Kernelized Active Subspaces

Benjamin P. Russo
Department of Mathematics
Farmingdale State College
Farmingdale, NY 11735 USA
russobp@farmingdale.edu

Joel A. Rosenfeld
Department of Mathematics and Statistics
University of South Florida
Tampa, Fl 33620
rosenfeldj@usf.edu

Abstract

Given a C^1 function over a potentially high dimensional domain, the active subspace method seeks an affine subspace inside which the functions changes the most on average. This is done by finding the eigenvectors of a covariance matrix incorporating gradient information. In a similar vein, the active manifold method finds a manifold γ and if information on f is recovered along γ then f can be recovered on the connected component of a level set touching γ. An inherent limitation of the Active subspace techinique is that it only considers affine subspaces (which may still be high dimensional). Inspired by methods in occupation kernel dynamic mode decomposition, we develop a notion of active subspace taking place in a Hilbert space which contains sufficient complexity to describe highly nonlinear level sets. In this learning problem, only function values along trajectories following the gradient direction of the function are required to determine this decomposition.

1 Introduction

Active subspaces is a dimension reduction technique originated by (Russi [2010]) and developed extensively by (Constantine [2015]). Current implementations of the active subspace method can capture "active" subspaces for functions whose level sets are subspaces of $\mathbb{R}^n$. However, for functions that have very nonlinear levelsets, their applicability is limited to a small neighborhood, where the functions may be effectively linearized. This limitation manifests theoretically in the covariance matrix leveraged in the construction of active subspaces, C given below. This matrix is finite dimensional, and lacks the complexity necessary for discovering level-sets that are nonlinear manifolds.

To address this limitation, the present manuscript introduces a new operator valued kernel to stand in place of the covariance matrix of Constantine [2015]. This operator valued kernel, based on the Liouville operator given in Section 4, acts on a vector valued RKHS, which is infinite dimensional. This kernel makes available an infinite collection of eigenvalues and eigenvectors that can be leveraged to decompose a function into "active" and "inactive" components, where the active component will effectively represent level sets of the original function.

In this learning problem, only function values along trajectories following the gradient direction of the function are required to determine this decomposition. One particular application is to learn the optimal value function from a collection of observed costs and optimal control trajectories. Significantly, the gradient of the original function need not be computed using the present method, which differs from established work (cf. Bridges et al. [2019]).

Preprint. Under review.

2 Review of Active Subspace Methods

Let $X \subseteq \mathbb{R}^m$ be a compact subset equipped with a probability measure ρ and suppose that $f : X \to \mathbb{R}$ is a $C^1(X)$ function. Let, $\nabla_x f = [\partial f/\partial x_1, \ldots, \partial f/\partial x_n]^\top$ and define the $n \times n$ matrix C given by

$$C = \mathbb{E}[(\nabla_x f)(\nabla_x f)^\top] = \left(\int_X \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \, d\rho \right)_{i,j=1,1}^{n,n}. \tag{1}$$

Since C is a positive semidefinite square matrix it admits an eigenvalue decomposition

$$C = W \Lambda W^\top, \quad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \quad \lambda_1 \geq \ldots \geq \lambda_n \geq 0.$$

For a given $m < n$ we can further decompose Λ into a block diagonal matrix $\Lambda = \operatorname{diag}(\Lambda_1, \Lambda_2)$ where $\Lambda_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $\Lambda_2 = \operatorname{diag}(\lambda_{m+1}, \ldots, \lambda_n)$. Likewise W can be decomposed as $W = [W_1 \, W_2]$, where W_1 is $n \times m$, and

$$C = [W_1 \, W_2] \operatorname{diag}(\Lambda_1, \Lambda_2) [W_1 \, W_2]^\top.$$

Let w_i^1 for $i = 1, \ldots, m$ be the column vectors defined by W_1, For our choice of m we define $\mathcal{A} = \operatorname{span}\{w_i^1 \mid i = 1, \ldots, m\}$ as the active subspace of f. We can interpret C as the uncentered covariance for the gradient and since the eigenvalues in the decomposition are listed in decreasing order the active subspace A corresponds to directions which have greater average variation for the function f. In this paper we develop a generalized version of the active subspace method that enables the learning of nonlinear level sets through the introduction of an operator valued kernel.

