Operator Approximations for Inverse Problems*

Joel A. Rosenfeld[1] and Benjamin P. Russo[2] and Rushikesh Kamalapurkar[3]

Abstract— This manuscript presents a framework for resolving inverse problems through the use of operator approximations over vector valued RKHSs. This generalizes Koopman based methods for data driven analysis and identification of dynamical systems. Three examples of this framework are presented to highlight its generality and effectiveness.

I. Introduction

Over the past decade, a variety of operator theoretic methods have been developed for the study of data driven methods in dynamical systems (cf. [1], [2], [3]). These methods cast an unknown dynamical system into a (hopefully) compact operator, and then leverages the interactions between the operator and certain observables within a Hilbert space to gain a finite rank approximation of that operator. This finite rank approximation of the infinite dimensional operator is then decomposed into its spectral decomposition, where then the full state observable is projected onto the eigenbasis, and ultimately, this results in a model for the system state.

For the Koopman and Liouville operators in particular, the eigenfunctions of these operators "observe" the state of a nonlinear system as an exponential involving the corresponding eigenvalue, and hence, a projection onto these eigenfunctions results in a linear approximation of the nonlinear system after composition with the system state.

Specifically, if $\{\varphi_i\}_{i=1}^M$ is a collection of eigenfunctions of a Liouville operator, A_f, in a Hilbert space H corresponding to eigenvalues $\{\lambda_i\}_{i=1}^M$, then if the full state observable, $g_{id} \in H$, given as $g_{id}(x) = x$, is projected onto the eigenfunctions, as $Pg_{id} = \sum_{i=1}^M \xi_i \varphi_i$, we have $x(t) \approx \sum_{i=1}^M \xi_i e^{\lambda_i t} \varphi_i(x(0))$ (cf. [3]).

To achieve convergence of the model given by the eigenfunctions, there are several operator theoretic conditions that should be met. First, if the Hilbert space is a space of L^2 functions, then convergence in this space does not imply pointwise convergence, so theoretically, there might be an uncountable number of points where this model does not converge. A RKHS helps mitigate this issue, where norm convergence implies pointwise (and often uniform) convergence ([4]). Second, it is desirable for the spectrum of the finite rank approximation to be close to that of the original operators, which can be achieved when the operators are compact.

It was shown in [4] and [3] that there are a variety of different adjustments that can be made to Liouville operators to achieve compactness of the operator. One of these methods is first composing with another function before the application of the Liouville operator to an observable. This results in what was called a scaled Liouville operator [4].

Weighted composition operators are defined in a similar manner, and they can effectively modify an unbounded multiplication operator into a compact operator, given the right conditions (cf. [5]).

Here we are going to generalize the modeling framework used for studying dynamical systems to other contexts and with different operators. To do this properly, we will use RKHSs, and select our operators to be compact. First, we will discuss vector valued RKHSs, which will allow us to give the presentation in the greatest generality, and we will then give that general framework for the methodology. The remaining sections will give examples of this framework, both for data driven methods in dynamical systems as well as for other inverse problems.

II. Vector Valued Reproducing Kernel Hilbert Spaces

In this section we give a review of vector valued RKHSs. A more complete coverage may be found in [6].

Definition 1: A vector valued RKHS from a set X to a Hilbert space $\mathcal{Y}$ is a Hilbert space, H, of functions with domain X and co-domain $\mathcal{Y}$, for which given any $x \in X$ and $\nu \in \mathcal{Y}$, the mapping $h \to \langle h(x), \nu\rangle_{\mathcal{Y}}$ is a bounded linear functional.

Just as the Reisz theorem provides for the reproducing kernels corresponding to a point $x \in X$, given any $x \in X$ and $\nu \in \mathcal{Y}$, the Reisz theorem guarantees that there is a function, $K_{x,\nu}$ in H such that $\langle h(x), \nu\rangle_{\mathcal{Y}} = \langle h, K_{x,\nu}\rangle_H$.

Note that if $\mathcal{Y}$ is $\mathbb{R}^n$, then ν may be selected as a member of the standard basis. For a vvRKHS from a set X to $\mathbb{R}^n$, the evaluation operator $E_x : H \to H$, given as $E_x h = h(x)$ is a bounded linear operator. The concept of the evaluation operator can be extended to general vvRKHSs. In fact, the mapping $\nu \to K_{x,\nu}$ is

*This work was partially supported by the following grants: AFOSR Program Awards FA9550-21-1-0134 and FA9550-20-1-0127 and NSF awards ECCS-2027976 and 2027999. It was also supported by the DOE/NIH/NSF "Collaborative Research in Computational Neuroscience (CRCNS): Decomposing Neural Dynamics." 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.

[1](Corresponding Author) Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620 rosenfeldj@usf.edu

[2]Riverside Research, New York, NY 10038, USA. russobp@riversideresearch.org

[3]Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 rkamalapurkar@ufl.edu

linear, which means we can think of that assignment from $\mathcal{Y}$ to H as a linear operator.

