=1}^\infty$ are dense in H_d and H_r, respectively, then $\lim_{M \to \infty} \left(\sup_{x \in X} \left\| \hat{f}_M(x) - f(x) \right\|_2 \right) = 0$.

Proof. Since the space $F_{\mu_d}^2(\mathbb{R}^n)$ contains $(h_{\mathrm{id}})_j$ for $j = 1, \cdots, n$, the functions $\hat{f}_{M,j} := P_r A_f P_d (h_{\mathrm{id}})_j$ and $f_j := A_f(h_{\mathrm{id}})_j$ that denote the j−th row of $\hat{f}_M$ and f, respectively, exist as members of H_r. Since $x \mapsto K_r(x, x) = \exp\left(\frac{x^\top x}{\mu_r} \right)$ is continuous and X is compact, there exists a real number $\overline{K}$ such that $\sup_{x \in X} K_r(x, x) = \overline{K}$. Since $\mu_r < \mu_d$, Theorem 1 implies that $A_f : F_{\mu_d}^2(\mathbb{R}^n) \to F_{\mu_r}^2(\mathbb{R}^n)$ is compact. Proposition 6 can then be used to conclude that for all $\epsilon > 0$ and $j = 1, \ldots, n$, there exists $M(j) \in \mathbb{N}$ such that for all $i \geq M(j)$, $\left\| \hat{f}_{i,j} - f_j \right\|_{H_r}^2 \leq \frac{\epsilon^2}{n\overline{K}}$. Using the reproducing property, for $i \geq \overline{M} := \max_j M(j)$,

$$\left\| \hat{f}_i(x) - f(x) \right\|_2^2 = \sum_{j=1}^n \left\langle \left(\hat{f}_{i,j} - f_j \right), K_r(\cdot, x) \right\rangle_{H_r}^2$$

$$\leq \sum_{j=1}^n \left\| \hat{f}_{i,j} - f_j \right\|_{H_r}^2 \| K_r(\cdot, x) \|_{H_r}^2$$

$$\leq \sum_{j=1}^n \frac{\epsilon^2}{n\overline{K}} \langle K_r(\cdot, x), K_r(\cdot, x) \rangle_{H_r}^2 = \frac{\epsilon^2}{\overline{K}} K_r(x, x)^2.$$

As a result, for all $\epsilon \geq 0$ there exists $\overline{M}$ such that for all $i \geq \overline{M}$,

$$\sup_{x \in X} \left\| \hat{f}_i(x) - f(x) \right\|_2 \leq \sqrt{\frac{\epsilon^2}{\overline{K}^2} \sup_{x \in X} K_r(x,x)^2} = \epsilon,$$

which completes the proof. $\qquad\square$

This convergence result, typically unattainable for Koopman-based dynamic mode decomposition, is a key feature of this algorithm.

6. OCCUPATION KERNEL REGRESSION

Another approach to modeling is to directly use the occupation kernels as basis functions for regression.

If the components of the vector field f reside in the RKHS, i.e., $f = [f_1 \ \ldots \ f_n]^\top$, with $f_i \in H$ for $i = 1, \ldots, n$, then using the defining characteristic of the occupation kernel, the inner product $\langle f_i, \Gamma_\gamma \rangle_H$ may be expressed as

$$\langle f_i, \Gamma_\gamma \rangle_H = \int_0^T f_i(\gamma(t))\mathrm{d}t = (\gamma(T))_i - (\gamma(0))_i$$

for each $i = 1, \ldots, n$. As such, given a set of solutions $\{\gamma_j : [0, T_j] \rightarrow \mathbb{R}^n\}_{j=1}^M$, of $\dot{x} = f(x)$, components of the function f can be estimated by minimizing the error between the inner products $\langle f_i, \Gamma_\gamma \rangle_H$ and the displacement $(\gamma(T))_i - (\gamma(0))_i$ of the i-th component of the trajectory.

A regularized regression problem to determine an approximation $\hat{f}_i$ of the i-th row of f can thus be formulated as

$$\min_{\hat{f}_i \in H} \sum_{j=1}^M \left(\langle \hat{f}_i, \Gamma_{\gamma_j} \rangle_H - (\gamma_j(T_j))_i - (\gamma_j(0))_i \right)^2 + \lambda \|\hat{f}_i\|_H^2, \tag{9}$$

where $\lambda > 0$ is a regularization parameter. Using the Represeter theorem for occupation kernels (Li and Rosenfeld, 2020, Proposition 1), the minimizer of (9) may be expressed as a linear combination of occupation kernels

$$\hat{f}_i = A_{1,i}\Gamma_{\gamma_1} + \cdots + A_{M,i}\Gamma_{\gamma_M} = A_i r(x),$$

where $A_i := [A_{1,i} \ \ldots \ A_{M,i}]$. Using this representation of the minimizer, the inner products and the norm in the optimization problem can be computed as

$$\langle \hat{f}_i, \Gamma_{\gamma_j} \rangle_H = \left\langle \sum_{k=1}^M A_{k,i}\Gamma_{\gamma_k}, \Gamma_{\gamma_j} \right\rangle_H = (G_r^j)^\top A_i^\top$$

and

$$\|\hat{f}_i\|_H^2 = A_i G_r A_i^\top,$$

where G_r^j denotes the j-th column of G_r.