3 Mathematical Preliminaries

The main setting for this generalized approach is vector valued reproducing kernel Hilbert spaces (RKHS). For completeness we will also define reproducing kernel Hilbert spaces.

Definition 1. A RKHS, H, over a set X is a Hilbert space of real valued functions over the set X such that for all $x \in X$ the evaluation functional $E_x g := g(x)$ is bounded. As such, the Riesz representation theorem guarantees, for all $x \in X$, the existence of a function $k_x \in H$ such that $\langle g, k_x \rangle_H = g(x)$, where $\langle \cdot, \cdot \rangle_H$ is the inner product for H.

The function k_x is called the reproducing kernel function at x, and the function $k(x, y) = \langle k_y, k_x \rangle_H$ is called the kernel function corresponding to H.

Vector valued RKHSs began to appear in learning theory over the past decade Carmeli et al. [2010], though their inception dates back at least as far as the 1950s (e.g. Pedrick [1957]).

Definition 2. 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 $x \in X$ the evaluation mapping $E_x : H \to \mathcal{Y}$ given by $E_x(f) = f(x)$ is bounded. The operator valued kernel for a vector valued reproducing kernel Hilbert space is given by $K : X \times X \to \mathcal{B}(Y)$, $K(x, y) = E_x E_y^*$, here $\mathcal{B}(\mathcal{Y})$ denotes the bounded operators on $\mathcal{Y}$.

Boundedness of the functional $E_x : H \to \mathcal{Y}$ is equivalent to the boundedness of the functional $H \ni g \mapsto \langle g(x), y \rangle_{\mathcal{Y}}$ for each $x \in X$ and $y \in \mathcal{Y}$. 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 $g \in H$.

Example 3. Let $\mathscr{H}$ be a real-valued reproducing kernel Hilbert space on a set X, with kernel $k : X \times X \to \mathbb{R}$. We can associate $H = \bigoplus_{i=1}^n \mathscr{H}$ with an $\mathbb{R}^n$ valued reproducing kernel Hilbert space R^n by identifying $f \in H$ with the function from X to $\mathbb{R}^n$ given by the formula, $f(x) := (f_1(x), \ldots, f_n(x))^\top \in \mathbb{R}^n$. Moreover, $E_x(f) = (f_1(x), \ldots, f_n(x))^\top$. Note that,

$$\langle f, E_x^*(v) \rangle_H = \langle E_x(f), v \rangle_{\mathbb{R}^n} = \sum_{i=1}^n f_i(x) v_i = \langle f, (v_1 k_x, \ldots, v_n k_x)^\top \rangle_H = \langle f, k_x v \rangle_H$$

where k_x is again the *scalar*-valued kernel for $\mathscr{H}$. For $v \in \mathbb{R}^n$ we define $K_{x,v} = k_x v \in H$, the reproducing property for H is given by

$$\langle f, K_{x,v} \rangle_H = \langle f, k_x v \rangle_H = \langle f(x), v \rangle_{\mathbb{R}^n}.$$

Under the standard basis for $\mathbb{R}^n$, the kernel function $K : X \times X \to \mathcal{B}(\mathbb{R}^n)$ is given by $k(x, y) \operatorname{Id}_n$ where Id_n is the $n \times n$ identity matrix.

Remark 4. In general, for a vector valued reproducing kernel Hilbert space H with kernel K, we can define $K_x v(\cdot) := K(\cdot, x)v \in H$. We note that $K_{x,v} = K_x v = E_x^*(v)$ since

$$(E_x^*(v))(y) = E_y E_x^*(v) = K(y,x)v = (K_x v)(y)$$

for all y and the reproducing property is given by

$$\langle f(x), v\rangle_{\mathcal{Y}} = \langle E_x(y), v\rangle_{\mathcal{Y}} = \langle f, E_x^*(v)\rangle_H = \langle f, K_{x,v}\rangle_H.$$

Fundamental to the evaluation of an operator theoretic kernel that will be introduced subsequently is the concept of an occupation kernel . In the same manner that a reproducing kernel embeds points from $\mathbb{R}^n$ into an infinite dimensional space through the mapping $x \mapsto k_x$, an occupation kernel embeds a continuous signal into the RKHS through the representation of a integral over the signal. The following definition generalizes that found in for vector valued RKHSs.

