Weighted Composition Operators for Learning Nonlinear Dynamics

Benjamin P. Russo * **Daniel A. Messenger** ** **David Bortz** *** **Joel A. Rosenfeld** ****

* *Oak Ridge National Laboratory, Computer Science and Mathematics Division, Oak Ridge, TN 37830*
** *Department of Applied Mathematics, University of Colorado, Boulder, CO 80309*
*** *Department of Applied Mathematics, University of Colorado, Boulder, CO 80309*
**** *(Corresponding Author) Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620*

Abstract: Operator theoretic methods in dynamical system have been dominated by the use of Koopman operators and their continuous time counterparts, such as Koopman Generators and Liouville Operators. The advantage gained from their use primarily stems from the ability to extract subspaces and eigenfunctions within a space of observables that are invariant with respect to the Koopman operator over that space. When this occurs, a dynamic mode decomposition of the systems state provides a linear model for the dynamical system.
Not all Koopman operators have eigenfunctions that may be exploited in this manner. However, the framework can still be leveraged for approximations using other operators. In this setting, we present a different operator for the study of dynamical systems, the weighted composition operator. These operators are compact for a wide range of dynamics and spaces, and through their interactions with occupation kernels and vector valued kernels, they admit an estimation of the underlying dynamics.
This manuscript presents a new algorithm for the data driven study of dynamical systems from data, and also provides two numerical experiments where convergence is achieved as a proof of concept.

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

1. INTRODUCTION

Operator theoretic methods in learning dynamical systems has been dominated by the use of the Koopman operator since Mezic, Rowley, and others made the connection between Dynamic Mode Decompositions (introduced in Schmid (2010)) and the Koopman operator (cf. Budišić et al. (2012) and Kevrekidis et al. (2016)). Rosenfeld et al. (2019a) developed an approach to learning dynamical systems using Koopman generators, or more generally Liouville operators, which leverages a strong relationship between observables representing trajectories corresponding to an unknown dynamical system and Liouville operators corresponding to that system. In principle, this relationship follows from the classical chain rule and the fundamental theorem of calculus combined with notions from reproducing kernel Hilbert spaces (RKHSs), and through this combination of relationships, Rosenfeld et al. (2019a) was able to achieve a point-wise convergent approximation of the flow field over a designated compact set, given a rich enough collection of data Rosenfeld et al. (2019b).

The present manuscript aims to achieve point-wise convergence to the flow field of an unknown dynamical system from observed trajectory data through the use of a different operator, the weighted composition operator (cf. Rosenfeld et al. (2022)). This operator uses a collection of test functions to leverage integration by parts to study unknown dynamical systems, analogizing the weak SINDy algorithm (cf. Messenger and Bortz (2021b), Messenger and Bortz (2021a)). As has been demonstrated, a specific class of weightings provides substantial robustness with regard to noise corrupted data in the weak SINDy algorithm (cf. Messenger and Bortz (2022)). This operator will be approximated through the use of finite rank operators that are constructed from the observed data and functional relationships between the weighted composition operators and occupation kernels.

A weighted composition operator is given formally as $W_{f,\phi} g = (g \circ \phi) \cdot f$. The principle advantage of using weighted composition operators is that they have been well studied over a variety of function spaces, including RKHSs, and for a large collection of weighted composition operators, the relationships required on f and ϕ that gives rise to a compact weighted composition operator have been established (Ueki, 2007). For example, if $\phi(x) = ax$ with $|a| < 1$, and f a polynomial then $W_{f,\phi}$ is compact over the Fock space and the Hardy space (Ueki, 2007). This suggests that the operator $W_{f,\phi}$ is approximable by finite rank operators for a rich collection of multiplication symbols, f, which is desirable for the estimation of unknown continuous f.

This manuscript will study a generalization of weighted composition operators for the estimation of an unknown dynamical system to that corresponding to vector valued RHKSs (cf. Paulsen and Raghupathi (2016)). Section 2 will give a brief review of vector valued RHKSs and occupation kernels. Section 3 will introduce a new weighted composition operator that corresponds with vector valued symbols, and it will present relationships between this new operator and a variety of observables in a RKHSs. Section 4 will present a construction of a finite rank approximation of weighted composition operators from observed trajectory data, which will leveraged ideas from Messenger and Bortz (2021b). Finally, Section 5 will perform numerical experiments and compare the results with that of Liouville operators, which will be discussed in Section 6.

2. VECTOR VALUED RKHS AND OCCUPATION KERNELS

Vector valued RKHSs are Hilbert spaces of functions that map a domain to a Hilbert space, and were introduced to the learning community through Micchelli and Pontil (2005). Note that in this manuscript, we are only considering real valued function spaces.

