Dynamic mode decomposition of higher order systems *

Joel A. Rosenfeld [a], Benjamin P. Russo [b], Rushikesh Kamalapurkar [c]

[a] *Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA*

[b] *Riverside Research, New York, NY 10038, USA.*

[c] *Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA*

Abstract

Conventionally, data driven identification and control problems for higher order dynamical systems are solved by augmenting the system state by the derivatives of the output to formulate first order dynamical systems in higher dimensions. However, solution of the augmented problem typically requires knowledge of the full augmented state, which requires numerical differentiation of the original output, frequently resulting in noisy signals. This manuscript develops the theory necessary for a direct analysis of higher order dynamical systems using higher order Liouville operators. Fundamental to this theoretical development is the introduction of signal valued RKHSs and new operators posed over these spaces. Ultimately, it is observed that despite the added abstractions, the necessary computations are remarkably similar to that of first order DMD methods using occupation kernels.

Key words: operator theory, system identification, dynamic mode decomposition, model fitting, Koopman operator

1 Introduction

Data driven methods for dynamical systems have developed significantly over the past 20 years (cf. [2, 7, 13,

14, 17]). Principal among them are those that leverage Koopman operators (also known as composition operators) over Hilbert function spaces to give a representation of finite dimensional discrete time dynamics as an operator over an infinite dimensional Hilbert space [2, 31]. When a continuous time dynamical system is forward complete, it may be discretized by fixing a timestep, $\Delta t > 0$, to yield a discrete time system. In this setting, the semigroup of Koopman operators has been demonstrated as an effective tool for extracting the underlying governing principles of a dynamical system, and for providing a model for the state which performs well over short time horizons via dynamic mode decomposition (DMD). However, many dynamical systems of interest are not forward complete (e.g., models of the form $\dot{x} = 1 + x^2$, often encountered in mass-action kinetics in thermodynamics [5, Section 6.3], chemical reactions [29, Section 8.4], and species populations [6, Section 4.2]), and consequently, the breadth of applicability of Koopman based methods is limited.

In [21], the concept of Liouville operators and occupation kernels was introduced. Given a signal $\gamma : [0, T] \to \mathbb{R}^n$, occupation kernels are functions Γ_γ inside of a reproducing kernel Hilbert space (RKHS) that represent the functional $g \mapsto \int_0^T g(\gamma(t))\mathrm{d}t$. Occupation kernels generalize the concept of an occupation measure (cf. [8]) by

changing the setting from a collection of measures to that of a Hilbert space. Thus, occupation kernels can be leveraged as a basis in a Hilbert space for function approximation and projections (cf. [11, 19, 22, 24]).

Liouville operators extend the idea of Koopman generators to dynamical system that do not need to be forward complete. Significant to the problem at hand, if the Liouville operator, A_f, is densely defined over an RKHS over $\mathbb{R}^n$ of continuously differentiable functions, and $\gamma : [0, T] \to \mathbb{R}^n$ is a trajectory satisfying $\dot{\gamma} = f(\gamma)$, then $\langle A_f g, \Gamma_\gamma \rangle_H = g(\gamma(T)) - g(\gamma(0)) = \langle g, K_{\gamma(T)} - K_{\gamma(0)} \rangle_H$. Hence, $A_f^* \Gamma_\gamma = K_{\gamma(T)} - K_{\gamma(0)}$, where $K_x := K(\cdot, x)$, and $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ denotes the reproducing kernel of the RKHS. This relationship between A_f and K allows for the construction of a finite rank representation of A_f, and subsequently, a dynamic mode decomposition of the system, using the trajectories of the system as the central unit of information.

The advantage gained by taking the Liouville operator/occupation kernel perspective is twofold. The dynamics are now free from the requirements of forward completeness, and instead, the dynamics are required to admit a densely defined Liouville operator over a given RKHS. Different selections of RKHSs yield new classes of dynamics that meet these requirements, which adds considerably to the flexibility of the resulting methods. The other advantage is that by the nature of occupation kernels as representatives of integration, they take the form of $\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t)) \mathrm{d}t$. Thus, Γ_γ interprets the trajectory data through an integral and is robust to sensor noise [22].

One limitation still present in the theory of data driven methods for dynamical systems is that of high order dynamics. Conventionally in systems theory, higher order dynamics are converted to first order systems of augmented state variables. For example, $\ddot{x} = f(x)$ can be adjusted to $z := \begin{pmatrix} x & \dot{x} \end{pmatrix}^\top$ with $\dot{z} = \begin{pmatrix} z_2 & f(z_1) \end{pmatrix}^\top$. Theoretically, the augmentation is well justified, but it is computationally problematic in data driven methods. To estimate the new state variable of z, data driven methods must compute an approximation of the first derivative of x. Numerical derivatives can be noisy, and if the order of the system exceeds 2, they are unreliable. For example, in [1], where numerical differentiation is used for parameter fitting, a considerable amount of filtering was required to get good results. The sensitivity of numerical differentiation techniques to noise motivates the development of methods that avoid numerical differentiation altogether.

This manuscript introduces the necessary theoretical components for the development of a DMD routine for higher order dynamical systems that avoids the use of numerical derivatives and augmented state variables.

The method developed here is is not related to the higher order DMD methods in results such as [9, 10, 30]. Higher order DMD applies to discrete time systems where the snapshot at time k is expressed as a linear combination of snapshots at times $k - 1, \ldots, k - m$ for some order m. Higher order DMD methods can be applied to higher order dynamical systems via discretization; however, unlike the developed technique, measurements of derivatives of the output are needed for implementation.

Underlying the subsequent development are vector valued Reproducing Kernel Hilbert spaces (vvRKHSs), for which the relevant theory is presented in Section 2. Using vvRKHSs as a tool, Section 4 develops a signal valued RKHS, which is a Hilbert space of functions that map d times continuously differentiable signals to a scalar valued RKHS over $[0, T]$. The signal valued RKHS framework allows for the formulation of well defined second order Liouville operators over the Hilbert space beyond the formal expression given in Section 6. Once the essential elements are established, Section 8 presents a second order DMD (SoDMD) method for the modeling of a second order dynamical system, which avoids the use of numerical derivatives. Implementation related details are provided in Section 9, and prediction of system trajectories is discussed in Section 10. Three numerical experiments are presented and discussed in Sections 11 and 12, respectively, and Section 13 contains concluding remarks. A path towards extension of the method to systems of order higher than two is provided in the Appendix.

2 Vector Valued Reproducing Kernel Hilbert Spaces

This section presents the concept of vector valued RKHSs, which recently came to prominence with [3], though their roots extend further back (e.g. [16]).

Definition 1 *Given a Hilbert space $\mathcal{Y}$ and a set X, a vector valued reproducing kernel Hilbert space, H, is a Hilbert space of functions mapping X to $\mathcal{Y}$, where for each $x \in X$ and $v \in \mathcal{Y}$ the functional $g \mapsto \langle g(x), v \rangle_{\mathcal{Y}}$ is bounded.*

The Riesz representation theorem guarantees for each $x \in X$ and $v \in \mathcal{Y}$ the existence of a function $K_{x,v} \in H$ such that $\langle g, K_{x,v} \rangle_H = \langle g(x), v \rangle_{\mathcal{Y}}$ for all $g \in H$. It is readily apparent that the map $K_x : \mathcal{Y} \to H$, that maps v to $K_{x,v}$, is linear, and as such, is expressed as $K_x v := K_{x,v}$. It should be noted that for each fixed $v \in \mathcal{Y}$, the collection of functions $\{h : h(x) = \langle g(x), v \rangle_{\mathcal{Y}}\}$ is a scalar valued RKHS. In the context of this manuscript the Hilbert space $\mathcal{Y}$ will be a scalar valued RKHS, which will facilitate the description of a function space on signals.

3 Problem Statement

Consider a collection of trajectories, $\{\gamma_i : [0,T] \to \mathbb{R}^n\}_{i=1}^M$, corresponding to an m-th order dynamical system $\gamma^{(m)} = f(\gamma)$, starting from known initial conditions $\gamma^{(j)}(0) = \gamma_0^j$, $j = 1, \cdots, m$ where the superscript (m) denotes the m-th order time-derivative, the order $m \in \mathbb{N}$ is known, and the function f is unknown. The objective is to build a predictive model to estimate trajectories starting from given initial conditions $x^{(j)}(0) = x_0^j$ for $j = 1, \cdots, m$.

The exposition in this paper will be focused on second order dynamical systems, however, the developed methods may be readily adapted to higher order dynamical systems using the higher order Liouville operator formulated in the Appendix. In the following, the model is constructed from a finite rank representation of a second order Liouville operator, B_f, obtained via adjoint relations between the operator B_f, and a collection of second order occupation kernels. The class of functions f that are amenable to approximation using the developed technique is defined indirectly through restrictions placed on the operator B_f, see Assumption 9.

4 Signal valued RKHSs

The objective of this section is to provide a definition of *signal valued* RKHSs that act on a set X of signals. These Hilbert spaces consist of functions that map from signals in X to signals in an RKHS $\mathcal{Y}([0,T], \mathbb{C})$. Signal valued RKHSs treat a signal from X as the fundamental unit of information, and as such, allow for the treatment of operators acting on functions of signals, which include higher order Liouville operators. Signal valued RKHSs are a special case of vector valued RKHSs, and arise from a mapping on scalar valued RKHSs.

In the following, three different RKHSs are under consideration (see Fig. 1). The range space, $\mathcal{Y}([0,T], \mathbb{C})$, is selected to be the complexified native RKHS of a real-valued reproducing kernel $\mathscr{K} : [0,T] \times [0,T] \to \mathbb{R}$. To construct a vvRKHS of functions that map signals from X to $\mathcal{Y}$, we define an auxiliary scalar valued RKHS, $\tilde{H}$, consisting of functions from $\mathbb{R}^n$ to $\mathbb{C}$. For each $g \in \tilde{H}$, a map from X to $\mathcal{Y}$ is obtained as $\phi_g[\theta](t) := g(\theta(t))$ for all $\theta \in X$ and $t \in [0,T]$. Theorem 2 shows that the space of all such maps is a signal-valued RKHS, denoted by H.