Hence, the resolution (9) reduces to the finite dimensional convex optimization problem

$$\min_{A_i \in \mathbb{R}^{1 \times M}} \sum_{j=1}^M \left((G_r^j)^\top A_i^\top - (\gamma_j(T_j) - \gamma_j(0))_i \right)^2 + \lambda A_i G_r A_i^\top. \tag{10}$$

Solutions of the finite dimensional optimization problem coincide with solutions of the linear system $A_i(G_r + \lambda I_M)G_r = D_i G_r$, where D_i is the i-th row of D and I_M denotes the $M \times M$ identity matrix.

Using the fact that the solution of $A_i(G_r + \lambda I_M) = D_i$ is one of the solutions of $A_i(G_r + \lambda I_M)G_r = D_i G_r$ (the
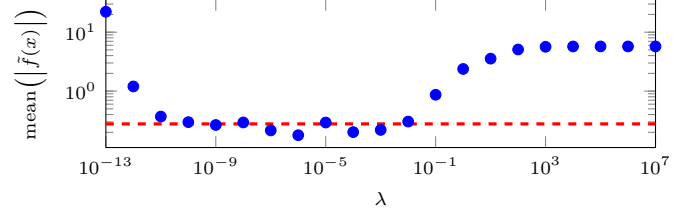

Fig. 1. The blue marks are the $\text{mean}\left(\left| \tilde{f}(x) \right| \right)$ over $x \in [-3, 3]$ for different values of λ using OKR trained with noisy trajectories. the dashed red line is the $\text{mean}\left(\left| \tilde{f}(x) \right| \right)$ over $x \in [-3, 3]$ calculated using the SLDMD method trained with noisy trajectories.

only one if G_r is nonsingular), the estimated model can be expressed as $\dot{x} \approx Ar(x)$, where $A \in \mathbb{R}^{n \times M}$ is the solution of

$$A(G_r + \lambda I_M) = D. \tag{11}$$

Interestingly, the model obtained via OKR coincides with that obtained using DMD if the regularization parameter is set to zero. As such, if occupation kernels are used for regression, the resulting model encodes additional structure, i.e., the singular functions and the singular values of the underlying Liouville operator, as provided by SLDMD. Furthermore, theorem 5, also applicable to the regression model without regularization, provides an alternative to the cumulative prediction error convergence guarantees (see, for example Zhdanov and Kalnishkan (2013) and Chapter 6 of Steinwart and Christmann (2008)) that are typically available for regression problems.

7. NUMERICAL EXPERIMENTS

The purpose of the numerical experiment is to demonstrate the effectiveness of the SLDMD algorithm, the OKR algorithm, and to compare the two methods. Additionally, the numerical experiment demonstrates the effects of the regularization constant λ on the OKR method.

This experiment utilizes the nonlinear model of the Duffing oscillator given by

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= (x_1 - x_1^3). \end{aligned} \tag{12}$$

To approximate the system dynamics for any given value of $\lambda > 0$, 169 trajectories of the system are recorded starting from a grid of initial conditions. In the first trial, each trajectory is corrupted by Gaussian measurement noise with standard deviation 0.001. The initial velocities are obtained by numerically differentiating the measured noisy trajectories. In the second trial, the trajectories are not corrupted by noise, and the initial velocities are similarly obtained by numerically differentiating the measured trajectories. The recorded trajectories are then utilized to approximate f. The SLDMD algorithm from Section 5 is implemented using the Moore-Penrose pseudoinverse. The OKR algorithm from Section 6 is implemented using the usual matrix inverse. Both methods use the kernel $K(x, y) = \exp\left(\frac{x^\top y}{\mu} \right)$ with $\mu = 5$.
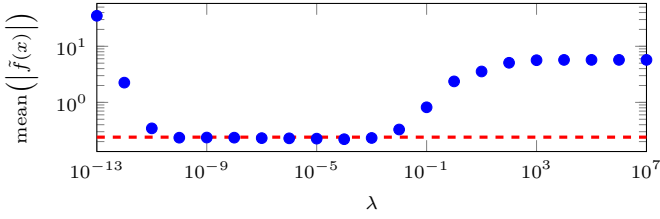
Fig. 2. The blue marks are the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3,3]$ for different values of λ using OKR trained with noise free trajectories. the dashed red line is the mean$\left(\left|\tilde{f}(x)\right|\right)$ over $x \in [-3,3]$ calculated using the SLDMD method trained with noise free trajectories.