Definition 1. Let X be a set and $\mathcal{Y}$ be a Hilbert space. Let H be a Hilbert space of functions that map X into $\mathcal{Y}$. H is called a vector valued RKHS if for every $x \in X$ and $\nu \in \mathcal{Y}$ the mapping $g \mapsto \langle g(x), \nu\rangle_{\mathcal{Y}}$ is a bounded linear functional over H.

The Reisz representation theorem implies that for each $x \in X$ and $\nu \in \mathcal{Y}$ there is a function $K_{x,\nu} \in H$ for which $\langle g(x), \nu\rangle_{\mathcal{Y}} = \langle g, K_{x,\nu}\rangle_H$. Note that the mapping $\nu \mapsto K_{x,\nu}$ is a bounded linear map from $\mathcal{Y}$ to H, where boundedness follows from the closed graph theorem. Hence, for each $x \in X$ we can define an operator $K_x : \mathcal{Y} \to H$ as $K_x\nu = K_{x,\nu}$. The adjoint of the operator K_x is the evaluation map, where $K_x^* g = g(x)$ for all $g \in H$, which follows from the definition of $K_{x,\nu}$ with regards to the inner products.

Definition 2. The operator valued kernel function associated with a vector valued RKHS is given as $K(x, y) = K_x^* K_y$ and is an operator from $\mathcal{Y}$ to itself for each pair $x, y \in X$.

Example 3. In the case where $\mathcal{Y} = \mathbb{R}^n$, $K(x, y)$ is a square matrix of size n.

Example 4. Let $\tilde{H}$ be a scalar valued RKHS over a set X with kernel function $k(x, y)$. Let $H = \tilde{H}^n$ for $n \in \mathbb{N}$. Take $g = (g_1 \cdots g_n)^T$ and $h = (h_1 \cdots h_n)^T$ be two elements in H, and define the inner product between them as $\langle g, h\rangle_H = \sum_{j=1}^n \langle g_j, h_j\rangle_{\tilde{H}}$. The operator valued kernel function corresponding to H is given as $K(x, y) = k(x, y)I_n$ where I_n is the $n \times n$ identity matrix.

A core component of the following work is the concept of an occupation kernel. Occupation kernels embed a continuous signal into a RKHS. Occupation kernels first appeared in the literature via Rosenfeld et al. (2019a) and are a generalization of occupation measures (see Lasserre et al. (2008)) to RKHSs . Vector valued RKHSs first appeared in the literature in Micchelli and Pontil (2005), and we define them below.

Definition 5. Let $\theta : [0, T] \to \mathbb{R}^n$ be a bounded measurable signal. Let H be a vector valued RKHS from a $\mathbb{R}^n$ to $\mathbb{R}^n$. The occupation kernel $\Gamma_{\theta,\nu} \in H$ is the representative of the bounded linear functional $g \mapsto \left\langle \int_0^T g(\theta(t))dt, \nu \right\rangle_{\mathbb{R}^n} = \langle g, \Gamma_{\theta,\nu}\rangle_H$.

Just as with the kernel functions, the mapping $\nu \mapsto \Gamma_{\theta,\nu}$ is a bounded linear mapping from $\mathbb{R}^n$ to H. Hence, we may define an operator $\Gamma_\theta : \mathbb{R}^n \to H$ as $\Gamma_\theta \nu = \Gamma_{\theta,\nu}$.

Example 6. If H is defined as in 1 with $\mathcal{Y} = \mathbb{R}^n$ and $X = \mathbb{R}^n$, then $\Gamma_\theta = \tilde{\Gamma}_\theta I_n$, where $\tilde{\Gamma}_\theta$ is the occupation kernel for the scalar valued RKHS, and $\tilde{\Gamma}_\theta(x) = \int_0^T k(x, \theta(t))dt$.

3. WEIGHTED COMPOSITION OPERATORS CORRESPONDING TO VECTOR VALUED SYMBOLS

This section gives a generalization of weighted composition operators that correspond to vector valued symbols. This definition is also similar to that of the multiplication operator defined in Wendland (2004) for the study of nonlinear control affine systems. One significant different in the definition given here is that operator is mapping a space of vector valued functions of a single scalar value to scalar valued functions of several variables, which will be important later when relationships between these weighted composition operators and occupation kernels are established.

Definition 7. Let H be a vector valued RKHS consisting of functions mapping $\mathbb{R}$ to a Hilbert space $\mathcal{Y}$ and let $\tilde{H}$ be a scalar valued RKHS over $\mathbb{R} \times \mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathcal{Y}$ and $\phi : \mathbb{R} \to \mathbb{R}$. Define the weighted composition operator, $W_{f,\phi} : \mathcal{D}(W_{f,\phi}) \to \tilde{H}$, with symbols f and ϕ as $W_{f,\phi}g(t, x) = \langle g \circ \phi(t), f(x)\rangle_{\mathcal{Y}}$ for all $x \in \mathbb{R}$ where $g \in \mathcal{D}(W_{f,\phi}) := \{g \in H : \langle g \circ \phi(\cdot), f(\cdot)\rangle_{\mathcal{Y}} \in \tilde{H}\}$.