Thus, we may write $K_{x,\nu}$ as $K_x\nu$ where K_x is an operator from $\mathcal{Y}$ to H. The operator, K_x is then a bounded linear operator from $\mathcal{Y}$ to H. Therefore, its adjoint is well defined over all of H and maps to $\mathcal{Y}$. Moreover, $K_x^* h = h(x)$ for all $h \in H$, and is the evaluation map.

In [6], the operator $K : X \times X \to B(\mathcal{Y})$ given as $K(x,y) = K_x^* K_y$ is called a $\mathcal{Y}$-reproducing kernel, we will frequently refer to K as an operator valued kernel.

Example 1: For a scalar valued RKHS, H, with kernel function $\tilde{K}$, we can define a vector valued RKHS H^n where the inner product between two vector valued functions, $g = \begin{pmatrix} g_1 & \cdots & g_n \end{pmatrix}^T$ and $h = \begin{pmatrix} h_1 & \cdots & h_n \end{pmatrix}^T$, is given as $\langle g, h\rangle_{H^n} = \sum_{j=1}^n \langle g_j, h_j\rangle_H$. If $v = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^T \in \mathbb{C}^n$ and $x \in X$, then

$$\begin{aligned} &\langle g(x), v\rangle_{C^n} = g_1(x)v_1 + \cdots + g_n(x)v_n \\ &= \langle g_1, v_1\tilde{K}_x\rangle_H + \cdots + \langle g_n, v_n\tilde{K}_x\rangle_H = \langle g, \tilde{K}_x v\rangle_{H^n}. \end{aligned}$$

Hence, $K_{x,v} = \tilde{K}_x v$, where $\tilde{K}_x$ is the kernel for the scalar valued RKHS. Moreover, the matrix representation of $K(x,y)$ with respect to the standard basis, is $K(x,y) = \tilde{K}(x,y)I_n$ where I_n is the identity matrix.

III. Operator Approximations for Inverse Problems

This section will be kept abstract to provide a general framework for operator decomposition methods as applied to inverse problems. This framework encompasses Koopman and Liouville based Dynamic Mode Decompositions, and will also apply to a broader class of inverse problems, some involving dynamics.

In the context of this problem, we have a series of measurements, which are stored as a collection of functionals over an a priori selected vvRKHS, H, from a set X to a Hilbert space $\mathcal{Y}$ and each of these functionals are represented as $\{b_1, \ldots, b_M\} \subset H$. In a possibly different vvRKHS, $\tilde{H}$, from a set X to a Hilbert space $\mathcal{Y}$, another collection of measurements is collected as $\{d_1, \ldots, d_M\} \subset \tilde{H}$. Let $f : X \to \mathcal{Y}'$ be a model for a system that generated these measurements, where f will be treated as unknown. Let $\{a_1, \ldots, a_M\} \subset H$ be a collection of functions for which an operator will be approximated over.

Let $Q_f : H \to \tilde{H}$ be a compact linear operator for which:

1) $Q_f^* d_j = b_j$
2) Given a collection of projection operators $\{P_\ell : \mathcal{Y} \to \mathcal{Y}\}_{\ell=1}^\infty$ for which $\oplus_\ell P_\ell = I_{\mathcal{Y}}$ and a compact subset $B \subset X$, for each ℓ there is a sequence of functions $\{g_{m,\ell}\}_{m=1}^\infty$ such that $Q_f g_{m,\ell} \to P_\ell f$ pointwise over B. In the simplest case, we could have $P_1 = I_{\mathcal{Y}}$, or for an orthonormal basis $\{e_\ell\}_{m=1}^\infty$ each P_ℓ could project to each span of e_ℓ (or its coefficient).

The operator decomposition method estimates the operator Q_f with a finite rank operator, $\tilde{Q}_f = P_\beta Q_f P_\alpha$, with a matrix representation from $\alpha = span\{a_1, \ldots a_M\}$ to $\delta = span\{d_1, \ldots, d_M\}$ as

$$[\tilde{Q}_f]_\alpha^\delta = \begin{pmatrix} \langle d_1, d_1\rangle_{\tilde{H}} & \cdots & \langle d_1, d_M\rangle_{\tilde{H}} \\ \vdots & \ddots & \vdots \\ \langle d_M, d_1\rangle_{\tilde{H}} & \cdots & \langle d_M, d_M\rangle_{\tilde{H}} \end{pmatrix}^{-1} \times \begin{pmatrix} \langle Q_f a_1, d_1\rangle_{\tilde{H}} & \cdots & \langle Q_f a_M, d_1\rangle_{\tilde{H}} \\ \vdots & \cdots & \vdots \\ \langle Q_f a_1, d_M\rangle_{\tilde{H}} & \cdots & \langle Q_f a_M, d_M\rangle_{\tilde{H}} \end{pmatrix}.$$

Note that in the literal computation of $[\tilde{Q}_f]_\alpha^\delta$, the inner products on the right matrix should be $\langle P_\delta Q_f P_\alpha a_i, d_j\rangle_{\tilde{H}}$, but since $a_i \in \alpha$ we have $P_\alpha a_i = a_i$ and projections are self adjoint, which means

$$\langle P_\delta Q_f P_\alpha a_i, d_j\rangle_{\tilde{H}} = \langle Q_f P_\alpha a_i, P_\delta d_j\rangle_{\tilde{H}} = \langle Q_f a_i, d_j\rangle_{\tilde{H}},$$

since $d_j \in \delta$.