Definition 5. Let $X \subset \mathbb{R}^n$ be compact, $\mathscr{H}$ be a RKHS of continuous real valued functions over X, and $\gamma : [0,T] \to X$ be a bounded measurable trajectory. The functional $g \mapsto \int_0^T g(\gamma(\tau))d\tau$ is bounded, and may be represented as $\int_0^T g(\gamma(\tau))d\tau = \langle g, \Gamma_\gamma\rangle_{\mathscr{H}}$, for some $\Gamma_\gamma \in H$ by the Riesz representation theorem. The function Γ_γ is called the occupation kernel corresponding to γ in H.

Proposition 6. *Let $\mathscr{H}$ be a RKHS over a compact set X consisting of continuous functions and let $\gamma : [0,T] \to X$ be a continuous trajectory. The occupation kernel corresponding to γ in $\mathscr{H}$, Γ_γ, may be expressed as*

$$\Gamma_\gamma(x) = \int_0^T k(x, \gamma(t))dt. \tag{2}$$

We can now extend the above definitions to the vector valued case.

Definition 7. Let $X \subset \mathbb{R}^n$ be compact, H be a $\mathbb{R}^n$ valued RKHS of continuous functions over X, and $\gamma : [0,T] \to X$ be a bounded measurable trajectory. For every $v \in \mathbb{R}^n$, the functional $g \mapsto \left\langle \int_0^T g(\gamma(\tau))d\tau, v \right\rangle_{\mathbb{R}^n}$ is bounded, and may be respresented as $\left\langle \int_0^T g(\gamma(\tau))d\tau, v \right\rangle_{\mathbb{R}^n} = \langle g, \Gamma_{\gamma,v}\rangle_H$, for some $\Gamma_{\gamma,v} \in H$ by the Riesz representation theorem. The function $\Gamma_{\gamma,v}$ is called the occupation kernel corresponding to γ in H and $v \in \mathbb{R}^n$

Definition 8. Let $X \subset \mathbb{R}^n$ be compact, H be a $\mathcal{Y}$-valued RKHS of continuous functions over X, and $\gamma : [0,T] \to X$ be a bounded measurable trajectory. Define the operator E_γ by

$$E_\gamma : H \to \mathcal{Y}, \quad g \mapsto \int_0^T g(\gamma(t))dt \in \mathcal{Y}.$$

Moreover,

$$E_\gamma^* : \mathcal{Y} \to H, \quad v \mapsto \Gamma_{\gamma,v}.$$

For consistent notation with the above discussion of vector-valued RKHS's we denote $E_\gamma^*(v) = \Gamma_{\gamma,v}$ by $\Gamma_\gamma v$.

Let P be the set of bounded measurable trajectories. We call

$$\Gamma(x,\gamma) : X \times P \to \mathcal{B}(\mathcal{Y}), \quad \Gamma(x,\gamma) := E_x E_\gamma^*$$

the operator valued occupation kernel.

Proposition 9. *For all $i \in \{1,\dots,n\}$ let $\mathscr{H}_i$ be real-valued reproducing kernel Hilbert spaces on a set X and $H = \bigoplus_{i=1}^n \mathscr{H}_i$ be the associated $\mathbb{R}^n$-valued reproducing kernel Hilbert space. For a given path $\gamma : [0,T] \to X \subset \mathbb{R}^k$, let $\Gamma_\gamma^i \in \mathscr{H}_i$ be the scalar valued occupation kernel associated to γ. The function $\Gamma_{\gamma,v} \in H$ is given by $\Gamma_\gamma(x) \odot v$ where $\Gamma_\gamma(x) = (\Gamma_\gamma^1(x), \dots, \Gamma_\gamma^n(x))^\top$. Here, $\odot$ represents the Hadamard product.*