Note that $W_{f,\phi}$ is a closed operator, given this particular domain $\mathcal{D}(W_{f,\phi})$. When $W_{f,\phi}$ is densely defined, this means that the adjoint is also densely defined and closed. In this setting, we can determine several functions in $\tilde{H}$ in the domain of $W_{f,\phi}^*$.

Proposition 8. In the setting of Definition 7 and setting K as the operator valued kernel of H and $\tilde{k}$ as the scalar valued kernel of $\tilde{H}$, then $\tilde{k}_{t,x} \in \mathcal{D}(W_{f,\phi}^*)$ and $W_{f,\phi}^* \tilde{k}_{(t,x)} = K_{\phi(t)} f(x)$.

Proposition 9. In the setting of Definition 7 and setting $X = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}^n$, $\theta : [0, T] \to \mathbb{R}^n$ is a bounded measurable signal, then $\tilde{\Gamma}_\theta \in \mathcal{D}(W_{f,\phi}^*)$ and $W_{f,\phi}^* \tilde{\Gamma}_\theta(t) = \int_0^T K(t, \phi(\tau)) f(\theta(\tau))d\tau$ for all $x \in \mathbb{R}^n$.

Examples of this operator being compact can be quickly determined when $H = \tilde{H}^n$ for appropriate $\tilde{H}$, since the norm of functions $g = (g_1 \cdots g_n)^T \in H$ are given as $\|g\|_H = \sqrt{\sum_{j=1}^n \|g_j\|_{\tilde{H}}^2}$. Hence, if f is a vector valued function where each component corresponds to a polynomial, then $W_{f,\phi}$ is a compact operator when $\phi(x) = ax$ with $|a| < 1$ and $\tilde{H}$ is the Fock space.

Henceforth, we will assume $W_{f,\phi}$ is a compact operator over a given RKHS, and we will take $\phi(t) = at$. In this context, we will write simply $W_{f,\phi} = W_{f,a}$ as an abuse of notation.

4. LEARNING NONLINEAR DYNAMICS VIA WEIGHTED COMPOSITION OPERATORS

In this section we describe a methodology similar to that of Dynamic Mode Decomposition as presented in Rosenfeld et al. (2022) to attain an approximation of the dynamics of an unknown dynamical system over a compact subset of $\mathbb{R}^n$ from observed trajectory data.

Suppose that $\{\gamma_i : [0, T_i] \to \mathbb{R}^n\}_{i=1}^M$ is a collection of observed trajectories corresponding to a dynamical system, $\dot{\gamma}_i = f(\gamma_i)$. Our objective is to leverage these trajectories to gain an estimation of $f = (f_1 \cdots f_n)^T$. We will further assume that, given a vector valued RKHS, H, consisting of continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}^n$ and a scalar valued space of continuously differentiable functions, $\tilde{H}$, over $\mathbb{R} \times \mathbb{R}^n$, the weighted composition operator, $W_{f,a}$, is compact for $|a| < 1$. We will further assume that the constant function $g_\ell(t) \equiv e_\ell$ (the standard basis element of $\mathbb{R}^n$) is in H.

One of the key ideas here is that $W_{f,a} g_\ell(0, x) = \langle e_\ell, f(x)\rangle_{\mathbb{R}^n} = f_\ell(x)$ for $\ell = 1, 2, \ldots, n$. Hence, if a sufficiently good estimate, $\hat{W}_{f,a}$, of $W_{f,a}$ is determined, then $\hat{W}_{f,a} g_\ell(0, x) \approx W_{f,a} g_\ell(0, x) = f_\ell(x)$, and thus, an estimate of f_ℓ is determined for each ℓ. More concretely $|\hat{W}_{f,a} g_\ell(0, x) - f(x)| = |\hat{W}_{f,a} g_\ell(0, x) - W_{f,a} g_\ell(0, x)| \leq \|\hat{W}_{f,a} - W_{f,a}\| \|g_\ell\|_H \sqrt{k((0, x), (0, x))}$. Since k is continuous, the term $\sqrt{k((0, x), (0, x))}$ may be bounded uniformly over a compact set and finite time, and $\|g_\ell\|_H$ is a fixed quantity. Thus, the quality of the approximation is controlled by how well $\hat{W}_{f,a}$ approximates $W_{f,a}$. The compactness here is important, since our $\hat{W}_{f,a}$ will be finite rank operators arising from our observed trajectories.