Theorem 2 *Let $X = C^d([0,T], \mathbb{R}^n)$ for some $d \in \mathbb{N}$. Let $\tilde{H}$ be a scalar valued RKHS over $\mathbb{R}^n$. Suppose there exists an RKHS $\mathcal{Y}$ over $[0,T]$ where the composition operator $C_\theta : \tilde{H} \to \mathcal{Y}$ is a bounded operator for all symbols $\theta \in X$. Define the vector space $H := \{\phi_g : g \in \tilde{H}\}$ of mappings $\phi_g : X \to \mathcal{Y}$ given by $\phi_g[\theta] := g(\theta(\cdot))$, together with the inner product induced by $\tilde{H}$, $\langle \phi_{g_1}, \phi_{g_2} \rangle_H = \langle g_1, g_2 \rangle_{\tilde{H}}$. Then H is a vvRKHS.*
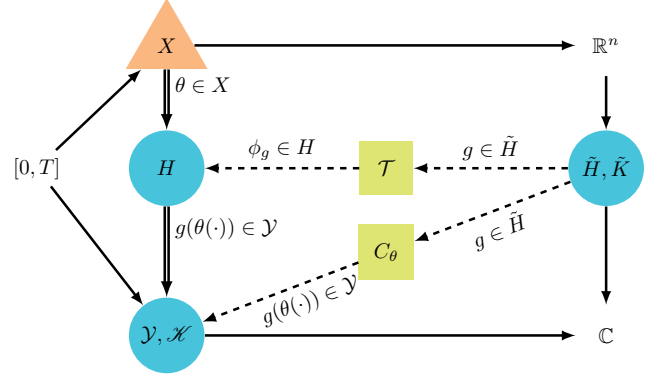

Fig. 1. A visualization of relationships between vector spaces and operators defined in Theorem 2. Symbols in blue filled circles denote RKHSs, with range space and domain defined by incoming and outgoing arrows. Symbols in orange filled triangles are sets and symbols in green filled squares are operators with operands and results of the operation denoted by incoming and outgoing dotted arrows.

PROOF. Let $\mathcal{T}$ be the mapping from $\tilde{H}$ to H given as $\mathcal{T}g = \phi_g$. It is immediate that $\mathcal{T}$ is an injection into H. Indeed if $g_1 \neq g_2$ for g_1 and g_2 in $\tilde{H}$, then there is a point $\omega \in \mathbb{R}^n$ such that $g_1(\omega) \neq g_2(\omega)$. Letting $\theta_\omega \in C^d([0,T], \mathbb{R}^n)$ denote the constant signal $\theta_\omega(t) \equiv \omega$, then $\phi_{g_1}(\theta_\omega) \neq \phi_{g_2}(\theta_\omega)$. Selecting the inner product on H to be $\langle \phi_{g_1}, \phi_{g_2} \rangle_H = \langle g_1, g_2 \rangle_{\tilde{H}}$, H can be seen to be complete with respect to the induced norm. Since $\tilde{H}$ is a Hilbert space, it is complete by definition. Hence, a Cauchy sequence in H, say $\{\phi_{g_n}\}$ gives rise to a Cauchy sequence in $\tilde{H}$, namely $\{g_n\}$. Since $\tilde{H}$ is complete, the sequence converges to a $g \in \tilde{H}$ and thus ϕ_{g_n} converges to ϕ_g. Since $H = \mathcal{T}(\tilde{H})$, $\mathcal{T}$ is surjective. What is left to resolve is to demonstrate that for any $\theta \in C^d([0,T], \mathbb{R}^n)$ and $v \in \mathcal{Y}$, the functional $\psi \mapsto \langle \psi[\theta], v \rangle_\mathcal{Y}$ is bounded. Note that there is a $g \in \tilde{H}$ such that $\mathcal{T}g = \psi$ and $\|\psi\|_H = \|g\|_{\tilde{H}}$. Now consider $|\langle \psi[\theta], v \rangle_\mathcal{Y}| = |\langle g \circ \theta, v \rangle_\mathcal{Y}| \leq \|g \circ \theta\|_\mathcal{Y} \|v\|_\mathcal{Y} \leq \|C_\theta\| \|g\|_{\tilde{H}} \|v\|_\mathcal{Y} = \|C_\theta\| \|\psi\|_H \|v\|_\mathcal{Y}$.

Definition 3 *The vvRKHS, H, given in Theorem 2 will be called the signal valued RKHS from $C^d([0,T], \mathbb{R}^n)$ to $\mathcal{Y}$ derived from $\tilde{H}$, more succinctly a signal valued RKHS, when the other quantities are understood from context.*

While a general characterization of tuples of function spaces $(X, \tilde{H}, \mathcal{Y})$ that admit bounded composition operators $C_\theta : \tilde{H} \to \mathcal{Y}$ is difficult, the following example analyzes one such tuple.

Example 4 *A possible choice for $\tilde{H}$ would be the native RKHS of a positive definite kernel $\tilde{K} \in \mathcal{C}^1(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$, for example, the Gaussian RBF reproducing kernel, $\tilde{K}(x,y) = \exp\left(-\frac{\|x-y\|_2^2}{\mu}\right)$ or the exponential dot product kernel $\tilde{K}(x,y) = \exp\left(\frac{x^\top y}{\mu}\right)$. Let $X = C^1([0,T], \mathbb{R}^n)$ and let $\mathcal{Y}$ be the Sobolev space $H^1([0,T], \mathbb{R})$.*

3

Corollary 5 *Example 4 provides a signal valued RKHS of functions from $C^1([0,T],\mathbb{R}^n)$ to $H^1([0,T],\mathbb{R})$.*

PROOF. Let $\theta \in C^1([0,T],\mathbb{R}^n)$ and $g \in \tilde{H}$. The norm on $\mathcal{Y}$, which consists of absolutely continuous functions over $[0,T]$, is given as $\|v\|_{H^1}^2 = \int_0^T |v(t)|^2 + |v'(t)|^2 \mathrm{d}t$ (cf. [18]). Consider the quantity $z = \int_0^T |[g \circ \theta](t)|^2 + |[g \circ \theta]'(t)|^2 \mathrm{d}t$. Note that for each t, $[g \circ \theta](t) = g(\theta(t)) = \langle g, \tilde{K}_{\theta(t)}\rangle_{\tilde{H}}$ and $[g \circ \theta]'(t) = \nabla g(\theta(t))\dot{\theta}(t) = \langle g, \nabla_2 \tilde{K}(\cdot,\theta(t))\dot{\theta}(t)\rangle_{\tilde{H}}$ (see [26, p. 132]), where ∇_2 denotes gradient with respect to the second argument and $\tilde{K}_y(x) := \tilde{K}(x,y)$, for $x,y \in \mathbb{R}^n$. Hence, via the Cauchy Schwarz inequality, $z \leq \|g\|_{\tilde{H}}^2 \int_0^T \|\tilde{K}_{\theta(t)}\|_{\tilde{H}}^2 + \|\nabla_2 \tilde{K}(\cdot,\theta(t))\dot{\theta}(t)\|_{\tilde{H}}^2 \mathrm{d}t$. Note that $\|\tilde{K}_{\theta(t)}\|_{\tilde{H}} = \sqrt{\tilde{K}(\theta(t),\theta(t))}$ and $\tilde{K} \in \mathcal{C}^1(\mathbb{R}^n \times \mathbb{R}^n,\mathbb{R})$. Hence, $\|\tilde{K}_{\theta(t)}\|_{\tilde{H}}$ is continuous, and as θ is continuous and $[0,T]$ is compact, $\int_0^T \|\tilde{K}_{\theta(t)}\|_{\tilde{H}}^2 \mathrm{d}t$ is bounded. A similar argument reveals that continuity of $\dot{\theta}$ and $\nabla \tilde{K}$ implies that $\int_0^T \|\nabla_2 \tilde{K}(\cdot,\theta(t))\dot{\theta}(t)\|_{\tilde{H}}^2 \mathrm{d}t$ is bounded. As a result, $g \circ \theta \in \mathcal{Y}$ and $z = \|g \circ \theta\|_{H^1}^2$. Thus the composition operator $g \mapsto g \circ \theta$ is bounded from $\tilde{H}$ to $\mathcal{Y}$.

Remark 6 *A similar argument can be used to show that the tuples $\left(X, \tilde{H}, H_{\mathbb{C}}^k([0,T],\mathbb{C})\right)$ also give rise to a vvRKHS for any $k \geq 1$ provided the functions in X and $\tilde{H}$ are at least $k-$times continuously differentiable, where $H_{\mathbb{C}}^k([0,T],\mathbb{C}) := \{f + \mathrm{i}h \mid f,h \in H^k([0,T],\mathbb{R})\}$ denotes the complexification of $H^k([0,T],\mathbb{R})$. For $g = f + \mathrm{i}h \in \mathcal{Y}$, we use the notation ∇g to denote $\nabla f + i\nabla h$.*

In this paper, unless otherwise specified, we select X to be $C^d([0,T],\mathbb{R}^n)$ for a sufficiently large $d \in \mathbb{N}$ (see Remark 6), $\tilde{H}$ to be the native space of a kernel $\tilde{K} \in C^\infty([0,T] \times [0,T],\mathbb{R})$, and $\mathcal{Y}$ to be the complexified Sobolev space $H_{\mathbb{C}}^3([0,T],\mathbb{C})$. Using the general Sobolev inequality [4, Section 5.6.3, Theorem 6], $y \in \mathcal{Y}$ implies $t \mapsto \Re(y(t)) \in \mathcal{C}^2([0,T],\mathbb{R})$ and $t \mapsto \Im(y(t)) \in \mathcal{C}^2([0,T],\mathbb{R})$.