Figures 1 and 2 show the mean of the difference between the actual values $f_2(x_1)$ and the estimated values $\hat{f}_2(x_1)$ over $x_2 \in [-3,3]$ using different values of λ. Figure 1 shows the results for the system identification methods using noisy trajectories, where OKR clearly outperforms SLDMD for specific values of λ. Whereas, in Figure 2, which shows the results of the two system identification methods using noise free trajectories, the OKR technique performs almost identically to the SLDMD technique for λ values between 10^{-10} and 10^{-3}.

If the Gram matrix is not full rank, which is the case for both trials of the numerical experiments, then inversion of $G_r + \lambda I_M$ is numerically unstable for small values of λ. As such, consistent with Figures 1 and 2, it is expected that the approximation $\hat{f}$ computed using OKR with a small value of λ would be poor. On the other hand, the regularized inverse $(G_r + \lambda I_M)^{-1}$ converges to the zero matrix as λ increases, which explains the error plateau seen in Figures 1 and 2 for $\lambda > 10^2$.

In theory, in the case where the trajectories are noise free and the inner products in the Gram matrix G_r are evaluated exactly, the SLDMD algorithm should outperform OKR for any value of λ, since the regularization introduces a bias. The authors speculate that the slight improvement, seen in Figures 1 and 2 for $10^{-10} < \lambda < 10^{-3}$ is due to the integration errors introduced when computing the entries of the Gram matrix using Simpson's rule. Since regularization can prevent over-fitting when the underlying data are noisy, a more accurate approximation may be obtained for some values of λ. In the case where the trajectories are corrupted with measurement noise, there are errors in both G_r and D of (11), which make the effects of λ even more significant, consistent with Figure 1.

8. CONCLUSION

This paper introduces a novel approach towards the construction of a finite rank representation of the Liouville operator. New results on the construction of singular values and functions of the finite rank operator using singular values and vectors of a matrix representation are also obtained. The singular values and functions give rise to a new dynamic mode decomposition model that is shown to be equivalent to regression using occupation kernels.

Numerical experiments that study the effect of the regularization parameter indicate that regularization may yield better results when the data are corrupted by integration errors and measurement noise. In order to support this conjecture a detailed Monte Carlo simulation is needed, which is a part of future work.

REFERENCES

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

Budišić, M., Mohr, R., and Mezić, I. (2012). Applied Koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4), 047510.

Das, S., Giannakis, D., and Slawinska, J. (2021). Reproducing kernel Hilbert space compactification of unitary evolution groups. *Appl. Comput. Harmon. Anal.*, 54, 75–136.

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

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

Li, X. and Rosenfeld, J.A. (2020). Fractional order system identification with occupation kernel regression. *IEEE Control Syst. Lett.*, 6, 19–24.

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

Pedersen, G.K. (2012). *Analysis now*, volume 118. Springer Science & Business Media.

Roman, S. (2008). *Advanced linear algebra*, volume 135 of *Graduate Texts in Mathematics*. Springer New York, third edition edition.

Rosenfeld, J.A. and Kamalapurkar, R. (2023). Singular dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.*, 22(3), 2357–2381.

Rosenfeld, J.A., Kamalapurkar, R., and Dixon, W.E. (2019a). The state following approximation method. *IEEE Trans. Neural Netw. Learn. Syst.*, 30(6), 1716–1730.

Rosenfeld, J.A., Kamalapurkar, R., Gruss, L.F., and Johnson, T.T. (2022). Dynamic mode decomposition for continuous time systems with the Liouville operator. *J. Nonlinear Sci.*, 32(1), 1–30.

Rosenfeld, J.A., Kamalapurkar, R., Russo, B., and Johnson, T.T. (2019b). Occupation kernels and densely defined Liouville operators for system identification. In *Proc. IEEE Conf. Decis. Control*, 6455–6460.

Rosenfeld, J.A., Russo, B., Kamalapurkar, R., and Johnson, T. (2024). The occupation kernel method for nonlinear system identification. *SIAM J. Control Optim.*, 62(3), 1643–1668. See arXiv:1909.11792.

Steinwart, I. and Christmann, A. (2008). *Support vector machines.* Information Science and Statistics. Springer, New York.

Wendland, H. (2004). *Scattered data approximation.* Cambridge University Press.

Zhdanov, F. and Kalnishkan, Y. (2013). An identity for kernel ridge regression. *Theor. Comput. Sci.*, 473, 157–178.