The Dynamic Mode Decomposition with respect weighted composition operators obtains a finite rank approximation of the weighted composition operator, and then the function f is then decomposed with respect to the modes determined from projections onto the right singular vectors of the resultant finite rank linear mapping.

The finite rank approximation of $W_{f,a}$ is determined first from the action of its adjoint on the occupation kernels in the scalar valued space, written as $\tilde{\Gamma}_{(\tau,\theta(\tau))}$ for a bounded measurable signal $\theta : [0, T] \to \mathbb{R}^n$. For clarity, for each $h : \mathcal{R} \times \mathcal{R}^n \to \mathbb{R}^n$ in $\tilde{H}$, $\langle h, \tilde{\Gamma}_{(\tau,\theta(\tau))}\rangle_{\tilde{H}} = \int_0^T h(\tau, \theta(\tau)) d\tau$.

Proposition 10. Suppose that $W_{f,\phi} : \mathcal{D}(W_{f,\phi}) \to \tilde{H}$ is a densely defined weighted composition operator from a vector valued RKHS, H, consisting of continuously differentiable functions mapping $\mathbb{R}$ to $\mathbb{R}^n$ and suppose that $\tilde{H}$ is a scalar valued RKHS over $\mathbb{R} \times \mathbb{R}^n$ consisting of continuously differentiable functions. Let $\gamma : [0, T] \to \mathbb{R}^n$ such that $\dot{\gamma} = f(\gamma)$, then $W_{f,a}^* \tilde{\Gamma}_{(\tau,\gamma(\tau))}(t) = \int_0^T K(t, \phi(\tau)) f(\gamma(\tau)) d\tau \in H$

Proof. Consider the functional

$$g \mapsto \int_0^T \langle g(\phi(t)), f(\gamma(t))\rangle_{\mathbb{R}^n} dt = \int_0^T \langle g(\phi(t)), \dot{\gamma}(t)\rangle_{\mathbb{R}^n} dt$$
$$= \langle g(\phi(T)), \gamma(T)\rangle_{\mathbb{R}^n} - \langle g(\phi(0)), \gamma(0)\rangle_{\mathbb{R}^n}$$
$$- \int_0^T \langle g'(\phi(t))\phi'(t), \gamma(t)\rangle_{\mathbb{R}^n} dt. \quad (1)$$

This mapping is bounded by similar arguments to those found in Rosenfeld et al. (2019b). Consequently, $\Gamma_\gamma \in \mathcal{D}(W_{f,\phi}^*)$ and the evaluation of $W_{f,\phi}^* \Gamma_\gamma$ at t is realized through the inner product with $K_{t,v}$ for $t \in \mathbb{R}$ and $\nu \in \mathcal{Y}$.

The finite rank approximation, $\hat{W}_{f,a}$, is determined by selecting a finite dimensional subspace, α_ℓ, of H and another, β, in $\tilde{H}$. The selection of α_ℓ plays a similar role to the selection of test functions in Messenger and Bortz (2021b). Indeed, any convenient observable in the Hilbert space H may be selected for this role with the caveat that the inner product with each of these functions is computable. Here $\alpha = span\{K_{t_1,e_\ell}, K_{t_1,e_\ell}, \ldots, K_{t_M,e_\ell}\}$ for $t_1 < t_2 < \ldots < t_M$, and $\beta = span\{\tilde{\Gamma}_{\gamma_1}, \ldots, \tilde{\Gamma}_{\gamma_M}\}$, and $\hat{W}_{f,a,\ell} = P_\beta W_{f,a} P_\alpha$. The selection of β is more constrained, where we aim to utilize integration by parts inside the integral provided by the occupation kernels. Here we select a diagonal kernel operator for H, since this will result in block diagonal matrices, each corresponding to a different dimension. This will reduce the required storage for the overall approximation considerably and allow the treatment the approximation of f to occur one dimension at a time. The matrix representation of $\hat{W}_{f,a,\ell}$ over the finite dimensional subspace α is given as