If we have a countable collection of measurements for which the corresponding b_i's and d_i's have a dense span in their respective spaces, this approximation of Q_f converges to Q_f in norm as we add the measurements to the approximation, owing to the compactness of Q_f.

Note that,

$$\begin{aligned} &\|\tilde{Q}_f g_m(x) - f(x)\|_{\mathcal{Y}} \\ &\leq \|\tilde{Q}_f g_m(x) - Q_f g_m(x)\|_{\mathcal{Y}} + \|Q_f g_m(x) - f(x)\|_{\mathcal{Y}} \\ &\leq \|K_x^*\|\|\tilde{Q}_f - Q_f\|\|g_m\| + \|Q_f g_m(x) - f(x)\|_{\mathcal{Y}}. \end{aligned}$$

Hence, for close norm approximation of Q_f by $\tilde{Q}_f$ and large enough m, $\tilde{Q}_f g_m(x)$ is close to $f(x)$. If $\|K_x^*\|$ is bounded over B, then it follows that this estimate of f can also be made uniform over B.

Of course, this is an estimate of Q_f and not an operator decomposition. The decomposition occurs when we look at either the Singular Value Decomposition of $[\tilde{Q}_f]_\alpha^\delta$ or its eigendecomposition, where the eigendecomposition is only possible when $\alpha \subset \delta$. This can happen if $\tilde{H} = H$ or when $H \subset \tilde{H}$.

Let $\tilde{\varphi}_s$ be a normalized right singular function of $\tilde{Q}_f$, corresponding to the singular value $\tilde{\sigma}_s \geq 0$ and right singular function $\tilde{\psi}_s$. Since Q_f is compact, the quantity $\|Q_f\tilde{\varphi}_s(x) - \tilde{\sigma}_s\tilde{\psi}_s(x)\|_{\mathcal{Y}} \leq \|K_x^*\|\|Q_f - \tilde{Q}_f\|$ can be made small for sufficiently close estimate with $\tilde{Q}_f$. In other words, $\tilde{\varphi}_m$ behaves point-wise closely to a singular function of Q_f when $\tilde{Q}_f$ closely estimates Q_f. Analogous statements can be derived for eigenfunctions of $\tilde{Q}_f$ without any adjustment.

Remark: In the setting of DMD, this inequality suggests that for normalized eigenfunctions of A_f, $|\frac{d}{dt}\varphi(x) - \lambda\varphi(x)| < \varepsilon$ when $\|A_f - \tilde{A}_f\| < \varepsilon$ and K is the Gaussian RBF. In turn, this means that $\varphi(x(t)) \approx \varphi(x(0))e^{\lambda t}$, which was leveraged for the reconstruction formula in [3].

The singular functions or eigenfunctions of $\tilde{Q}_f$ represent a feature extraction based on the available data, the form of the operator Q_f, and the selected Hilbert spaces. This is similar to how PCA and SVDs work in typical data science applications.

To complete the approximation via the spectral decomposition of $\tilde{Q}_f$, we project g_m onto the span of a collection of right singular functions of $\tilde{Q}_f$, $\{\varphi_1, \ldots, \varphi_M\}$, as $P_S g_s = \sum_{s=1}^M \xi_s \varphi_s$, where ξ_m are the operator modes corresponding to the data, and $S = span\{\varphi_1, \ldots, \varphi_M\}$. The estimate of f is then obtained as

$$f(x) \approx Q_f g_m(x) \approx \tilde{Q}_f g_m(x) = \sum_{s=1}^{M} \xi_s \sigma_s \psi_m(x).$$

The last equality follows, since $\tilde{Q}_f = P_\beta Q_f P_\alpha$, and $S = \beta$ but with a different basis. The estimations in the equation above can be quantified by selecting sufficiently large m and with sufficiently rich data so that $\tilde{Q}_f \approx Q_f$.