Proof. Let $g = (g_1, \dots, g_n)^\top$ be in H, and $v = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$ then

$$\langle g, \Gamma_{\gamma,v}\rangle_H = \left\langle \int_0^T g(\gamma(t))dt, v \right\rangle_{\mathbb{R}^n} = \sum_{i=1}^n \int_0^T v_i g_i(\gamma(t))dt = \sum_{i=1}^n \langle g_i, \Gamma_\gamma^i v_i\rangle_{\mathscr{H}_i} = \langle g, \Gamma_\gamma \odot v\rangle_H.$$

□

4 Theoretical Description of the Method

For this section we will assume the functions are over a compact domain $X \subset \mathbb{R}^n$ and are $\mathbb{R}^n$ valued.

Definition 10. For an $\mathbb{R}^n$ valued reproducing kernel Hilbert space H and a given symbol $\psi : \mathbb{R}^n \to \mathbb{R}^n$ we define the Liouville operator with symbol ψ as $A_\psi(g) = D(g)\psi$.

Modally, A_ψ is a closed, densely defined, and unbounded operator, owing to the inclusion of the differentiation operator D. Throughout this manuscript, a heuristic assumption that this operator is bounded and even compact will be leveraged in the development of the numerical methods of this paper. This heuristic assumption is justified in several contexts, where the selection of appropriate Hilbert spaces in the range and domain of A_ψ can make it a bounded operator, and the use of scaled versions of this operator (cf. Rosenfeld et al.) can in fact produce a compact operator that agrees to computational precision with A_ψ on a compact subset of a given workspace.

Given a $f, h \in C^1(\mathbb{R}^n, \mathbb{R})$, define the Liouville operator $A_{\nabla f} : \mathcal{D}(A_{\nabla f}) \to H$. Let $\mathcal{D}_h(A_{\nabla f}) \subset \mathcal{D}(A_{\nabla f})$ be defined as those vectors $g \in \mathcal{D}(A_{\nabla f})$ such that $A_{\nabla f} g \in \mathcal{D}(A^*_{\nabla h})$. Under the heuristic assumption discussed above, $\mathcal{D}_h(A_{\nabla f}) = H$. Definition 11 yields an operator theoretic replacement for the covariance matrix, C, in (1). The advantage gained through this perspective is that the operator valued kernel in Definition 11 provides a potentially infinite collection of eigenvalues and eigenfunctions that can be leveraged to decompose $\mathbb{R}^n$ in a nonlinear manner.

Definition 11. Let H be a $\mathbb{R}^n$- valued reproducing kernel Hilbert space, we define

$$C(\nabla f, \nabla h) : \mathcal{D}_h(A_f) \to H, \qquad C(\nabla f, \nabla h) := A^*_{\nabla h} A_{\nabla f}.$$

If it is assumed that $A_{\nabla f}$ is compact, and consequently $C(\nabla f, \nabla f)$ is compact (compact operators form an ideal in the algebra of bounded operators), then as a self adjoint operator, $C(\nabla f, \nabla f)$ is diagonalizable. That is, the eigenfunctions of $C(\nabla f, \nabla f)$ form an orthonormal basis of H. Moreover, $A_{\nabla f}$ has a singular value decomposition, where the right singular vectors are the eigenfunctions of $C(\nabla f, \nabla f)$. In the subsequent numerical methods, finite rank representations for $C(\nabla f, \nabla f)$ will be extracted from finite rank representations of $A_{\nabla f}$.