$$[\hat{W}_{f,a,\ell}]_{\alpha_\ell}^{\beta} =$$
$$\begin{pmatrix} \langle \tilde{\Gamma}_{\gamma_1(\tau)}, \tilde{\Gamma}_{\gamma_1(\tau)}\rangle_{\tilde{H}} & \cdots & \langle \tilde{\Gamma}_{\gamma_1(\tau)}, \tilde{\Gamma}_{\gamma_M(\tau)}\rangle_{\tilde{H}} \\ \vdots & \ddots & \vdots \\ \langle \tilde{\Gamma}_{\gamma_M(\tau)}, \tilde{\Gamma}_{\gamma_1(\tau)}\rangle_{\tilde{H}} & \cdots & \langle \tilde{\Gamma}_{\gamma_M(\tau)}, \tilde{\Gamma}_{\gamma_M(\tau)}\rangle_{\tilde{H}} \end{pmatrix}^{-1}$$
$$\times$$
$$\begin{pmatrix} \langle \hat{W}_{f,a} K_{t_1,e_\ell}, \tilde{\Gamma}_{\gamma_1(\tau)}\rangle_{\tilde{H}} & \cdots & \langle \hat{W}_{f,a} K_{t_M,e_\ell}, \tilde{\Gamma}_{\gamma_1(\tau)}\rangle_{\tilde{H}} \\ \vdots & \ddots & \vdots \\ \langle \hat{W}_{f,a} K_{t_1,e_\ell}, \tilde{\Gamma}_{\gamma_M(\tau)}\rangle_{\tilde{H}} & \cdots & \langle \hat{W}_{f,a} K_{t_M,e_\ell}, \tilde{\Gamma}_{\gamma_M(\tau)}\rangle_{\tilde{H}} \end{pmatrix}.$$

Note that

$$\langle \tilde{\Gamma}_{(\tau,\gamma_i(\tau))}, \tilde{\Gamma}_{(\tau,\gamma_j(\tau))}\rangle_{\tilde{H}} =$$
$$\int_0^{T_i} \int_0^{T_j} \tilde{K}((t, \gamma_i(t)), (\tau, \gamma_j(\tau))) d\tau dt,$$

where $\tilde{K}$ is the user selected kernel of $\tilde{H}$, and the double integral can be estimated using numerical methods. The terms $\langle \hat{W}_{f,a,\ell} K_{t_i,e_\ell}, \tilde{\Gamma}_{(\tau,\gamma_j(\tau))}\rangle_{\tilde{H}} = \langle W_{f,a} K_{t_i,e_\ell}, \tilde{\Gamma}_{(\tau,\gamma_j(\tau))}\rangle_{\tilde{H}}$, since P_β is self adjoint and acts as the identity on β, and P_α also acts as the identity on α. Moreover,

$$\langle W_{f,a} K_{t_i,e_\ell}, \tilde{\Gamma}_{(\tau,\gamma_j(\tau))}\rangle_{\tilde{H}} =$$
$$\langle K_{t_i,e_\ell}(aT_j), \gamma_j(T_j)\rangle_{\mathbb{R}^n} - \langle K_{t_i,e_\ell}(a0), \gamma_j(0)\rangle_{\mathbb{R}^n}$$
$$- \int_0^{T_j} \langle \frac{d}{d\tau} K_{t_1,e_\ell}(a\tau), \gamma_j(\tau)\rangle_{\mathbb{R}^n} d\tau$$
$$= k(t_i, aT_j)\langle e_\ell, \gamma_j(T_j)\rangle_{\mathbb{R}^n} - k(t_i, 0)\langle e_\ell, \gamma_j(0)\rangle_{\mathbb{R}^n}$$
$$- \int_0^{T_j} \frac{d}{d\tau} k(t_i, a\tau) \langle e_\ell, \gamma_j(\tau)\rangle_{\mathbb{R}^n} dt d\tau. \quad (2)$$

The estimation of the unknown dynamics f occurs one dimension at a time through the application of $\hat{W}_{f,a,\ell}$ to the constant function, $g_\ell(x) \equiv e_\ell$. To realize this through the matrix representation given above, g_ℓ must be projected onto α. This projection may be expressed as $g_\ell = \sum_{i=1}^{M} w_i K_{t_i,e_\ell}$ where the weights, $\mathbf{w} = (w_1 \ \cdots \ w_M)^T$ are determined through the application of the inverse of Gram matrix,

$$G_\alpha = \begin{pmatrix} \langle K_{t_1,e_\ell}, K_{t_1,e_\ell}\rangle_H & \cdots & \langle K_{t_1,e_\ell}, K_{t_M,e_\ell}\rangle_H \\ \vdots & \ddots & \vdots \\ \langle K_{t_M,e_\ell}, K_{t_1,e_\ell}\rangle_H & \cdots & \langle K_{t_M,e_\ell}, K_{t_M,e_\ell}\rangle_H \end{pmatrix},$$

to the vector $(\langle g_\ell, K_{t_1,e_\ell}\rangle_H \ \cdots \ \langle g_\ell, K_{t_M,e_\ell}\rangle_H)^T$, where

$$\langle g_\ell, K_{t_i,e_\ell}\rangle_H = \langle e_\ell, e_\ell\rangle_{\mathbb{R}^n} dt = 1.$$

Hence, the estimation of f_ℓ is given as

$$\hat{f}_\ell(x) = \left(\tilde{\Gamma}_{(\tau,\gamma_1(\tau))}(t,x) \ \cdots \ \tilde{\Gamma}_{(\tau,\gamma_M(\tau))}(t,x)\right) [\hat{W}_{f,a,\ell}]^{\beta}_{\alpha_\ell} \mathbf{w}$$

for any t.