5 Higher order occupation kernels

DMD relies on the action of Liouville operators on occupation kernels. As such, it is necessary to define second order occupation kernels in the context of vvRKHSs of the form in Theorem 2. To motivate the definition of higher order occupation kernels, recall Cauchy's formula for iterated integrals given as $h^{(-m)}(T) = \frac{1}{(m-1)!}\int_0^T (T-t)^{m-1}h(t)\mathrm{d}t$, where $h^{(-m)}(t) := \int_0^t \int_0^{\tau_1} \cdots \int_0^{\tau_{m-1}} h(\tau_m)\mathrm{d}\tau_m \cdots \mathrm{d}\tau_2\mathrm{d}\tau_1$. For

an RKHS of continuous functions, $\mathcal{Y}$, as given in Theorem 2, the mapping $h \mapsto h^{(-m)}(T)$ is a bounded functional. As a result, by the Riesz representation theorem, there exists a function denoted by $1^{(-m)} \in \mathcal{Y}$ such that $\langle h, 1^{(-m)}\rangle_{\mathcal{Y}} = \frac{1}{(m-1)!}\int_0^T (T-t)^{m-1}h(t)\mathrm{d}t$. Note that for $g \in \tilde{H}$, $\theta \in C^d([0,T],\mathbb{R}^n)$, and $\psi \in H$ such that $\psi = \mathcal{T}g$, the functional

$$\psi \mapsto \langle g \circ \theta, 1^{(-m)}\rangle_{\mathcal{Y}} = \frac{1}{(m-1)!}\int_0^T (T-t)^{m-1}g(\theta(t))\mathrm{d}t$$

is bounded. As such, by the Riesz representation theorem, there exists $\Gamma_\theta^{(m)} \in H$ such that $\langle \psi, \Gamma_\theta^{(m)}\rangle_H = \langle g \circ \theta, 1^{(-m)}\rangle_{\mathcal{Y}}$. We define $\Gamma_\theta^{(m)}$ as the *m-th order occupation kernel corresponding to $\theta \in C^d([0,T],\mathbb{R}^n)$ in H.*

Due to the fact that $\langle g \circ \theta, 1^{(-m)}\rangle_{\mathcal{Y}} = \langle \psi[\theta], 1^{(-m)}\rangle_{\mathcal{Y}} = \langle \psi, K_{\theta, 1^{(-m)}}\rangle_H$, the m-th order occupation kernel corresponding to θ can be identified with the reproducing kernel $K_{\theta, 1^{(-m)}} \in H$ of the vvRKHS.

Thus, in contrast with [22], where occupation kernels are integrals of the reproducing kernel of an RKHS along trajectories, the m-th order occupation kernels defined here are a subset of the set of vector valued kernels in a vvRKHS.

Example 7 *As an example application of signal valued RKHSs, consider, for a function $f : \mathbb{R}^n \to \mathbb{R}^n$, the action of the Liouville operator on $\psi \in H$, given as $A_f\psi[\theta](t) = \nabla\psi[\theta](t)f[\theta](t)$, where $f[\theta](t) := f(\theta(t))$, and $\nabla\psi[\theta] := \nabla g(\theta(\cdot))$ where $\mathcal{T}g = \psi$. Assume that A_f is densely defined over H. If $\gamma : [0,T] \to \mathbb{R}^n$ satisfies $\dot{\gamma} = f(\gamma)$ it follows that*

$$\langle A_f\psi, \Gamma_\gamma^{(1)}\rangle_H = \int_0^T \nabla\psi[\gamma](t)f[\gamma](t) =$$
$$\int_0^T \nabla g(\gamma(t))f(\gamma(t))\mathrm{d}t = \int_0^T \frac{\mathrm{d}}{\mathrm{d}t}g(\gamma(t))\mathrm{d}t =$$
$$g(\gamma(T)) - g(\gamma(0)) = \psi[\gamma](T) - \psi[\gamma](0) =$$
$$\langle \psi, K_{\gamma,\mathscr{K}_T} - K_{\gamma,\mathscr{K}_0}\rangle_H,$$

where, $\mathscr{K}_t(s) := \mathscr{K}(s,t)$ for $t,s \in [0,T]$. Hence, $A_f^*\Gamma_\gamma^{(1)} = K_{\gamma,\mathscr{K}_T} - K_{\gamma,\mathscr{K}_0}$, which coincides with the relation between occupation kernels and Liouville operators over RKHSs introduced in [22].

6 Higher order Liouville operators

The structure of Liouville operators, given formally as $A_fg(x) = \nabla g(x)f(x)$, is derived from the orbital derivative. In particular, suppose that $\gamma : [0,T] \to \mathbb{R}^n$ satisfies $\dot{\gamma} = f(\gamma)$, then $A_fg(\gamma(t)) = \nabla g(\gamma(t))\dot{\gamma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}g(\gamma(t))$. Consequently, higher order Liouville operators may

be derived via the same process, where g is composed with γ and higher order derivatives with respect to time are taken. To wit, letting $\mathcal{H}[g]$ denote the Hessian of $g : \mathbb{R}^n \to \mathbb{R}$, $\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(\gamma(t)) = \dot{\gamma}(t)^\top \mathcal{H}[g](\gamma(t))\dot{\gamma}(t) + \nabla g(\gamma(t))\ddot{\gamma}(t)$. In the case that γ satisfies $\ddot{\gamma} = f(\gamma)$, the derivative becomes

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(\gamma(t)) = \nabla g(\gamma(t))f(\gamma(t)) +$$
$$\left(\dot{\gamma}(0) + I_{f\circ\gamma}(t)\right)^\top \mathcal{H}[g](\gamma(t))\left(\dot{\gamma}(0) + I_{f\circ\gamma}(t)\right),$$

where $I_{f\circ\gamma}(t) := \int_0^t f(\gamma(\tau))\mathrm{d}\tau$. If g is complex-valued, then a similar relationship holds for the real and the complex components of g.

Fixing H as a signal valued RKHS as in Theorem 2, let $f : \mathbb{R}^n \to \mathbb{R}^n$ be the symbol for a second order Liouville operator, $B_f : \mathcal{D}(B_f) \to H$, defined as

$$B_f\psi[\theta](t) := \nabla\psi[\theta](t)f(\theta(t))$$
$$+ \left(\dot{\theta}(0) + I_{f\circ\theta}(t)\right)^\top \mathcal{H}\psi[\theta](t)\left(\dot{\theta}(0) + I_{f\circ\theta}(t)\right), \quad (1)$$

where $\mathcal{D}(B_f)$ is precisely the collection of ψ for which $B_f\psi \in H$. Note that since $\psi = \phi_g$ for some $g \in \tilde{H}$, $\frac{\partial}{\partial x_i}\psi[\theta](t)$ is defined as $\frac{\partial}{\partial x_i}g(\theta(t))$ for $i = 1, \ldots, n$, which facilitates the definitions of the gradient and Hessian of ψ. Hence, when $\ddot{\gamma} = f(\gamma)$, $B_f\phi_g[\gamma](t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}g \circ \gamma(t)$. Owing to the integral appearing in (1), the operator B_f needs to be posed over a Hilbert space consisting of functions of trajectories. Additionally, in contrast to the first order Liouville operator, B_f is linear in ψ but not in the symbol, f.

7 Interaction between higher order occupation kernels and higher order Liouville operators

The operator B_f is connected to second order occupation kernels in the following manner. If $\ddot{\gamma} = f(\gamma)$, then

$$\langle B_f\psi, \Gamma_\gamma^{(2)}\rangle_H =$$
$$\int_0^T (T-t)B_f\psi[\gamma](t)\mathrm{d}t = \int_0^T (T-t)\ddot{\psi}[\gamma](t)\mathrm{d}t =$$
$$\psi[\gamma](T) - \psi[\gamma](0) - T\nabla\psi[\gamma](0)\dot{\gamma}(0) =$$
$$\langle \psi, K_{\gamma,\mathscr{K}_T} - K_{\gamma,\mathscr{K}_0} - TK_{\gamma,\mathscr{K}_0'}\rangle_H,$$

where $\mathscr{K}_0' := s \mapsto \left.\mathrm{d}(\mathscr{K}(s,t))/\mathrm{d}t\right|_{t=0} \in \mathcal{Y}$ follows from [26, p. 132]. Hence, the functional $\psi \mapsto \langle B_f\psi, \Gamma_\gamma^{(2)}\rangle_H$ is bounded, and the following proposition is established.

Proposition 8 *Let f be the symbol for a densely defined[1] second order Liouville operator, B_f, over a signal valued RKHS H from X to $\mathcal{Y}$ derived from $\tilde{H}$, where $\mathcal{Y}$ is a scalar-valued RKHS with a continuously differentiable kernel, and let $\gamma \in X$ be such that $\ddot{\gamma} = f(\gamma)$. Then, $\Gamma_\gamma^{(2)} \in \mathcal{D}(B_f^*)$, and $B_f^*\Gamma_\gamma^{(2)} = K_{\gamma,\mathscr{K}_T} - K_{\gamma,\mathscr{K}_0} - TK_{\gamma,\mathscr{K}_0'}$.*

8 Dynamic Mode Decomposition for Second Order Dynamical Systems

The objective of this section is to develop a data driven model for a system governed by an unknown second order nonlinear ordinary differential equation. The developed SoDMD algorithm is closely connected to the occupation kernel DMD algorithm detailed in [20]. The approach is to determine a finite rank representation of B_f over H and to perform an eigendecomposition on this representation to obtain eigenfunctions and eigenvectors for the representation. Following this, the full state observable is decomposed with respect to the eigenfunctions, which ultimately allows for a model to be extracted for the dynamical system.