Definition 12. Suppose $\varepsilon > 0$ is a given threshold and that $A_{\nabla f}$ is diagonalizable. Define,

$$\mathcal{A}_\varepsilon := \operatorname{span}\{\varphi \mid C(\nabla f, \nabla f)\varphi = \lambda\varphi, \quad \sqrt{|\lambda|} > \varepsilon\} \qquad \mathcal{I}_\varepsilon := \mathcal{A}_\varepsilon^\perp.$$

Note, our Hilbert space H can be orthogonally decomposed as $H = \mathcal{A}_\varepsilon \oplus \mathcal{I}_\varepsilon$. Furthermore, for notational convenience write $\Sigma := \{\varphi \in H : \exists \lambda \in \mathbb{C} \text{ such that } C(\nabla f, \nabla f)\varphi = \lambda\varphi\}$ and $\Sigma^* := \{\psi \in H : \exists \lambda \in \mathbb{C} \text{ such that } A_{\nabla f} A^*_{\nabla f}\psi = \lambda\psi\}$. Each $\varphi \in \Sigma$ is a right singular vector of $A_{\nabla f}$ and maps to the left singular vectors under$A_{\nabla f}$.

Moreover, the identity function, $g_{id}(x) = x$, which is assumed to be in H, admits the decomposition

$$x = g_{id}(x) = \sum_{\varphi \in \Sigma} \langle g, \varphi \rangle_H \varphi(x).$$

Set $M = \max_{\varphi \in \Sigma} |\langle \varphi, g_{id} \rangle_H|$ and set $\varepsilon_0 = \varepsilon / M$. Write $\Sigma_\varepsilon := \{\varphi \in \Sigma : \sqrt{|\lambda|} > \varepsilon\}$. Then,

$$\begin{aligned}
\nabla f(x) &= D(g_{id}) \nabla f(g_{\mathrm{id}}(x)) \\
&= D(P_{\mathcal{A}_{\varepsilon_0}} g_{\mathrm{id}}(x)) \nabla f(x) + D(P_{\mathcal{I}_{\varepsilon_0}} g_{\mathrm{id}}(x)) \nabla f(x) \\
&= \left(\sum_{\varphi \in \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H D\varphi(x) \right) \nabla f(x) + \left(\sum_{\varphi \in \Sigma \setminus \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H D\varphi(x) \right) \nabla f(x) \\
&= \sum_{\varphi \in \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H D\varphi(x) \nabla f(x) + \sum_{\varphi \in \Sigma \setminus \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H D\varphi(x) \nabla f(x) \\
&= \sum_{\varphi \in \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H A_{\nabla f}\varphi(x) + \sum_{\varphi \in \Sigma \setminus \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H A_{\nabla f}\varphi(x) \\
&= \sum_{\varphi \in \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H \sqrt{\lambda}\psi(x) + \sum_{\varphi \in \Sigma \setminus \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H \sqrt{\lambda}\psi(x),
\end{aligned}$$

where ψ are the right singular vectors of $A_{\nabla f}$ corresponding to $\varphi \in \Sigma$.

Definition 13. The subspace $\mathcal{A}_{\varepsilon_0}$ described above will be called the *active subspace* within H for f and $P_{\mathcal{A}_{\varepsilon_0}} g_{\mathrm{id}}$ is called the *active* component of g_{id} (of order ε)

Each term of the active component satisfies $\langle g_{id}, \varphi \rangle \sqrt{|\lambda|} > \varepsilon$.

5 Finite Rank Representations via Occupation Kernels

To computationally determine an estimation of the eigendecomposition of $C(\nabla f, \nabla f)$, a finite rank representation of A_f will be determined using a collection of trajectories, $\{\gamma_i : [0,T] \to \mathbb{R}^n\}_{i=1}^M$, that satisfy $\dot{\gamma}_i = \nabla f(\gamma_i)$ and their corresponding operator valued occupation kernels, $\{\Gamma_\gamma\}_{i=1}^M$. Significant to the method is the following proposition:

Proposition 14. *Let H be a $\mathbb{R}^n$ valued RKHS consisting of continuously differentiable funcitons, and let $\gamma : [0,T] \to \mathbb{R}^n$ be a trajectory satisfying $\dot{\gamma} = \nabla f(\gamma)$ for a function $f : \mathbb{R}^n \to \mathbb{R}$ which yields a densely defined Liouville operator, $A_{\nabla f}$, over H. Then the following relation holds*

$$\langle A_{\nabla f} g, \Gamma_\gamma v \rangle_H = \langle g, (K_{\gamma(T)} - K_{\gamma(0)}) v \rangle_H. \tag{3}$$

Hence, $A^*_{\nabla f} \Gamma_{\gamma,v} = (K_{\gamma(T)} - K_{\gamma(0)})v$.