The last computational challenge in the implementation of this method is to apply the finite rank approximation to $g_{m,\ell}$. Since the matrix is defined to be acting on the α basis, $g_{m,l}$ must first be projected onto that basis before the application of the matrix to achieve the approximation. This means we are looking for the weights that satisfy $P_\alpha g_{m,\ell} = \sum_{j=1}^M w_j a_j$, which is the closest element of α to $g_{m,\ell}$. These weights can be determined as

$$\vec{w} = \begin{pmatrix} \langle a_1, a_1\rangle_H & \cdots & \langle a_1, a_M\rangle_H \\ \vdots & \ddots & \vdots \\ \langle a_M, a_1\rangle_H & \cdots & \langle a_M, a_M\rangle_H \end{pmatrix}^{-1} \begin{pmatrix} \langle g_{m,\ell}, a_1\rangle_H \\ \vdots \\ \langle g_{m,\ell}, a_M\rangle_H \end{pmatrix}.$$

Hence the approximation manifests from the matrix representation as $\vec{u} = [\tilde{Q}_f]_\alpha^\delta \vec{w}$, where $\vec{u}$ is a vector of components that, when placed with the basis δ, yield the approximation of f as $\sum_{j=1}^M u_j d_j$.

The above framework generalizes the DMD methodology to a broader collection of inverse problems, beyond but including dynamical systems. The next several sections will examine some new and some extant decomposition methods, and extol the connections between the above framework and the applications.

IV. Function Approximation with Weighted Composition Operators

In this section, we will demonstrate how the operator decomposition method could be leveraged for function approximation via point samples,

$$\{(x_1, f(x_1)), \ldots, (x_M, f(x_M))\}.$$

The operator we will using is the weighted composition operator. Point samples will be represented by the corresponding reproducing kernels centered at the samples, which aligns with scattered data approximation. The difference between scattered data approximation as typically done with RBFs in [7], is that this is frequently done using interpolation methods, which yields a projection onto the span of the kernels.

Definition 2: Let H be a vvRKHS from a set X to a Hilbert space $\mathcal{Y}$ and $\tilde{H}$ a scalar valued RKHS over X, and let $f : X \to \mathcal{Y}$ and $\phi : X \to X$ (note that ϕ is used here for the composition symbol as opposed to φ). The weighted composition operator $W_{f,\phi} : \mathcal{D}(W_{f,\phi}) \to \tilde{H}$, with $\mathcal{D}(W_{f,\phi}) := \{g \in H : \langle g(\phi(\cdot), f\rangle_{\mathcal{Y}} \in \tilde{H}\}$, is given as $W_{f,\phi} g = \langle g(\phi(\cdot)), f\rangle_{\mathcal{Y}}$.

Let $\{e_\ell\}_{\ell=1}^\infty$ be an orthonormal basis for $\mathcal{Y}$. For the purpose of this section, we will assume that the constant function $1_\ell(x) \equiv e_\ell$ is in the domain of $W_{f,\phi}$. Hence,

$$W_{f,\phi} 1_\ell(x) = \langle f(x), e_\ell\rangle_{\mathcal{Y}}.$$

Which fits the projection form on the methodology given in Section III.

We will further assume that $W_{f,\phi}$ is compact, which holds when $\mathcal{Y} = \mathbb{R}^n$, $H = F^2(\mathbb{R}^n)^n$, each component of f is a polynomial, and $\phi(x) = ax$ with $|a| < 1$.

The adjoints of weighted composition operators interact nicely with kernel functions, as $\langle W_{f,\phi} g, \tilde{K}_x\rangle_{\tilde{H}} = \langle g(\phi(x)), f(x)\rangle_{\mathcal{Y}} = \langle g, K_{\phi(x)} f(x)\rangle_H$ for all $g \in H$. Hence, $W_{f,\phi}^* \tilde{K}_x = K_{\phi(x)} f(x)$.

For each x_i, $\phi(x_i)$ is known, since ϕ is user selected, and $f(x_i)$ is known through measurement. Therefore, only the operator is unknown in $W_{f,\phi}^* \tilde{K}_{x_i} = K_{\phi(x)} f(x_i)$ and this can be treated as a sample of the operator. Let $d_i = \tilde{K}_{x_i}$ and $b_i = K_{\phi(x_i)} f(x_i)$, and $a_{i,j} = K_{x_i, e_j}$ in the above framework with corresponding α and δ subspaces.

V. Approximating Flow Fields

In the setting of learning an unknown dynamical system from data, we will consider the data as a collection of observed trajectories, $\{\gamma_i : [0, T_i] \to \mathbb{R}^n\}_{i=1}^M$, each satisfying the differential equation $\dot{\gamma}_i(t) = f(\gamma_i(t))$ for $t \in [0, T_i]$, for an unknown $f : \mathbb{R}^n \to \mathbb{R}^n$.

Definition 3: Given a RKHS, H, mapping $\mathbb{R}^n$ to $\mathbb{R}^n$, the operator in question here is the scaled Liouville operator, $A_{f,a} : \mathcal{D}(A_{f,a}) \to H$, given as $A_{f,a} g(x) = aDg(ax)f(x)$ where Dg is the matrix valued derivative of the vector valued observable $g \in \mathcal{D}(A_{f,a}) := \{h \in H : aDh(ax)f(x) \in H\}$.

Motivated by [4], it will be assumed here that scaled Liouville operators are compact for the selected parameters, dynamics, and Hilbert spaces.