Suppose that $\varphi \in \mathcal{D}(B_f)$ is an eigenfunction for B_f with eigenvalue $\lambda \in \mathbb{C}$. Then for $\ddot{\gamma} = f(\gamma)$, $\ddot{\varphi}[\gamma](t) = B_f\varphi[\gamma](t) = \lambda\varphi[\gamma](t)$. Hence,

$$\varphi[\gamma](t) = \frac{1}{2}\left(\varphi[\gamma](0) + \frac{\nabla\varphi[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda}}\right)e^{\sqrt{\lambda}t}$$
$$+ \frac{1}{2}\left(\varphi[\gamma](0) - \frac{\nabla\varphi[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda}}\right)e^{-\sqrt{\lambda}t}.$$

By slight abuse of notation, the full state observable ψ_{id} for the signal-valued case is defined as $\psi_{\mathrm{id}}[\theta] := \left[(\psi_{\mathrm{id}})_1[\theta] \ldots (\psi_{\mathrm{id}})_n[\theta]\right]^\top$, where $(\psi_{\mathrm{id}})_i[\theta] = \theta_i$, for all $i = 1, \ldots, n$. The objective is to decompose each dimension of the full state observable with respect to an eigenbasis, $\{\varphi_i\}_{i=1}^\infty$ with eigenvalues $\{\lambda_i\}_{i=1}^\infty$, of B_f, provided that one exists, so as to express $\psi_{id}[\gamma](t)$ as

$$\psi_{id}[\gamma](t) = \gamma(t) = \lim_{M\to\infty}\gamma_M(t). \quad (2)$$

Note that the limit in (2) converges in the norm topology of $\mathcal{Y}$ and as a result, uniformly over compact sets (see, e.g., [15, Section 3.1.11]). In (2),

$$\gamma_M(t) := \sum_{m=1}^M \xi_{m,M}\left(\frac{\left(\varphi_m[\gamma](0) + \frac{\nabla\varphi_m[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda_m}}\right)e^{\sqrt{\lambda_m}t}}{2}\right.$$
$$\left. + \frac{\left(\varphi_m[\gamma](0) - \frac{\nabla\varphi_m[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda_m}}\right)e^{-\sqrt{\lambda_m}t}}{2}\right).$$

---

[1] An operator $B_f : \mathcal{D}(B_f) \to H$ is called densely defined if $\mathcal{D}(B_f)$ is a dense subset of H.

Since the eigenfunctions may not be pairwise orthogonal, addition of each new eigenfunction to the linear combination in (2) may affect the coefficients corresponding to all other eigenfunctions. This dependence of the coefficients on the collection of basis functions is expressed through the second subscript M in the notation $\xi_{m,M}$. In the following, finite-rank representations of the coefficients $\xi_{m,M}$ are referred to as the second order Liouville modes for the dynamical system.

Since $\gamma_M \in \mathcal{Y}$, using [26, Corollary 4.36], we conclude that provided $\mathcal{K} \in C^2([0,T] \times [0,T], \mathbb{R})$, the real and imaginary parts of the sequence $\{\frac{d^2\gamma_M}{dt^2}\}_{M=1}^{\infty}$ converge uniformly over compact sets. Differentiating (2) twice and using [23, Theorem 7.17], a spectral representation of the vector field f can also be obtained as $f(x) = \frac{d^2\gamma}{dt^2}\big|_{\gamma(0)=x, t=0}$. That is,

$$f(x) = \lim_{M \to \infty} \sum_{m=1}^{M} \xi_{m,M} \lambda_m \tilde{\varphi}_m(x), \tag{3}$$

where the function $\tilde{\varphi} : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\tilde{\varphi}_m(x) := \varphi_m[\gamma](0)\big|_{\gamma(0)=x} = g_{\varphi_m}(x), \tag{4}$$

where $g_{\varphi_m} \in \tilde{H}$ satisfies $\mathcal{T} g_{\varphi_m} = \varphi_m$. Note that while the eigenfunction φ_m operates on the entire trajectory γ, the function $\tilde{\varphi}_m$ depends only on the initial condition, and as such, can be expressed as a function on $\mathbb{R}^n$.

Since f, and as a result, B_f, are unknown, a finite rank proxy of B_f needs to be constructed from the observed trajectories. In DMD, the eigenfunctions of a finite rank proxy of B_f are leveraged, as proxies of the eigenfunctions of B_f, to determine an estimate for the trajectory representation in (2) and the vector field representation in (3). Let $\{\gamma_i\}_{i=1}^{M} \subset X$ be a collection of observed trajectories for the second order dynamical system, and let $\alpha = \{\Gamma_{\gamma_i}^{(2)}\}_{i=1}^{M}$ be the corresponding collection of second order occupation kernels in H. Let P_α be the projection onto spanα. Similar to Liouville DMD, a projection, onto spanα, of the second order Liouville operator, restricted to spanα is the finite rank proxy considered for computation. In the following, the finite rank operator is denoted by $P_\alpha B_f|_\alpha$ and for any complex numbers $a_1, ..., a_M$, the notation $(a)_{i=1}^{M}$ is used to denote the column vector $(a_1, ..., a_M)^\top$.

Assumption 9 *The derivation of the Liouville modes relies the assumptions that, (1) the second order Liouville operator B_f admits eigenfunctions forming a basis of H, (2) the domain of B_f contains spanα, and (3) the closed span of the eigenfunctions of B_f contains the full state observables $(\psi_{id})_i$ for all $i=1,...,n$.*

For $h \in \mathcal{D}(B_f)$, the coefficients $\{b_i\}_{i=1}^{M}$ in the projection of $B_f h$ onto spanα, given by $P_\alpha B_f h = \sum_{i=1}^{M} b_i \Gamma_{\gamma_i}^{(2)}$, can be expressed as $(b_i)_{i=1}^{M} = G\left(\left\langle B_f h, \Gamma_{\gamma_i}^{(2)} \right\rangle_H\right)_{i=1}^{M}$, where $G := (\langle \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H)_{i,j=1}^{M}$ is the occupation kernel Gram matrix. Assuming that the occupation kernels are in the domain of the Liouville operator, i.e., $\alpha \subset \mathcal{D}(B_f)$, for $h \in$ spanα, given by $h = \sum_{i=1}^{M} a_i \Gamma_{\gamma_i}^{(2)}$, we have

$$\left\langle B_f h, \Gamma_{\gamma_j}^{(2)} \right\rangle_H =$$
$$\sum_{i=1}^{M} a_i \left\langle B_f \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \right\rangle_H = \sum_{i=1}^{M} a_i \left\langle \Gamma_{\gamma_i}^{(2)}, B_f^* \Gamma_{\gamma_j}^{(2)} \right\rangle_H$$
$$= \left(\left\langle \Gamma_{\gamma_1}^{(2)}, B_f^* \Gamma_{\gamma_j}^{(2)} \right\rangle_H, ..., \left\langle \Gamma_{\gamma_M}^{(2)}, B_f^* \Gamma_{\gamma_j}^{(2)} \right\rangle_H\right)(a_i)_{i=1}^{M}.$$

A representation of $P_\alpha B_f|_\alpha$, i.e., the matrix $[P_\alpha B_f]_\alpha^\alpha$ that maps the coefficients $\{a_i\}_{i=1}^{M}$ to the coefficients $\{b_i\}_{i=1}^{M}$, is then given as

$$[P_\alpha B_f]_\alpha^\alpha = G I^\top, \tag{5}$$

where $I := (\langle \Gamma_{\gamma_i}^{(2)}, B_f^* \Gamma_{\gamma_j}^{(2)} \rangle_H)_{i,j=1}^{M}$. A normalized eigenfunction of $P_\alpha B_f|_\alpha$ can be extracted from an eigenvector, ν_j, of $[P_\alpha B_f]_\alpha^\alpha$ with eigenvalue λ_j as

$$\hat{\varphi}_j = \frac{1}{\sqrt{\nu_j^\dagger G \nu_j}} \sum_{i=1}^{M} (\nu_j)_i \Gamma_{\gamma_i}^{(2)}, \tag{6}$$

where $(\cdot)^\dagger$ denotes the conjugate transpose. Indeed since ν_j is an eigenvector of $[P_\alpha B_f]_\alpha^\alpha$, and $[P_\alpha B_f]_\alpha^\alpha$ maps coefficients ν_j of $\hat{\varphi}_j \in$ spanα to the corresponding coefficients in $P_\alpha B_f|_\alpha \hat{\varphi}_j$, we have

$$P_\alpha B_f|_\alpha \hat{\varphi}_j = \frac{1}{\sqrt{\nu_j^\dagger G \nu_j}} P_\alpha B_f|_\alpha \left(\sum_{i=1}^{M} (\nu_j)_i \Gamma_{\gamma_i}^{(2)}\right)$$
$$= \frac{1}{\sqrt{\nu_j^\dagger G \nu_j}} \sum_{i=1}^{M} (\lambda_j \nu_j)_i \Gamma_{\gamma_i}^{(2)} = \lambda_j \hat{\varphi}_j.$$

Similar to the implementation of DMD for Koopman and Liouville operators, the eigenfunction can be leveraged as a proxy for a proper eigenfunction of B_f. In (6) and in the following development, $(x)_i$ denotes the projection onto the i-th coordinate of $x \in \mathbb{R}^n$. If the Gram matrix is rank-deficient, it can be regularized as $G = (\langle \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H)_{i,j=1}^{M} + l I_M$, where $l > 0$ is the regularization coefficient and I_M is the $M \times M$ identity matrix.