Proof. Note that

$$\begin{aligned}
\langle A_{\nabla f} g, \Gamma_{\gamma,v} \rangle_H &= \langle D(g) \nabla f, \Gamma_{\gamma,v} \rangle_H \\
&= \left\langle \int_0^T Dg(\gamma(t)) \nabla f(\gamma(t)) dt, v \right\rangle_{\mathbb{R}^n} \\
&= \left\langle \int_0^T \frac{d}{dt} g(\gamma(t)) dt, v \right\rangle_{\mathbb{R}^n} \\
&= \langle g(\gamma(T)) - g(\gamma(0)), v \rangle_{\mathbb{R}^n} \\
&= \langle g, \left(K_{\gamma(T)} - K_{\gamma(0)}\right) v \rangle_H.
\end{aligned}$$

□

Leveraging this relation, a finite rank representation of $A^*_{\nabla f}$ may be determined, and consequently the transpose of this representation will represent $A_{\nabla f}$ under the boundedness assumption.

In particular, replacing v with vectors from the standard basis in $\mathbb{R}^n$, and for a fixed $1 \leq i \leq n$ writing $\beta_i = \operatorname{span}\{\Gamma_{\gamma_j} e_i\}_{j=1}^M$

Proposition 15. *Let $\{e_k\}$ denote the standard basis for $\mathbb{R}^n$ and fix an $i \in \{1, \ldots, n\}$. For a finite dimensional subspace given by $\beta_i = \operatorname{span}\{\Gamma_{\gamma_j, e_i}\}_{j=1}^M = \operatorname{span}\{\Gamma_{\gamma_j} e_i\}_{j=1}^M$,*

$$\begin{aligned}
[P_{\beta_i} A^*_{\nabla f}]_{\beta_i}^{\beta_i} = &\begin{pmatrix} \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_1} e_i \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_M} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_M} e_i \rangle_H \end{pmatrix}^{-1} \\
\times &\begin{pmatrix} \langle \left(K_{\gamma_1(T)} - K_{\gamma_1(0)}\right) e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & \langle \left(K_{\gamma_M(T)} - K_{\gamma_M(0)}\right) e_i, \Gamma_{\gamma_1} e_i \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \left(K_{\gamma_1(T)} - K_{\gamma_1(0)}\right) e_i, \Gamma_{\gamma_M} e_i \rangle_H & \cdots & \langle \left(K_{\gamma_M(T)} - K_{\gamma_M(0)}\right) e_i, \Gamma_{\gamma_M} e_i \rangle_H \end{pmatrix}.
\end{aligned}$$

Proof. For $h \in \mathcal{D}(A^*_{\nabla f})$, the coefficients $\{a_j\}_{j=1}^M$ in the projection of $A^*_{\nabla f} h$ onto β_i, given by $P_{\beta_i} A^*_{\nabla f} h = \sum_{j=1}^M a_j \Gamma_{\gamma_j} e_i$, can be expressed as

$$\begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix} = \begin{pmatrix} \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_1} e_i \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_M} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_M} e_i \rangle_H \end{pmatrix}^{-1} \begin{pmatrix} \left\langle A^*_{\nabla f} h, \Gamma_{\gamma_1} e_i \right\rangle_H \\ \vdots \\ \left\langle A^*_{\nabla f} h, \Gamma_{\gamma_M} e_i \right\rangle_H \end{pmatrix}. \tag{4}$$

Assuming that the occupation kernels are in the domain of the Liouville operator, i.e., $\beta_i \subset \mathcal{D}(A^*_{\nabla f})$, for $h \in \beta_i$, given by $h = \sum_{j=1}^{M} c_j \Gamma_{\gamma_j} e_i$, for a fixed $k \in \{1, \ldots, M\}$, we have

$$\begin{aligned}\left\langle A^*_{\nabla f} h, \Gamma_{\gamma_k} e_i \right\rangle_H &= \sum_{j=1}^{M} c_j \left\langle A^*_{\nabla f} \Gamma_{\gamma_j} e_i, \Gamma_{\gamma_k} e_i \right\rangle_H \\ &= \begin{pmatrix} \left\langle A^*_{\nabla f} \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_k} e_i \right\rangle_H \ldots, \left\langle A^*_{\nabla f} \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_k} e_i \right\rangle_H \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix}\end{aligned}$$