If $\varphi_{1,\ell}, \ldots, \varphi_{M,\ell}$ are right singular functions for $\hat{W}_{f,a,\ell}$ with corresponding singular values, $\sigma_{1,\ell} \ldots, \sigma_{M,\ell}$, and left singular vectors $\psi_{1,\ell}, \ldots, \psi_{M,\ell}$, then the estimation of f may be represented as $\hat{f}(x) = \left(\sum_{i=1}^{M} \langle f_\ell, \varphi_{i,\ell}\rangle_H \sigma_{i,\ell}\psi_{i,\ell}(t,x)\right)_{\ell=1}^{n}$.

5. NUMERICAL EXPERIMENTS

In this section we present two numerical experiments as a proof of concept. One for a particular instance of the Duffing oscillator and another for an instance of the Lotka-Voltera equation. For each experiment, we selected H as the RKHS of functions from $\mathbb{R}$ to $\mathbb{R}^2$ as a version of the exponential dot product space (i.e. the real valued Fock space), with the kernel function given as $K(x,y) = \exp\left(\mu x^T y\right) I_2$ with $\mu = 1/1000$. The space $\tilde{H}$ is the space corresponding to $\tilde{k}((t,x),(s,y)) = \exp\left(\mu(t \cdot s + x^T y)\right)$, which is also an exponential dot product kernel, and the same μ was used for both spaces and both experiments. For the weighted composition operators, the parameter a was selected as 0.9.

In each experiment, a lattice of initial conditions were selected and from these initial conditions Runge-Kutta 4 was used with timestep 0.05 to generate synthetic trajectory data. The occupation kernels and inner products were computed using Simpson's rule.

5.1 Experiment 1: Duffing Oscillator

The Duffing oscillator for this experiment was selected as $\ddot{x} = x - x^3$. For simulations the system state was augmented as $z = (x \ \dot{x})^T$ which results in the system

$$\dot{z} = \begin{pmatrix} z_2 \\ z_1 - z_1^3 \end{pmatrix}.$$

The initial values used in generating the trajectories shown in Figure 1 was selected from a square lattice of side length 0.25 over $[-1,1]^2$. Each trajectory was generated with RK4, using a time step of 0.05, up to time $T = 2$.

From these trajectories and the given kernel functions, the method of Section 4 was implemented. This resulted in the approximations of the first and second dimensions of f displayed in Figure 2. The point-wise errors over $[-1,1]^2$ are plotted in Figure 3.

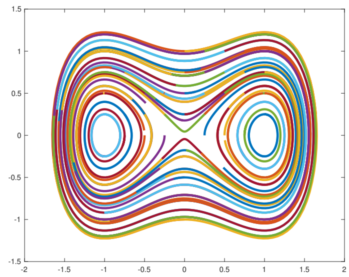

Fig. 1. Presented here are the trajectories that were leveraged for data in the approximation of the dynamics for the Duffing Oscillator. The initial values for the trajectories were selected from a square lattice over $[-1,1]$ with side length 0.25.

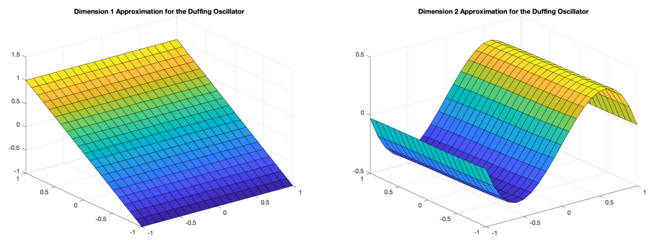

Fig. 2. This figure presents the approximations of the dynamics for the Duffing oscillator in Experiment 1.

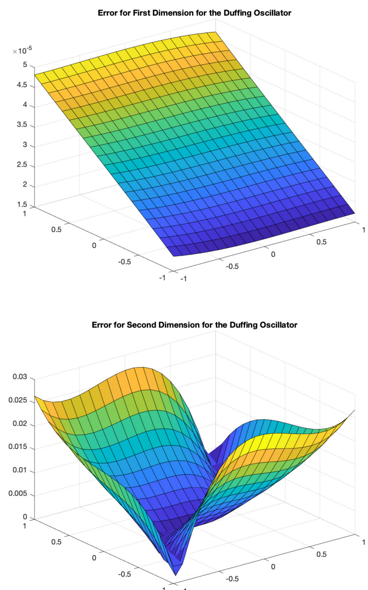

Fig. 3. This figure presents the point-wise errors produced by the approximations of the dynamics for the Duffing oscillator in Experiment 1.