The SoDMD modes can now be constructed by examining the inner products $\left\langle (\psi_{id})_i, \Gamma_{\gamma_j}^{(2)} \right\rangle_H$, where $(\psi_{id})_i$

Algorithm 1. Pseudocode for the SoDMD algorithm described in Sections 8 and 9. Once the SoDMD modes, the normalized eigenfunctions, and the eigenvalues are returned, either (14) can be used along with a numerical integration routine, or (3) can be used along with an ordinary differential equation solver to reconstruct trajectories of the system starting from any given initial condition. 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 $\{\gamma_j : [0,T] \to \mathbb{R}^n\}_{j=1}^M$

Require: reproducing kernel $\tilde{K}$ of the RKHS $\tilde{H}$

Require: A numerical integration routine

1: Construct the matrix $G := (\langle \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H)_{i,j=1}^M$ using (9) and a numerical integration routine

2: Construct the matrix $I := (\langle \Gamma_{\gamma_i}^{(2)}, B_f^* \Gamma_{\gamma_j}^{(2)} \rangle_H)_{i,j=1}^M$ using (10), (11), and a numerical integration routine

3: Construct the matrix $[P_\alpha B_f]_\alpha^\alpha$ using (5) and compute its eigenvalues, λ_j, and eigenvectors, ν_j

4: Use (7), (12) and a numerical integration routine to compute the SoDMD modes ξ_j

5: Use (6), (8), and a numerical integration routine to construct the eigenfunctions $\hat{\varphi}_j$

6: Use (15) and a numerical integration routine to construct the initial eigenfunction evaluations $\tilde{\varphi}_j$

7: **return** SoDMD modes, ξ_j, eigenfunctions, $\hat{\varphi}_j$, initial evaluation functions, $\tilde{\varphi}_j$ and eigenvalues λ_j for $j = 1, \cdots, M$

---

is the i-th component of the full state observable, i.e., $(\psi_{id})_i[\theta](t) := (\theta(t))_i$. The SoDMD modes $\{(\xi_m)_i\}_{m=1}^M$ are defined as the coefficients in the projection of $(\psi_{id})_i$ onto the span of the normalized eigenfunctions in (6) that is,

$$\left(\left\langle (\psi_{id})_i, \Gamma_{\gamma_j}^{(2)} \right\rangle_H \right)_{i=1}^M \approx \left(\left\langle \sum_{m=1}^M (\xi_m)_i \hat{\varphi}_m, \Gamma_{\gamma_j}^{(2)} \right\rangle_H \right)_{i=1}^M$$

$$\left(\left\langle \sum_{m=1}^M (\xi_m)_i \sum_{k=1}^M \frac{(\nu_m)_k \Gamma_{\gamma_k}^{(2)}}{\sqrt{\nu_m^\dagger G \nu_m}}, \Gamma_{\gamma_j}^{(2)} \right\rangle_H \right)_{i=1}^M$$

$$= \sum_{m=1}^M \xi_m \sum_{k=1}^M \frac{(\nu_m)_k}{\sqrt{\nu_m^\dagger G \nu_m}} \left\langle \Gamma_{\gamma_k}^{(2)}, \Gamma_{\gamma_j}^{(2)} \right\rangle_H = \sum_{m=1}^M \frac{\xi_m \nu_m^\dagger (G)^j}{\sqrt{\nu_m^\dagger G \nu_m}},$$

where $(G)^j$ denotes the j-th column of the Gram matrix. The matrix $\xi := \left(\xi_1 \ \cdots \ \xi_M \right)$ of second order Liouville modes is then given by

$$\xi = J \left(\left[\left(\frac{\nu_i^\top}{\sqrt{\nu_i^\dagger G \nu_i}} \right)_{i=1}^M \right] G \right)^{-1}, \tag{7}$$

where $J := \left(\left\langle (\psi_{id})_i, \Gamma_{\gamma_j}^{(2)} \right\rangle_H \right)_{i,j=1}^{n,M}$. Equations (5), (6), and (7) constitute the algorithm for obtaining the second order dynamic modes of a dynamical system from observed trajectories. The modes can be utilized via (2), to obtain a data driven model for the dynamical system that reconstructs trajectories of the system starting from given initial conditions, or via (3), to estimate the vector field f. The resulting SoDMD procedure is summarized in Algorithm 1. The following section discusses numerical computation of the inner products needed to implement the algorithm.

9 Evaluation of inner products and occupation kernels

To implement the algorithm, one needs to evaluate quantities $\Gamma_{\gamma_i}^{(2)}[\theta]$, $\langle \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H$, $\langle B_f^* \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H$, and $\left\langle (\psi_{id})_k, \Gamma_{\gamma_i}^{(2)} \right\rangle_H$, where $i,j = 1,...,M$, $k = 1,...,n$, and $\theta \in X$.

First note that as a surjective isometry, the operator $\mathcal{T}$, given in Theorem 2, is a unitary operator. Moreover, if $\psi \in H$ such that $g = \mathcal{T}^{-1}\psi$, then for $\theta \in X$ and $t \in [0,T]$, we have $\langle \psi, K_{\theta, \mathscr{K}_t} \rangle_H = \psi[\theta](t) = g(\theta(t)) = \langle g, \tilde{K}_{\theta(t)} \rangle_{\tilde{H}} = \langle \mathcal{T}g, \mathcal{T}\tilde{K}_{\theta(t)} \rangle_H = \langle \psi, \mathcal{T}\tilde{K}_{\theta(t)} \rangle_H$. Since ψ was arbitrary, it follows that $K_{\theta, \mathscr{K}_t} = \mathcal{T}\tilde{K}_{\theta(t)}$. Thus,

$$\langle K_{\theta_1, \mathscr{K}_{t_1}}, K_{\theta_2, \mathscr{K}_{t_2}} \rangle_H = \langle \mathcal{T}\tilde{K}_{\theta_1(t_1)}, \mathcal{T}\tilde{K}_{\theta_2(t_2)} \rangle_H =$$
$$\langle \tilde{K}_{\theta_1(t_1)}, \tilde{K}_{\theta_2(t_2)} \rangle_{\tilde{H}} = \tilde{K}(\theta_2(t_2), \theta_1(t_1)).$$

As a result,

$$\Gamma_{\gamma_i}^{(2)}[\theta](t) = \langle \Gamma_{\gamma_i}^{(2)}, K_{\theta, \mathscr{K}_t} \rangle_H = \int_0^T (T-\tau) K_{\theta, \mathscr{K}_t}[\gamma_i](\tau) \mathrm{d}\tau$$
$$= \int_0^T (T-\tau) \tilde{K}(\theta(t), \gamma_i(\tau)) \mathrm{d}\tau. \tag{8}$$

Consequently,

$$\langle \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H = \int_0^T \int_0^T (T-\tau)(T-t) \tilde{K}(\gamma_i(t), \gamma_j(\tau)) \mathrm{d}t \mathrm{d}\tau. \tag{9}$$

To compute $\langle B_f^* \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H$, recall that according to Proposition 8, when γ is trajectory satisfying $\ddot{\gamma} = f(\gamma)$, $B_f^* \Gamma_{\gamma_i}^{(2)} = K_{\gamma_i, \mathscr{K}_T} - K_{\gamma_i, \mathscr{K}_0} - T K_{\gamma_i, \mathscr{K}_0'}$. By linearity,

$$\langle B_f^* \Gamma_{\gamma_i}^{(2)}, \Gamma_{\gamma_j}^{(2)} \rangle_H = \langle K_{\gamma_i, \mathscr{K}_T}, \Gamma_{\gamma_j}^{(2)} \rangle_H - \langle K_{\gamma_i, \mathscr{K}_0}, \Gamma_{\gamma_j}^{(2)} \rangle_H$$
$$- \langle T K_{\gamma_i, \mathscr{K}_0'}, \Gamma_{\gamma_j}^{(2)} \rangle_H. \tag{10}$$

Using the fact that $\langle K_{\gamma_i, \mathscr{K}_T}, \Gamma_{\gamma_j}^{(2)} \rangle_H = \Gamma_{\gamma_j}^{(2)}[\gamma_i](T)$ and $\langle K_{\gamma_i, \mathscr{K}_T}, \Gamma_{\gamma_j}^{(2)} \rangle_H = \Gamma_{\gamma_j}^{(2)}[\gamma_i](0)$, which can be computed

using (8), the only unresolved quantity in (10) is $\langle K_{\gamma_i,\mathscr{K}_0'},\Gamma_{\gamma_j}^{(2)}\rangle_H$. To that end, we have $\langle \psi, K_{\gamma_i,\mathscr{K}_0'}\rangle_H = \psi'[\gamma_i](0) = g'(\gamma_i(0)) = \langle g, K'_{\gamma_i(0)}\rangle_{\tilde{H}}$. Hence, $\langle K_{\gamma_i,\mathscr{K}_0'},\Gamma_{\gamma_j}^{(2)}\rangle_H$ can be computed as

$$\langle K_{\gamma_i,\mathscr{K}_0'},\Gamma_{\gamma_j}^{(2)}\rangle_H = \int_0^T (T-t)\nabla_2 \tilde{K}(\gamma_j(t),\gamma_i(0))\dot{\gamma}_i(0)\mathrm{d}t. \tag{11}$$

The remaining computation for $\left\langle (\psi_{id})_k, \Gamma_{\gamma_i}^{(2)}\right\rangle_H$ can be accomplished by recalling that

$$\left\langle (\psi_{id})_k, \Gamma_{\gamma_i}^{(2)}\right\rangle_H = \left\langle (\psi_{id})_k[\gamma_i], 1^{(-2)}\right\rangle_{\mathcal{Y}} =$$
$$\left\langle (\gamma_i)_k, 1^{(-2)}\right\rangle_{\mathcal{Y}} = \int_0^T (T-\tau)(\gamma_i(\tau))_k \mathrm{d}\tau. \tag{12}$$

Equations (5), (6), and (7), along with the computational details in equations (8) - (12), complete the description of the SoDMD algorithm. The computational complexity of the Algorithm 1 is dominated by construction of the matrices in steps 1 and 2, the matrix inversion in (5), and the matrix inversion in (7). The complexity of matrix construction depends on the complexity of kernel evaluation. Assuming that kernel evaluation complexity is linear in n, the matrices in steps 1 and 2 can be constructed in $\mathcal{O}(M^2 s^2 n)$ computations, where s is the number of samples used for numerical integration. The matrix inversion is $\mathcal{O}(M^3)$, which gives us a total computational complexity of $\mathcal{O}(M^2 s^2 n + M^3)$. This bound is similar to first order DMD, with the exception that n is twice as large.