As a result, a finite rank representation of $A^*_{\nabla f}$ restricted to β_i, i.e., the matrix $[P_{\beta_i} A^*_{\nabla f}]^{\alpha}_{\alpha}$ that maps the coefficients $\{c_j\}_{j=1}^{M}$ to the coefficients $\{a_j\}_{j=1}^{M}$, is given as

$$\begin{aligned}[P_{\beta_i} A^*_{\nabla f}]^{\beta_i}_{\beta_i} = &\begin{pmatrix} \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_1} e_i \rangle_H \\ \vdots & \ddots & \vdots \\ \langle \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_M} e_i \rangle_H & \cdots & \langle \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_M} e_i \rangle_H \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \langle A^*_{\nabla f} \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & \langle A^*_{\nabla f} \Gamma_{\gamma_1} e_i, \Gamma_{\gamma_M} e_i \rangle_H \\ \vdots & \ddots & \vdots \\ \langle A^*_{\nabla f} \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_1} e_i \rangle_H & \cdots & A^*_{\nabla f} \Gamma_{\gamma_M} e_i, \Gamma_{\gamma_M} e_i \rangle_H \end{pmatrix}\end{aligned}$$

The proof is completed by an application of Proposition 14 □

Proposition 16. *Let $\{e_k\}$ denote the standard basis for $\mathbb{R}^n$ and for each $i \in \{1, \ldots, n\}$ let $\beta_i = \operatorname{span}\{\Gamma_{\gamma_j, e_i}\}_{j=1}^{M} = \operatorname{span}\{\Gamma_{\gamma_j} e_i\}_{j=1}^{M}$. Let $B = \oplus_{i=1}^{n} \beta_i$ then*

$$[P_B A^*_{\nabla f}]^B_B = \operatorname{diag}([P_{\beta_i} A^*_{\nabla f}]^{\beta_i}_{\beta_i})$$

Proof. Let ℓ and m be arbitrary indices in $\{1, \ldots, M\}$. We must show for $i, j \in \{1, \ldots, n\}$ with $i \neq j$ that

$$\langle A^*_{\nabla f} \Gamma_{\gamma_\ell} e_i, \Gamma_{\gamma_m} e_j \rangle = 0.$$

By an applications of Propositions 9 and 14 we get

$$\begin{aligned}\langle A^*_{\nabla f} \Gamma_{\gamma_\ell} e_i, \Gamma_{\gamma_m} e_j \rangle &= \langle \Gamma_{\gamma_\ell} e_j, (K_{\gamma_m(T)} - K_{\gamma_m(0)}) e_i \rangle_H \\ &= \langle \Gamma_{\gamma_\ell, e_j}(\gamma_m(T)) - \Gamma_{\gamma_\ell, e_j}(\gamma_m(0)), e_i \rangle_{\mathbb{R}^n} \\ &= (0, \ldots, \Gamma_{\gamma_\ell}(\gamma_m(T)) - \Gamma_{\gamma_\ell}(\gamma_m(0)), \ldots, 0) \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},\end{aligned}$$

where we get the difference of the *scalar* valued occupation kernel for γ_ℓ evaluated at $\gamma_m(T)$ and $\gamma_m(0)$ in the j-th spot and a 1 in the i-th spot. Hence, this is non-zero only when $i = j$. □