5.2 Experiment 2: Lotka-Voltera Equation

The Lotka-Voltera equation used for the second experiment was

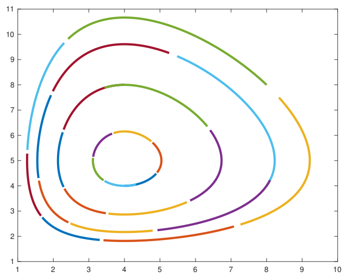

Fig. 4. Presented here are the trajectories that were leveraged for data in the approximation of the dynamics for the Lotka-Voltera equations. The initial values for the trajectories were selected as $\{(4,4)^T,(4,8)^T,(8,4)^T,(8,8)^T\}$, and each trajectory was segmented in to 6 smaller trajectories.

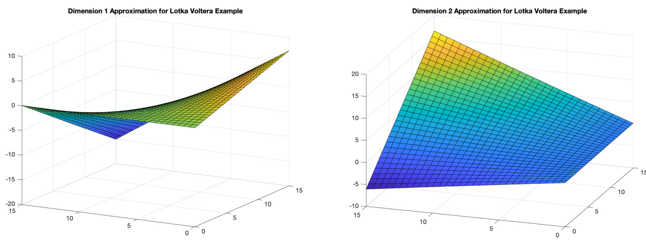

Fig. 5. This figure presents the approximations of the dynamics in Experiment 2.

$$\frac{dx}{dt} = \frac{1}{2}x - \frac{1}{10}xy; \qquad \frac{dy}{dt} = \frac{1}{10}xy - \frac{2}{5}y.$$

The initial values were $\{(4,4)^T,(4,8)^T,(8,4)^T,(8,8)^T\}$, resulting in four trajectories generated using RK4 from time $t=0$ to $T=15$. Each of these four trajectories were segmented into 6 smaller trajectories, resulting in a total of 24 trajectories used in the method of Section 4, which are displayed in Figure 4.

The approximations of the dynamics are given in Figure 5. The pointwise error in the approximation for each dimension are given in Figure 6.

6. DISCUSSION

6.1 *On convergence*

Weighted composition operators are compact for a wide range of dynamics over spaces of real analytic functions. This follows from the same proof that was leveraged for scaled Liouville operators in Rosenfeld et al. (2022). If we also have a collection of vector valued kernels in H and a complete set of occupation kernels in $\tilde{H}$, it follows that the sequence of finite rank operators created through the selection of one function at a time from these complete collections will converge to the weighted composition operator. Consequently, this means that the resultant subsequence of approximations of the dynamics will also converge pointwise (and uniformly over compact sets) to the unknown dynamics.

Complete sets of kernels can be quickly obtained for spaces of analytic functions of a single variable, since

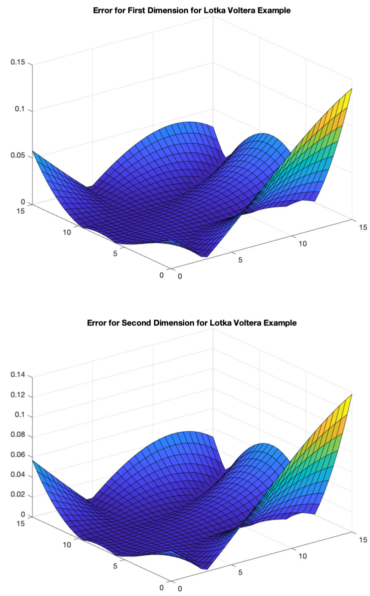

Fig. 6. This figure presents the point-wise errors produced by the approximations of the dynamics in Experiment 2. The overall shape of both errors are very similar, but not identical. This is because both dimensions are given by very similar quadratic equations in two variables. Significantly, the relative error in both cases is small.

any sequence $\{t_m\}_{m=1}^\infty$ with a limit point implies that $\{K_{t_m,e_\ell}\}_{m=1,\ell=1}^{\infty,\ell}$ is a complete set in H, and in general the collection of all kernel functions is complete within a RKHS. The completeness of the collection of all occupation kernels corresponding to a dynamical system follows from that of regular kernels, since short segments of occupation kernels approximate kernel functions centered at the initial points in norm. These considerations align with that found in Rosenfeld et al. (2022).