10 Data-driven representation of the dynamical system

The generated higher order Liouville modes, eigenfunctions, and eigenvalues can be used to constitute a data-driven representation of the underlying dynamical system, and can be used to generate trajectories of the system via two different approaches. In the direct approach, the trajectories $\hat{\gamma}_D$ are generated using the proxy eigenfunctions in (6) in a truncated version of (2), i.e.,

$$\hat{\gamma}_D(t) = \sum_{j=1}^M \frac{\xi_j}{2}\left(\left(\hat{\varphi}_j[\gamma](0) + \frac{\nabla\hat{\varphi}_j[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda_j}}\right)e^{\sqrt{\lambda_j}t}\right.$$
$$\left. + \left(\hat{\varphi}_j[\gamma](0) - \frac{\nabla\hat{\varphi}_j[\gamma](0)\dot{\gamma}(0)}{\sqrt{\lambda_j}}\right)e^{-\sqrt{\lambda_j}t}\right). \tag{13}$$

In the indirect approach, the trajectories $\hat{\gamma}_I$ are generated using a truncated version of the vector field in (3),

i.e., by solving the differential equation

$$\ddot{x} = \hat{f}(x) := \sum_{j=1}^M \xi_j \lambda_j \tilde{\varphi}_j(x), \tag{14}$$

starting from the same randomly selected initial conditions as above, using the MATLAB® ode45 solver. Note that the function $\tilde{\varphi}_j$ used in (14) was introduced in (4), and can be computed using the proxy eigenfunction from (6) in (8) to yield

$$\tilde{\varphi}_j(x) = \hat{\varphi}_j[\gamma](0)|_{\gamma(0)=x} = \sum_{i=1}^M \frac{(\nu_j)_i \, \Gamma_{\gamma_i}^{(2)}[\gamma](0)\big|_{\gamma(0)=x}}{\sqrt{\nu_j^\dagger G \nu_j}}$$
$$= \sum_{i=1}^M \frac{(\nu_j)_i}{\sqrt{\nu_j^\dagger G \nu_j}}\int_0^T (T-\tau)\tilde{K}(x,\gamma_i(\tau))\mathrm{d}\tau. \tag{15}$$

Once the modes are computed, the computational complexity of each evaluation of $\hat{f}$ is dominated by matrix multiplication, which gives us the bound $\mathcal{O}(M^2 + Mn)$. This bound is similar to first order DMD, with the exception that n is twice as large.

11 Numerical experiments

The following numerical experiments demonstrate the efficacy of the developed DMD method. Note that the range space $\mathcal{Y}$, while needed in the construction of the signal-valued RKHS, H, is not needed to implement Algorithm 1. Verification of Assumption 9 and similar assumptions made in the DMD literature for general dynamical systems is an open question [1].

Experiment 1: In this experiment, 16 trajectories of the undamped linear oscillator $\ddot{\gamma} = -2\gamma$, over the time horizon $[0,5]$, starting from initial conditions over a regular grid on the domain $[-1,1]\times[-1,1]$, are utilized to build a data-driven model in the form of Liouville modes and eigenfunctions and eigenvalues of the finite rank representation of the second order Liouville operator. A Gaussian RBF kernel $\tilde{K}(x,y) = \exp\left(\frac{\|x-y\|_2^2}{\mu}\right)$ with width parameter $\mu = 100$ is used to generate the data-driven model. The rank of the Gram matrix is found to be 6, which necessitates the use of a regularization coefficient $l = 10^{-8}$. The trajectories are sampled every 0.5s. To simulate measurement noise, a random number drawn from the zero mean Gaussian distribution with standard deviation 0.01 is added to every sample of the trajectories and the time derivatives at the initial time. The model generated using SoDMD is compared against a model obtained using first order DMD (same kernel with width parameter $\mu = 100$ and regularization coefficient 10^{-6}) by manually tuning parameters in both models for the best
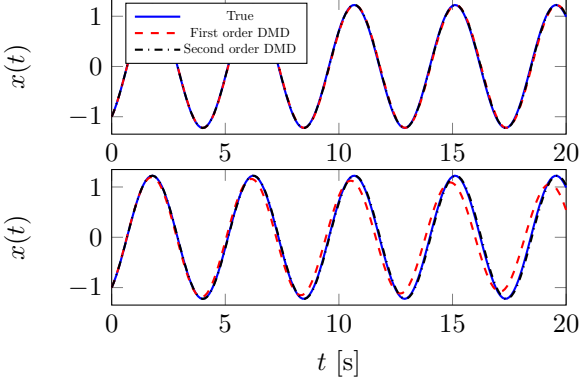
Fig. 2. The plot shows a true trajectory of a linear oscillator and the corresponding estimates generated by the data-driven model, obtained via the indirect approach, using first and second order DMD, starting from the same initial condition. The plot on the top uses noise-free data and the plot at the bottom uses noisy data.
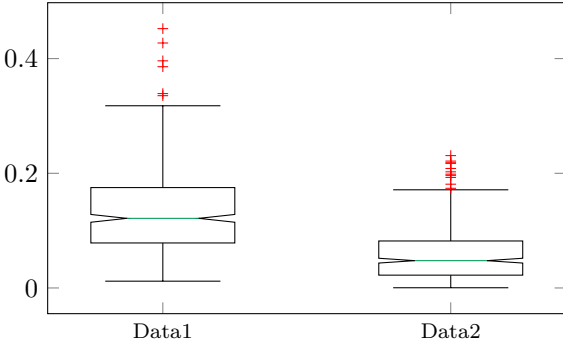

Fig. 3. Performance of the data-driven model for the linear oscillator over 500 trials. Each trial generates a noisy trajectory of the linear oscillator, over $t \in [0,20]$ s, starting from a randomly selected initial condition, using the indirect approach for first and second order DMD. The root mean square (RMS) value of the relative error signal $t \mapsto \|\gamma(t) - \hat{\gamma}(t)\| / |\gamma|_\infty$ is computed for first and second order DMD in each trial to generate this box plot.

noise-free reconstruction of a trajectory starting from $\gamma(0) = -1$ and $\dot{\gamma}(0) = 1$ (see Fig. 2.)

The Liouville modes and the finite rank eigendecomposition are then used to generate trajectories of the system, using the indirect approach detailed in Section 10, over a longer time horizon of $[0,20]$, starting from 500 randomly selected initial conditions in the domain $[-1,1] \times [-1,1]$.

The root mean square (RMS) value of the relative error signal $t \mapsto \|\gamma(t) - \hat{\gamma}(t)\| / \|\gamma\|_\infty$ is used to quantify the quality of the data-driven model. The results are compared against data-driven models generated by first order DMD. Fig. 3 shows the relative RMS errors in each of the 500 trials. Fig. 2 shows the true trajectory and the trajectories generated using first and second order DMD.
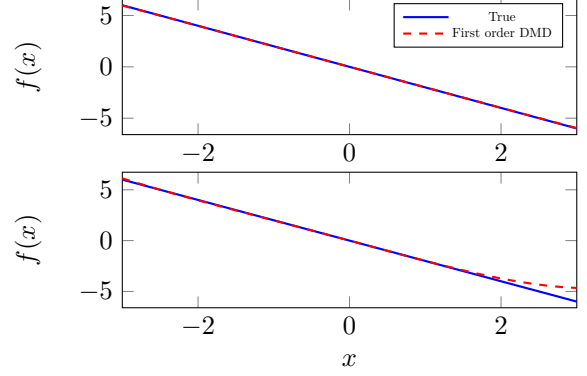

Fig. 4. The plot shows the true vector field of the linear oscillator and the data-driven estimate of the vector field generated using the indirect approach over the domain $[-3,3]$. Note that the data used for DMD were generated using trajectories starting in the domain $[-1,1] \times [-1,1]$. The plot on the top uses noise-free data and the plot at the bottom uses noisy data.

To test the generalization capabilities of the developed model, the Liouville modes and the finite rank eigendecomposition are used to approximate the vector field using (3) over a larger domain ($[-3,3]$). Fig. 4 shows the true vector field, f and the approximated vector field $\hat{f}$, introduced in (14), generated using the indirect approach in one representative trial.

Experiment 2: In this experiment, 100 trajectories of the Duffing oscillator $\ddot{\gamma} = \gamma - \gamma^3$, over the time horizon $[0,5]$, starting from initial conditions over a regular grid on the domain $[-5,5] \times [-5,5]$, are utilized to build a data-driven model in the form of Liouville modes and eigenfunctions and eigenvalues of the finite rank representation of the second order Liouville operator. An exponential dot product kernel $\tilde{K}(x,y) = \exp\left(\frac{x^\top y}{\mu}\right)$ with decay rate parameter $\mu = 25$ and is used to generate the data-driven model. The rank of the Gram matrix is found to be 11, which necessitates the use of a regularization coefficient $l = 10^{-5}$. The trajectories are sampled every 0.05s. To simulate measurement noise, a random number drawn from the zero mean Gaussian distribution with standard deviation 0.01 is added to every sample of the trajectories and the time derivatives. The model generated using SoDMD is compared against a model obtained using first order DMD (same kernel with decay rate parameter $\mu = 200$ and regularization coefficient 10^{-8}) by manually tuning parameters in both models for the best noise-free reconstruction of a trajectory starting from $\gamma(0) = -1$ and $\dot{\gamma}(0) = 1$ (see Fig. 5).