Let $g_{id}(x) = x$ which is known as the identity function or the full state observable. Note that $Dg_{id}(x) = I_n$ for all x and

$$\frac{1}{a} A_{f,a} g_{id}(x) = f(x).$$

Since $A_{f,a}$ is compact, so is $\frac{1}{a} A_{f,a}$, and this latter operator is the Q_f of this section.

The functionals in question here are occupation kernels. For a given bounded measurable signal $\theta : [0, T] \to \mathbb{R}^n$ the occupation kernel corresponding to θ and $\nu \in \mathbb{R}^n$ within a RKHS, H, is the unique function, $\Gamma_{\gamma,\nu}$ for which $\langle g, \Gamma_{\gamma,\nu}\rangle_H = \langle \int_0^T g(\theta(t))dt, \nu\rangle_{\mathbb{R}^n}$. Here we see that $\nu \mapsto \Gamma_{\gamma,\nu}$ is linear, and as such for each bounded measurable $\theta : [0, T] \to \mathbb{R}^n$ there is an operator $\Gamma_\theta : \mathbb{R}^n \to H$ such that $\Gamma_\theta \nu = \Gamma_{\theta,\nu}$ for all $\nu \in \mathbb{R}^n$. Note that $\langle \Gamma_{\theta,\nu}(x), \omega\rangle_{\mathbb{R}^n} = \langle \Gamma_\theta \nu, K_x \omega\rangle_H = \langle \int_0^T K_{x,\omega}(\theta(t))dt, \nu\rangle_{\mathbb{R}^n} = \langle \int_0^T K_{\theta(t)}^* K_x \omega dt, \nu\rangle_{\mathbb{R}^n}$. Thus, if $K(x,y) = \tilde{K}(x,y)I_n$, where $\tilde{K}$ is the kernel for a scalar valued RKHS, $\Gamma_\theta(x) = K_x^*\Gamma_\theta = \int_0^T \tilde{K}(x,\theta(t))dtI_n$. That is, the vector valued RKHS for this vector valued RKHS is a scalar valued occupation kernel times the identity matrix.

The adjoint of the scaled Liouville operator corresponding to the dynamics f applied to Γ_γ where $\dot{\gamma} = f(\gamma)$ can be expressed in terms of a difference of kernels. This is demonstrated quickly by examining, for arbitrary $g \in H$, $\langle A_{f,a} g, \Gamma_{\gamma,\nu}\rangle_H = \langle \int_0^T aDg(a\gamma(t))f(\gamma(t))dt, \nu\rangle_{\mathbb{R}^n} = \langle \int_0^T Dg(a\gamma(t))a\dot{\gamma}(t)dt, \nu\rangle_{\mathbb{R}^n} = \langle g(a\gamma(T)) - g(a\gamma(0)), \nu\rangle_{\mathbb{R}^n} = \langle g, (K_{a\gamma(T)} - K_{a\gamma(0)})\nu\rangle_H$. Hence, for each $\nu \in \mathbb{R}^n$ we have $A_{f,a}^*\Gamma_\gamma \nu = (K_{a\gamma(T)} - K_{a\gamma(0)})\nu$, hence $A_{f,a}^*\Gamma_\gamma = K_{a\gamma(T)} - K_{a\gamma(0)}$, where $A_{f,a}^*\Gamma_\gamma : \mathbb{R}^n \to H$ is a bounded operator.

VI. Approximating Flow Fields with Weighted Composition Operators

The operator in Section V is a natural generalization of the Liouville operator or Koopman generator that is used in DMD [8], [4], but which is also a compact operator, allowing for convergence guarantees based on the density of the occupation kernels within a RKHS. This section presents an approach for estimating flow fields that was introduced in [9], but which also is an example of this framework in execution.

Consider the weighted composition operator mapping from a space of a functions of a single time variable, H, to a space of functions of both time and the state variable, $\tilde{H}$, given formally as $W_{f,\phi} g(t,x) = \langle g(t), f(x)\rangle_{\mathbb{R}^n}$. The approach presented in Section V leverages the chain rule and the fundamental theorem of calculus which aligns well with the system identification method presented in [10]. Using the weighted composition operator gives an operator theoretic interpretation of the weak-SINDy method of [11], [12] in that integration by parts plays a central role for the interaction between the operator and the embedding of the trajectories through occupation kernels.

In particular, suppose that $\gamma : [0,T] \to \mathbb{R}^n$ is the solution to the dynamical system $\dot{x} = f(x)$, and consider $\langle W_{f,\phi} g, \Gamma_\gamma\rangle_H = \int_0^T \langle g(t), f(\gamma(t))\rangle_{\mathbb{R}^n} dt = \int_0^T \langle g(t), \dot{\gamma}(t)\rangle_{\mathbb{R}^n} dt = \langle g(T), \gamma(T)\rangle_{\mathbb{R}^n} - \langle g(0), \gamma(0)\rangle_{\mathbb{R}^n} - \int_0^T \langle \dot{g}(t), \gamma(t)\rangle_{\mathbb{R}^n} dt$.