6.2 *Similarities with weak-SINDy Methods*

The method presented here can be seen as an operator theoretic manifestation of the weak-SINDy algorithm Messenger and Bortz (2021b). In this algorithm, data from trajectories, $\{\gamma_i : [0,T] \to \mathbb{R}^n\}_{i=1}^M$, are combined with test functions, $\varphi_j : [0,T] \to \mathbb{R}^n$, that are compactly supported in $(0,T)$. The dynamics are then parameterized as $f(x) = \sum_{m=1}^M \theta_m Y_m(x)$, and the parameters are determined through the relation $-\int_0^T \varphi_j'(t)\gamma(t)dt = \int_0^T \varphi_j(t) f(\gamma_i(t))dt = \sum_{m=1}^M \theta_m \int_0^T \varphi_j(t) Y_m(\gamma_j(t))dt$, and since the left and right hand sides are computable with the available data, the parameters $\{\theta_m\}_{m=1}^M$ may be estimated using a regression procedure Messenger and Bortz (2021b). The fundamental relation for the weak-SINDy manifests through the use of integration by parts, and the boundary terms disappear since φ_i is compactly supported.

In (2), the kernel functions in α_ℓ replace the test functions in the weak-SINDy method. Since kernel functions aren't necessarily compactly supported, the boundary terms in (2) persist. However, there are kernel functions that are compactly supported, such as the Wendland radial basis functions (cf. Wendland (2004)). Moreover, if compactly supported is relaxed to vanishing at 0 and T, then any kernel function can be adjusted to vanish at those points by replacing the kernel function, K, with $K(s,t) = t\cdot(T-t)K(s,t)s\cdot(T-s)$, and thus the boundary terms disappear for this kernel function.

6.3 Discussion of Numerical Experiments

The estimations presented in Section 4 represent a form of scattered data approximation, where instead of point samples, as in Rosenfeld et al. (2022), trajectories are used as the central unit of data. Occupation kernels are centered at each of the trajectories, and pointwise estimation near a trajectory are better than estimations further away from a trajectory. This localized estimation property is shared with that of scattered data interpolation using kernel functions. This also explains the similarity in shape of the error plots in Figures 3 and 6.

6.4 Discussion of Dynamic Modes

Since each dimension carries with it different collection of singular functions in the estimation of the dynamics, a direct description of dynamics modes (which arise from each dimension sharing the same collection of basis functions) is difficult. However, the utilization of the singular value decomposition of the finite rank operators provides a collection of basis functions arising from the interaction of the data with the weighted composition operator. The exploration of the utilization of this basis will be explored in future work.

7. CONCLUSION

This manuscript introduced weighted composition operators as a tool for the estimation of a nonlinear continuous time dynamical system. The presented method utilizes the methodology introduced in Rosenfeld et al. (2022) to obtain an estimation of the weighted composition operator corresponding to an unknown dynamical system, and subsequently the application of this approximation is itself applied to the 1 function to obtain an approximation of the unknown dynamics.

The method described in this manuscript generalizes approaches using Koopman and Liouville operators to a new collection of operators, while simultaneously providing an operator theoretic analog to weak-SINDy methods given in Messenger and Bortz (2021b).

ACKNOWLEDGEMENTS

This work was partially supported by the following grants: AFOSR Young Investigator Research Program Award FA9550-21-1-0134, NSF grant ECCS-2027976, NSF grant MCB-2054085, DOE grant DE-SC0023346, NIH grant R35GM149335, and AFOSR Award FA9550-20-1-0127. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agencies.

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the work for publication, acknowledges that the US government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the submitted manuscript version of this work, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

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

Kevrekidis, I., Rowley, C.W., and Williams, M. (2016). A kernel-based method for data-driven koopman spectral analysis. *Journal of Computational Dynamics*, 2(2), 247–265.

Lasserre, J.B., Henrion, D., Prieur, C., and Trélat, E. (2008). Nonlinear optimal control via occupation measures and lmi-relaxations. *SIAM journal on control and optimization*, 47(4), 1643–1666.

Messenger, D.A. and Bortz, D.M. (2021a). Weak sindy for partial differential equations. *Journal of Computational Physics*, 443, 110525.

Messenger, D.A. and Bortz, D.M. (2021b). Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3), 1474–1497.

Messenger, D.A. and Bortz, D.M. (2022). Asymptotic consistency of the wsindy algorithm in the limit of continuum data. *arXiv preprint arXiv:2211.16000.*

Micchelli, C.A. and Pontil, M. (2005). On learning vector-valued functions. *Neural computation*, 17(1), 177–204.

Paulsen, V.I. and Raghupathi, M. (2016). *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge university press.

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

Rosenfeld, J.A., Kamalapurkar, R., Russo, B., and Johnson, T.T. (2019a). Occupation kernels and densely defined liouville operators for system identification. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 6455–6460. doi: 10.1109/CDC40024.2019.9029337.

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

Schmid, P.J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656, 5–28.

Ueki, S.I. (2007). Weighted composition operator on the fock space. *Proceedings of the American Mathematical Society*, 135(5), 1405–1410.

Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.