Similar to the linear oscillator, the results are compared against data-driven models generated by first order DMD. Fig. 6 shows the relative RMS errors in 500 trials starting from 500 randomly selected initial conditions in the domain $[-5,5] \times [-5,5]$. Generalization capabilities of the developed model are also evaluated by plotting
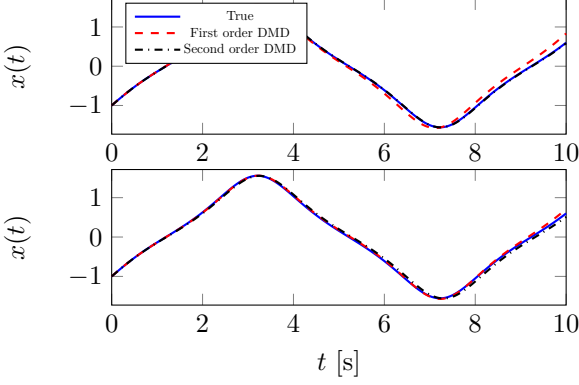
9

Fig. 5. The plot shows a true trajectory of the Duffing oscillator and the corresponding estimates generated by the data-driven model, obtained via the indirect approach, using first and second order DMD, starting from the same initial condition. The plot on the top uses noise-free data and the plot at the bottom uses noisy data.
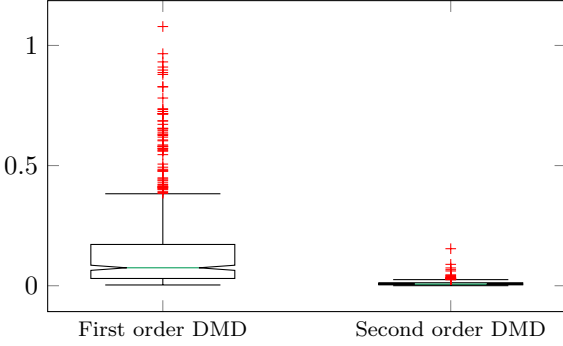

Fig. 6. Performance of the data-driven model of the Duffing oscillator over 500 trials. Each trial generates a noisy trajectory of the Duffing oscillator, over $t \in [0,10]$s, starting from a randomly selected initial condition in the domain $[-5,5] \times [-5,5]$, using the indirect approach for first and second order DMD. The root mean square (RMS) value of the relative error signal $t \mapsto \|\gamma(t) - \hat{\gamma}(t)\| / |\gamma|_\infty$ is computed for first and second order DMD in each trial to generate this box plot. The first order DMD method produced NaN results for a few initial conditions that were removed from the dataset.

the true and the estimated vector fields in the larger domain $[-10,10]$.

Fig. 5 shows the true trajectory and the trajectories generated using first and second order DMD in the *noise-free* tuning run. Fig. 7 shows the true vector field, f and the approximated vector field $\hat{f}$, introduced in (14), generated using the indirect approach in one representative trial.

Experiment 3: This experiment comprises of a two dimensional undamped cantilever beam that is initially assumed to be bent under a 10 N load applied at the end. The load is removed at the initial time, resulting in sustained oscillations. The beam model is a linear plane-stress elasticity problem from [28]. A single trajectory
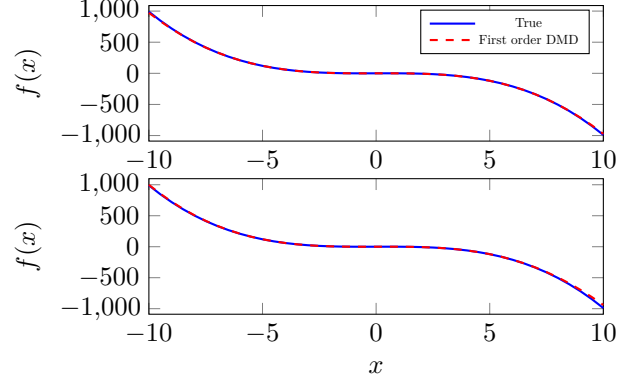

Fig. 7. The plot shows the true vector field of the Duffing oscillator and the data-driven estimate of the vector field generated using the indirect approach over the domain $[-10,10]$. Note that the data used for training were generated using trajectories starting in the domain $[-5,5] \times [-5,5]$. The plot on the top uses noise-free data and the plot at the bottom uses noisy data.
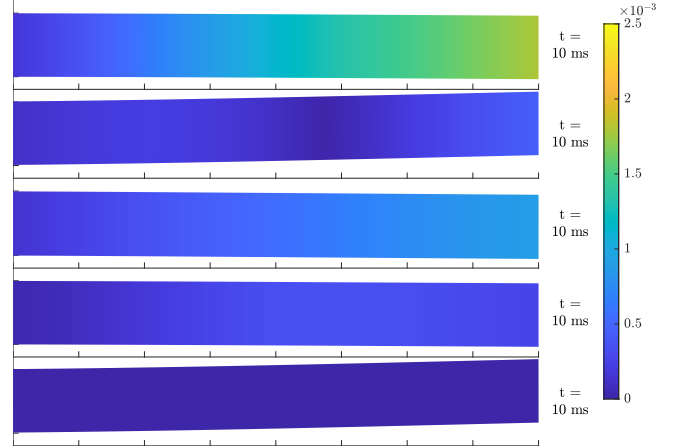

Fig. 8. Reconstruction of oscillations of an undamped two dimensional cantilever beam fixed at the left end. True positions of the finite element nodes are compared against their reconstructed positions, computed using a DMD model trained from trajectory segments 2.3 ms long. The plot shows the shape of the beam as computed by the finite element solver and color of the beam indicates the magnitude of the local reconstruction error, between the FEA calculation and the DMD calculation, scaled according to the colorbar on the right.

of the beam, comprised of displacements in the x and y direction of 6511 nodes is computed over the time horizon $[0,0.0236]$ using the MATLAB® Partial Differential Equation Toolbox [27]. The trajectory is sampled every 0.0788 ms. The resulting dataset of 301 snapshots, each with dimension 13022, is segmented to yield 271 trajectories, 2.3 ms (31 snapshots) long. The 271 trajectories are utilized for SoDMD using a $\tilde{K}(x,y) = x^\top y / \mu$ with slope parameter $\mu = 0.0005$. The rank of the Gram matrix is found to be 5, which necessitates the use of a regulariza-
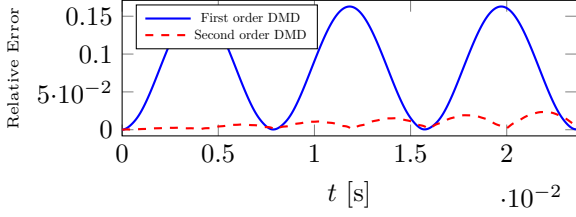
10

Fig. 9. Relative error $t \mapsto \|\gamma(t) - \hat{\gamma}(t)\| / |\gamma|_\infty$ between the FEA trajectory and the DMD estimate, computed using first and second order DMD, for the undamped two dimensional cantilever beam fixed at the left end, plotted over $t \in [0,25]$ ms. To test if the DMD models can extrapolate, the trajectory starts from an initial condition that is outside of the domain of the training data.

tion coefficient $l=10^{-14}$.

The Liouville modes and the finite rank eigendecomposition are then used in (13) and (14) to reconstruct the original trajectory $\hat{\gamma}$ over the time horizon $[0, 0.0102]$, starting from the same initial condition as the original trajectory. Fig. 8 shows true positions and the reconstruction error of the nodes in the finite element analysis (FEA).

To test the generalization capabilities of the developed model, a new initial condition for the beam, bent under a 15N load applied at the end, is computed using the MATLAB® Partial Differential Equation Toolbox [27]. Using the Liouville modes, the finite rank eigendecomposition, and the models in (13) and (14), a prediction of the beam trajectory is generated starting from the new initial condition. The prediction is then compared against the FEA estimation of the trajectory starting from the new initial condition, also computed using the the MATLAB® Partial Differential Equation Toolbox [27]. Fig. 9 shows relative RMS errors for trajectory prediction using the direct and the indirect approaches.

12 Discussion

This manuscript presents a framework for SoDMD methods by introducing signal valued RKHSs. These Hilbert spaces may be constructed from traditional scalar valued RKHSs via the construction given in Theorem 2. The mapping $T:\tilde{H} \to H$ is a unitary operator and H can be seen to be unitarily equivalent to $\tilde{H}$. However, this extra dressing on top of $\tilde{H}$ that enables the consideration of a signal as the fundamental input is necessary for the definition of higher order Liouville operators, such as B_f. Significantly, as can be observed in Section 9, the range space $\mathcal{Y}$ is not accessed for any of the evaluations. This means that $\mathcal{Y}$ is necessary *only* for theoretical completeness and not for practical implementation. While only second order Liouville operators were presented, signal valued RKHSs are sufficient for the analysis of higher

order Liouville operators provided the smoothness d of the signal space X is selected to be large enough.

Section 8 gives an algorithm for the direct DMD analysis of second order systems that avoids augmenting the system state and numerically estimating the time derivative of the system state. Accounting for the computations in Section 9, it can be seen that the implementation of the algorithm differs only slightly from that of the DMD method in [20]. However, it can be observed that it is necessary to numerically estimate the time derivative only at a single point in the time domain, $t=0$, which is a dramatic improvement over numerically differentiating the entire trajectory. However, higher order Liouville operators will require additional numerical derivatives at the origin, which may limit the applicability of the method when initial values of high order system is unknown. The effects of noisy derivatives are apparent in Figures 5 and 6, where without noise, the trajectory estimates of both first and second order methods are identical, but with noise, a significantly larger fraction of the second order predictions are better than the first order predictions.

Section 11 includes three numerical experiments, two that demonstrates the robustness of second order Liouville DMD to measurement noise in linear and nonlinear systems and another that demonstrates application to systems with a high dimensional state space. The results of Experiments 1 and 2 demonstrate that given a small amount of noisy data, the models generated via second-order Liouville DMD can accurately reconstruct and predict system behavior starting from given initial conditions within and outside of the domain covered by the recorded data.

The results of Experiment 3 demonstrate the applicability of second order Liouville DMD to a real-world problem with a very high-dimensional state space. The generated direct and indirect data-driven models are able to accurately characterize the transient dynamics of an undamped two dimensional cantilever beam starting from the same initial condition as the recorded data and from initial conditions outside the domain covered by the recorded data.

As indicated by Figures 3, 6 and 9, the models generated using SoDMD perform better than those generated using first order DMD. The simulation results depend on tuning of parameters such as the kernel width and the regularization coefficient, and as such, are not definitive. However, the authors postulate that the improved performance is attributed to the fact that the structure of the system, where the state is equal to the derivative of the output, is built into the SoDMD model. The first order DMD model makes no such assumptions, and instead has to infer the relationship between the output and the state from the training data.