Recognizing that $g \mapsto \langle g(T), \gamma(T)\rangle_{\mathbb{R}^n} - \langle g(0), \gamma(0)\rangle_{\mathbb{R}^n} - \int_0^T \langle \dot{g}(t), \gamma(t)\rangle_{\mathbb{R}^n} dt$ is a bounded linear functional, this may be represented, via the Riesz theorem, as a function inside of the Hilbert space, $\tilde{H}$ as Φ_γ.

This relation was leveraged in [9] to produce a method of recovering a flow field from trajectory data.

VII. Convolution Operators and Recovering Impulse Response Function

In this setting we will assume we have an unknown linear differential operator L and a collection of inputs (or forcing functions) $\{G_1, \ldots, G_M\}$ and outputs $\{y_1, \ldots, y_M\}$ satisfying $Ly_m = G_m$. The objective is to find the impulse response function for this system, h, so that given a new forcing function, G, the output y_G may be predicted.

The relation between the input and the outputs naturally falls into a function theoretic operator, such as $G \mapsto h \star G$ where $\star$ represents convolution. The operator we are looking for needs to leverage convolution in some nontrivial way.

Let $\tilde{H}$ be a RKHS of scalar valued functions from $\mathbb{R}^n$ to $\mathbb{R}$. A signal valued RKHS, H is a Hilbert space of functions that take $C([0,T], \mathbb{R}^n)$ signals to $C([0,T], \mathbb{R}^n)$ signals, and H is constructed from functions in $\tilde{H}$. Specifically, for each $\phi \in H$, there exists a unique $g \in \tilde{H}$ such that given a signal $\theta : [0,T] \to \mathbb{R}^n$, $\phi[\theta](t) = \phi_g[\theta](t) := g(\theta(t))$ for $t \in [0,T]$.

Consequently, the inner product of two elements $\phi_g, \phi_{g'} \in H$ is given as $\langle \phi_g, \phi_{g'}\rangle_H = \langle g, g'\rangle_{\tilde{H}}$. More details about signal valued RKHSs may be found in [cite].

The occupation kernel corresponding to a signal, $G : [0,T] \to \mathbb{R}^n$, is then given as the unique function, Γ_G, such that $\int_0^T \phi_g[G](t)dt = \langle \phi_g, \Gamma_G\rangle_H$. However, this means that $\int_0^T g(G(t))dt = \langle g, \Gamma_G\rangle_{\tilde{H}}$, where we use Γ_G to mean either the occupation kernel for the ordinary RKHS or the occupation kernel that takes signals as inputs in the signal valued case.

The function $\theta \mapsto g(h \star \theta(\cdot))$ is well defined as a mapping that takes continuous functions over some interval, $[0,T]$, to continous functions from $[0,T]$, given continuity of g and h. Hence, the operator, Q_h, formally defined as

$$Q_h \phi_g = g(h \star (\cdot))$$

makes sense as an operator acting on a signal valued RKHS.

In this setting, letting $a_i = d_i = \Gamma_{G_i}$ and $b_i = \Gamma_{y_i}$, we can derive a matrix representation of a finite rank approximation of Q_h. The impulse response function can then be recovered as $\langle Q_h g_{id}, \Gamma_{\tilde{\delta}(\cdot - t)}\rangle_H \approx h(t)$, where $\tilde{\delta}$ is a continuous signal estimating the delta function.

VIII. Numerical Experiments

A. Scattered Data Approximation

For this experiment the function $f(x) = \sin(x) + \cos(3x)$ was sampled at a collection of equally spaces points in the interval $[0, 2\pi]$ with spacing 0.25, and an approximation was generated via a weighted composition operator, $W_{f,\phi}$ with $\phi(x) = 0.9x$. The kernel function selected was the exponential dot product kernel, $\exp(\mu x^T y)$ with $\mu = 0.2$.

A total of 26 modes were derived from the weighted composition operator's eigendecomposition. It can be

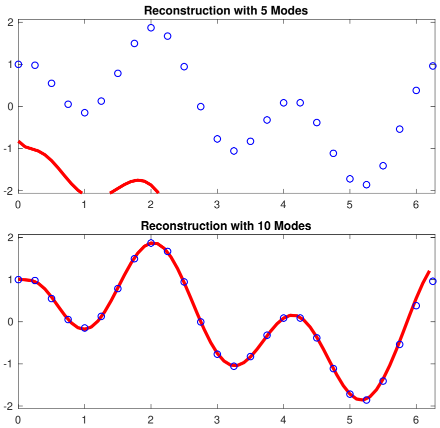

Fig. 1. A demonstration of the approximation of a function through modes determined through a data driven approximation of a weighted composition operator using the exponential dot product kernel. Out of a total of 26 modes, 10 are sufficient for determing a good approximation.

seen in Figure 1 that the function can be approximated well using 10 of the data driven modes.