Hence for each dimension of the workspace, a collection of approximate eigenfunctions for $A_{\nabla f}$ may be determined through the SVD of the matrix $([P_{\beta_i} A^*_{\nabla f}]^{\beta_i}_{\beta_i})^T$. For each singular vector of $([P_{\beta_i} A^*_{\nabla f}]^{\beta_i}_{\beta_i})^T$, $\eta = (\eta_1, \ldots, \eta_M)$, with singular value λ, the corresponding normalized singular function in H is given as $\varphi = \frac{1}{\sqrt{\eta^T G \eta}} \sum_{j=1}^{M} \eta_j \Gamma_{\gamma_j} e_i$, where G is the Gram matrix corresponding to the basis for β_i. Hence, to evaluate $\sum_{\varphi \in \Sigma_{\varepsilon_0}} \langle g_{id}, \varphi \rangle_H \sqrt{\lambda} \psi(x)$ we must be able to compute $\langle g_{id}, \Gamma_{\gamma_j} e_i \rangle$. By definition, this is given as

$$\langle g_{id}, \Gamma_{\gamma_j} e_i \rangle = \left\langle \int_0^T g_{id}(\gamma_j(t)) dt, e_i \right\rangle_{\mathbb{R}^n} = \int_0^T \gamma_j^i(t) dt,$$

i.e. it is the integral of the i-th component of $\gamma_j : [0, T] \to X \subset \mathbb{R}^n$

6 Discussion

This manuscript presents a new dimension reduction technique as a nonlinear version of the active subspace method. To address the limitation of the active subspace routine this manuscript introduces a new operator valued kernel to stand in place of the covariance matrix of Constantine [2015]. This operator valued kernel, based on the Liouville operator given in Section 4, acts on a vector valued RKHS, which is infinite dimensional. This kernel makes available an infinite collection of eigenvalues and eigenvectors that can be leveraged to decompose a function into "active" and "inactive" components, where the active component will effectively represent level sets of the original function. In this learning problem, only function values along trajectories following the gradient direction of the function are required to determine this decomposition. Significantly, the gradient of the original function need not be computed using the present method, which differs from established work (cf. Bridges et al. [2019]). Additionally, an overview of occupation kernels over vector valued reproducing kernel Hilbert spaces is given in section 3 and explicit formulas for computation are given in section 5.

7 Conclusion

This technique is influenced by occupation kernel dynamic mode decomposition seen in (Rosenfeld and Kamalapurkar [To Appear]). Most notably, under the assumption that $A_{\nabla f}$ is a self-adjoint operator, the eigenfunctions for $C(\nabla f, \nabla f)$ and $A_{\nabla f}$ become the same. Under this assumption, essentially the same computations as seen in occupation kernel dynamic mode decomposition are present. The authors expect the same challenges that appear in occupation kernel DMD to be present in this new technique as well. While, in principle any reproducing kernel Hilbert space can be used, choice of RKHS affects the operator theoretic properties of boundedness and compactness of $C(\nabla f, \nabla f)$ and $A_{\nabla f}$. (Russo and Rosenfeld [2021]) explores the properties of Liouville operators over the Hardy space and (Rosenfeld and Kamalapurkar [To Appear]) gives a modification of Liouville operators that allow for compactness. Moreover, it is found in dynamic mode decomposition that careful parameter selection is necessary and a poor choice of parameter can lead to bad reconstruction of the function. This in general is one of the major limitations of the technique.

Acknowledgments and Disclosure of Funding

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

References

Robert A. Bridges, Anthony D. Gruber, Christopher Felder, Miki Verma, and Chelsey Hoff. Active manifolds: A non-linear analogue to active subspaces, 2019.

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.

Paul G Constantine. *Active subspaces: Emerging ideas for dimension reduction in parameter studies*. SIAM, 2015.

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

Joel A. Rosenfeld and Rushi Kamalapurkar. Dynamic mode decomposition with control liouville operators. In *Proceedings of the 24th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2021)*, To Appear.

Joel A. Rosenfeld, Rushikesh Kamalapurkar, L. Forest Gruss, and Taylor T. Johnson. Dynamic mode decomposition for continuous time systems with the Liouville operator. *Under Review*.

Trent Michael Russi. *Uncertainty quantification with experimental data and complex system models*. PhD thesis, UC Berkeley, 2010.

Benjamin P. Russo and Joel A. Rosenfeld. Liouville operators over the hardy space, 2021.