The results demonstrated in Section 11 are obtained us-

ing the Gaussian RBF kernel for Experiment 1, the exponential dot product kernel for Experiment 2, and the linear dot product kernel for Experiment 3. While agnostic to the type of kernel, a successful implementation of the DMD technique does require careful selection of kernel shape parameters, e.g., the width parameter for Gaussian RBF kernels, the slope parameter for linear dot product kernels, and the decay rate parameter for exponential dot product kernels. In experiments presented in Section 11, the parameters were hand-tuned using trial and error.

13 Conclusion

This manuscript presented the necessary theoretical foundations for the definition of higher order Liouville operators. This manuscript proceeded to give a DMD algorithm for modeling a second order dynamical system that leveraged second order Liouville operators and occupation kernels. Through the examination of the components of the algorithm in Section 9, it was determined that the actual implementation of the SoDMD routine requires only a slight modification of DMD with occupation kernels in [20], and the range space was ultimately unnecessary for computations.

The experiments in Section 11 involve undamped oscillators and undamped beams because damped oscillators and damped beams typically cannot be cast into the form $\ddot{\gamma} = f(\gamma)$, but require the form $\ddot{\gamma} = f(\gamma, \dot{\gamma})$. An immediate future direction for this research is in exploring extension to such models where the higher order derivatives of the output depend not just on the output, but also on its derivatives.

While the primary advantage of SoDMD is that it does not require measurement of derivatives of the trajectory, the authors postulate that even when the derivatives are available, the smaller state dimension in SoDMD would lead to similar accuracy but faster execution, especially in systems with a large number of state variables. Confirmation of the above postulate via a detailed numerical comparison of the performance (in terms of computation time) of models generated using first order DMD with those generated using SoDMD is also a part of future research.

References

[1] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Nat. Acad. Sci. U.S.A.*, 113(15):3932–3937, 2016.

[2] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied Koopmanism. *Chaos*, 22(4):047510, 2012.

[3] Claudio Carmeli, Ernesto De Vito, Alessandro Toigo, and Veronica Umanitá. Vector valued reproducing kernel Hilbert spaces and universality. *Anal. Appl.*, 08(01):19–61, 2010.

[4] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2 edition, 2010.

[5] W. M. Haddad. *A dynamical systems theory of thermodynamics.* Princeton Series in Applied Mathematics. Princeton University Press, 2019.

[6] Thomas G. Hallam and Simon A. Levin. *Mathematical ecology: an introduction*, volume 17 of *Biomathematics*. Springer Science & Business Media, 2012.

[7] J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. *Dynamic mode decomposition - data-driven modeling of complex systems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[8] Jean B. Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat. Nonlinear optimal control via occupation measures and LMI-relaxations. *SIAM J. Control Optim.*, 47(4):1643–1666, 2008.

[9] Soledad Le Clainche and José M. Vega. Higher order dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.*, 16(2):882–925, 2017.

[10] Soledad Le Clainche and José M. Vega. Higher order dynamic mode decomposition to identify and extrapolate flow patterns. *Phys. Fluids*, 29(8):084102, 2017.

[11] Xiuying Li and Joel A. Rosenfeld. Fractional order system identification with occupation kernel regression. *IEEE Control Syst. Lett.*, 6:19–24, 2020.

[12] Jonathan H. Manton. Differential calculus, tensor products and the importance of notation. arXiv:1208.0197, 2013.

[13] Alexandre Mauroy and Jorge Goncalves. Koopman-based lifting techniques for nonlinear systems identification. *IEEE Trans. Autom. Control*, 65(6):2550–2565, 2020.

[14] Alexandre Mauroy and Igor Mezić. Global stability analysis using the eigenfunctions of the Koopman operator. *IEEE Trans. Autom. Control*, 61(11):3356–3369, 2016.

[15] Gert K Pedersen. *Analysis now*, volume 118. Springer Science & Business Media, 2012.

[16] George Pedrick. *Theory of reproducing kernels for Hilbert spaces of vector valued functions.* PhD thesis, University of Kansas, 1957.

[17] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.*, 15(1):142–161, 2016.

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

[19] Joel A. Rosenfeld and Rushikesh Kamalapurkar. Dynamic mode decomposition with control Liouville operators. *IEEE Trans. Autom. Control*, 69(12):8571–8586, December 2024.

[20] Joel A. Rosenfeld, Rushikesh Kamalapurkar, L. Forest Gruss, and Taylor T. Johnson. Dynamic mode decomposition for continuous time systems with the Liouville operator. *J. Nonlinear Sci.*, 32(1):1–30, 2022.

[21] Joel A. Rosenfeld, Rushikesh Kamalapurkar, Benjamin Russo, and Taylor T. Johnson. Occupation kernels and densely defined Liouville operators for system identification. In *Proc. IEEE Conf. Decis. Control*, pages 6455–6460, 2019.

[22] Joel A. Rosenfeld, Benjamin Russo, Rushikesh Kamalapurkar, and Taylor Johnson. The occupation kernel method for nonlinear system identification. *SIAM J. Control Optim.*, 62(3):1643–1668, January 2024.

[23] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3rd edition, 1976.

[24] Benjamin P. Russo, Rushikesh Kamalapurkar, Dongsik Chang, and Joel A. Rosenfeld. Motion tomography via occupation kernels. *J. Comput. Dyn.*, 9(1):27–45, 2022.

[25] Maciej Skorski. Chain rules for hessian and higher derivatives made easy by tensor calculus. arXiv:1911.13292, 2019.

[26] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[27] The MathWorks, Inc. Dynamic analysis of clamped beam. https://www.mathworks.com/help/pde/ug/dynamic-analysis-of-a-clamped-beam.html. Accessed: 2025-04-21.

[28] The MathWorks, Inc. Linear elasticity equations. https://www.mathworks.com/help/pde/ug/3-d-linear-elasticity-equations-in-toolbox-form.html, 2025. Accessed: 2025-04-21.

[29] János Tóth, Attila László Nagy, and Dávid Papp. *Reaction kinetics: exercises, programs and theorems*. Springer, 2018.

[30] Jose Manuel Vega and Soledad Le Clainche. *Higher order dynamic mode decomposition and its applications*. Academic Press, 2020.

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

A Extension to orders higher than two

Formulation of higher order Liouville operators requires computation of higher order derivatives using the chain rule in the multi-variable setting. Such computation can be achieved using high dimensional arrays or component-wise via a multivariable Faà di Bruno formula. For our purposes here we will use the high dimensional array approach. A good reference for the computational intricacies involved with the higher order chain rule can be found in [12, 25]. Using the tensor notation in [12] and letting D denote the Fréchet derivative, we have $D(fg) = (Df)(I \otimes g) + fDg$, where I is the identity matrix, $D(f \otimes g) = (Df \otimes g) + (f \otimes Dg)$, and $D(f \circ g) = (Df \circ g)(Dg \otimes I)$. Under this set of rules we can show that

$$D^2(f \circ g) = (D^2 f \circ g)(Dg \otimes Dg) + (Df \circ g)D^2 g$$

which has the form used in the above sections by symmetry [25] and

$$D^3(f \circ g) = (D^3 f \circ g)(Dg \otimes Dg \otimes Dg) + (D^2 f \circ g)[(D^2 g \otimes Dg) + 2(Dg \otimes D^2 g)] + (Df \circ g)D^3 g.$$

In the context of Section 6 we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}(g(\gamma(t)) = D^3 g(\gamma(t)) \left[\bigotimes^3 \left(\dot{\gamma}(0) + \right. \right.$$

$$\left. \left. \int_0^T \ddot{\gamma}(0)\mathrm{d}t + \int_0^T \int_0^t f(\gamma(\tau))\mathrm{d}\tau \mathrm{d}t \right) \right]$$

$$+ D^2 g(\gamma(t)) \left[\left(\ddot{\gamma}(0) + \int_0^T f(\gamma(t))\mathrm{d}t \right) \otimes \right.$$

$$\left. \left(\dot{\gamma}(0) + \int_0^T \ddot{\gamma}(0)\mathrm{d}t + \int_0^T \int_0^t f(\gamma(\tau))\mathrm{d}\tau \mathrm{d}t \right) \right]$$

$$+ 2D^2 g(\gamma(t)) \left[\left(\dot{\gamma}(0) + \int_0^T \ddot{\gamma}(0)\mathrm{d}t + \right. \right.$$

$$\left. \left. \int_0^T \int_0^t f(\gamma(\tau))\mathrm{d}\tau \mathrm{d}t \right) \otimes \left(\ddot{\gamma}(0) + \int_0^T f(\gamma(t))\mathrm{d}t \right) \right]$$

$$+ \nabla g(\gamma(t))f(\gamma(t)).$$

The above equation can be turned into the appropriate operator via the natural translation

$$B_f \psi[\theta](t) := D^3 \psi[\theta](t) \left[\bigotimes^3 \left(\dot{\theta}(0) + \right. \right.$$

$$\left. \left. \int_0^T \ddot{\theta}(0)\mathrm{d}t + \int_0^T \int_0^t f(\theta(\tau))\mathrm{d}\tau \mathrm{d}t \right) \right]$$

$$+ D^2 \psi[\theta](t) \left[\left(\ddot{\theta}(0) + \int_0^T f(\theta(t))\mathrm{d}t \right) \otimes \right.$$

$$\left. \left(\dot{\theta}(0) + \int_0^T \ddot{\theta}(0)\mathrm{d}t + \int_0^T \int_0^t f(\theta(\tau))\mathrm{d}\tau \mathrm{d}t \right) \right]$$

$$+ 2D^2 \psi[\theta](t) \left[\left(\dot{\theta}(0) + \int_0^T \ddot{\theta}(0)\mathrm{d}t + \right. \right.$$

$$\left. \left. \int_0^T \int_0^t f(\theta(\tau))\mathrm{d}\tau \mathrm{d}t \right) \otimes \left(\ddot{\theta}(0) + \int_0^T f(\theta(t))\mathrm{d}t \right) \right]$$

$$+ \nabla \psi[\theta](t)f(\theta(t))$$