B. Approximation of flow fields

In this experiment, the technique described in Section V is applied to estimate the dynamics of a damped, unforced Duffing oscillator

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -\delta\, x_2 - \beta\, x_1 - \alpha\, x_1^3, \tag{1}$$

with parameters $\alpha = 1$, $\beta = -1$, and $\delta = 0.1$.

To train the model, MATLAB ode45 solver is used to generate trajectories starting from initial conditions on a 12×12 uniform grid over $[-5, 5] \times [-5, 5]$ (144 trajectories). Each trajectory is 5 seconds long and sampled every 0.05 s. An exponential dot-product kernel with parameter $\mu = 190$ is used to define the underlying RKHS. $\lambda = 10^{-8}$. A scaling factor $s = 0.5$ is applied to ensure compactness of the scaled Liouville operator. The algorithm returns a learned vector field $\hat{f}(\cdot)$.

To evaluate the learned model, a reference trajectory is computed starting from the initial condition $x(0) = (1, 1)$. A predicted trajectory is the computed by integrating the learned dynamics $\dot{x} = \hat{f}(x)$. Figure 2 shows the true trajectory $x(\cdot 0$ and the predicted trajectory $\hat{x}(\cdot)$ and Figure 3 shows the the resulting error $\|x(t) - \hat{x}(t)\|_2$ as a function of time.

To assess the learned vector field, the true and learned vector fields are evaluated on a 25×25 grid over $[-5, 5] \times [-5, 5]$, and the relative error $\frac{\|f_{\text{true}}(x) - f_{\text{learned}}(x)\|_2}{\max_x \|f_{\text{true}}(x)\|_2}$ is visualized in Figure 4 as a function of x. Figures 2 - 4 demonstrate that the technique described in Section V is can accurately estimate vector fields for dynamical systems. An ablation study with respect to the scaling parameter a reveals that the results are identical for any $a \in (0, 1)$. That is, in practice, the technique is not sensitive to scaling.

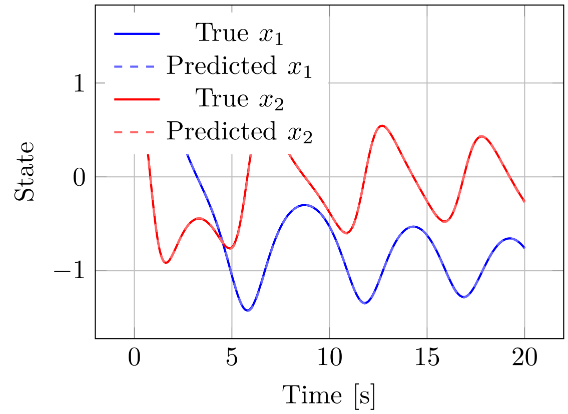

Fig. 2. True versus predicted state trajectories for the Duffing oscillator.

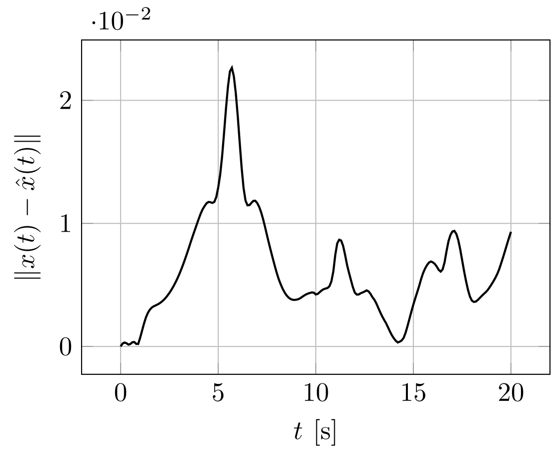

Fig. 3. Norm of the prediction error over time for the Duffing oscillator.

C. Approximating Flow Fields with Weighted Composition Operators

This numerical experiment estimates the duffing oscillator corresponding to $\alpha = 1$, $\beta = -1$, and $\delta = 0$ using a collection of observed trajectories initialized at a lattice within $[-1, 1]^2$ with spacing 0.25. The trajectories were generated using RK4 and a step-size of 0.05. The computation of both the occupation kernels and Φ_γ terms were executed leveraging the standard Simpson's rule using the exponential dot product kernel, $K(x, y) = \exp(\mu x^T y)$ with $\mu = 1/1000$ for both the space H and $\tilde{H}$, with dimensions 1 and 3 respectively. These results were first presented in [9], and are included here as another example of the generalized methodology presented in this manuscript.

The resultant approximations are presented in Figure ??, and the error plots given in Figure 5.

IX. Conclusion

This manuscript gives a generalized framework for solving a variety of inverse problems through operators and their decompositions. The method was inspired by

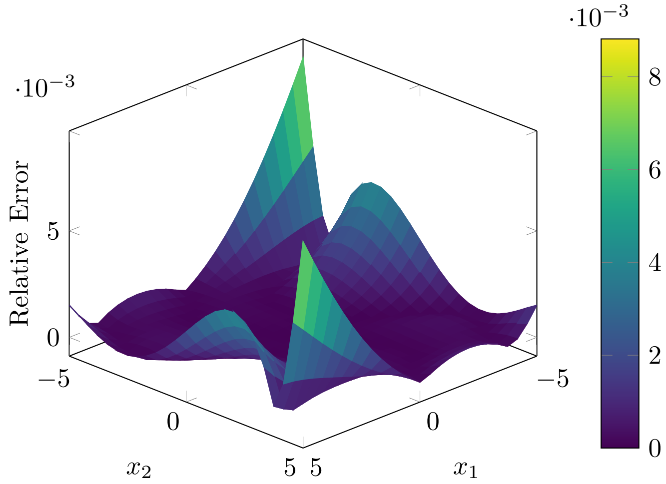

Fig. 4. Relative error between the true and the learned vector fields for the Duffing oscillator.

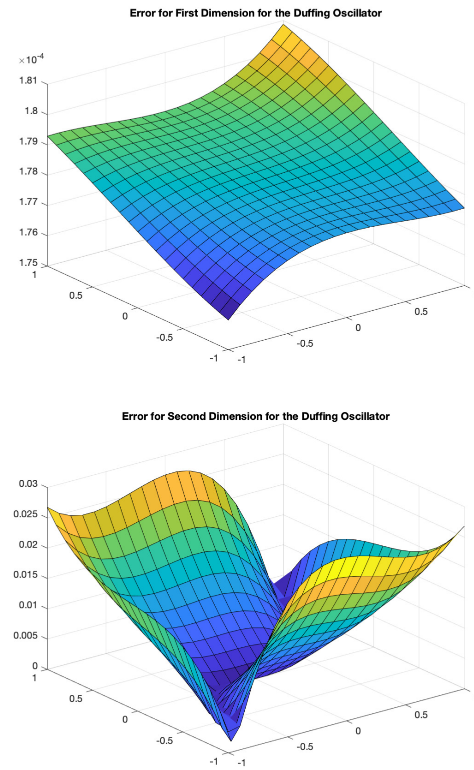

Fig. 5. Presented are the absolute error plots corresponding to approximations of the Duffing Oscillator flow field using the weighted composition operator framework. This experiment was first presented in [9].

Dynamic Mode Decompositions and Koopman operator methods for dynamical systems, and the generalized approach yielded several different operators and approaches to resolving inverse problems in dynamical systems as well as scattered data approximation. Several numerical experiments were presented, which utilized both weighted composition operators for scattered data approximation and flow field estimation, and scaled Liouville operators for flow field approximation. In each case, close estimation of the unknown functions were presented.

References

[1] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, Dynamic mode decomposition: data-driven modeling of complex systems. SIAM, 2016.

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

[3] J. A. Rosenfeld and R. Kamalapurkar, "Singular dynamic mode decomposition," SIAM Journal on Applied Dynamical Systems, vol. 22, no. 3, pp. 2357–2381, 2023.

[4] J. A. Rosenfeld, R. Kamalapurkar, L. F. Gruss, and T. T. Johnson, "Dynamic mode decomposition for continuous time systems with the liouville operator," Journal of Nonlinear Science, vol. 32, pp. 1–30, 2022.

[5] P. V. Hai and J. A. Rosenfeld, "Weighted composition operators on the mittag-leffler spaces of entire functions," Complex Analysis and Operator Theory, vol. 15, no. 1, p. 17, 2021.

[6] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel hilbert spaces of integrable functions and mercer theorem," Analysis and Applications, vol. 4, no. 04, pp. 377–408, 2006.

[7] H. Wendland, Scattered data approximation, vol. 17. Cambridge university press, 2004.

[8] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, and C. Schütte, "Data-driven approximation of the koopman generator: Model reduction, system identification, and control," Physica D: Nonlinear Phenomena, vol. 406, p. 132416, 2020.

[9] B. P. Russo, D. A. Messenger, D. Bortz, and J. A. Rosenfeld, "Weighted composition operators for learning nonlinear dynamics," IFAC-PapersOnLine, vol. 58, no. 17, pp. 97–102, 2024.

[10] J. A. Rosenfeld, B. P. Russo, R. Kamalapurkar, and T. T. Johnson, "The occupation kernel method for nonlinear system identification," SIAM Journal on Control and Optimization, vol. 62, no. 3, pp. 1643–1668, 2024.

[11] D. A. Messenger and D. M. Bortz, "Weak sindy for partial differential equations," Journal of Computational Physics, vol. 443, p. 110525, 2021.

[12] D. A. Messenger and D. M. Bortz, "Weak sindy: Galerkin-based data-driven model selection," Multiscale Modeling & Simulation, vol. 19, no. 3, pp. 1474–1497, 2021.