Geometric Considerations of a Good Dictionary for Koopman Analysis of Dynamical Systems

Erik M. Bollt

Department of Electrical and Computer Engineering and C^3S^2 the Clarkson Center for Complex Systems Science, Clarkson University, Potsdam, New York 13699

December 20, 2019

Abstract

Representation of a dynamical system in terms of simplifying modes is a central premise of reduced order modelling and a primary concern of the increasingly popular DMD (dynamic mode decomposition) empirical interpretation of Koopman operator analysis of complex systems. In the spirit of optimal approximation and reduced order modelling the goal of DMD methods and variants are to describe the dynamical evolution as a linear evolution in an appropriately transformed lower rank space, as best as possible. However, as far as we know there has not been an in depth study regarding the underlying geometry as related to an efficient representation. To this end we present that a good dictionary, that quite different from other's constructions, we need only to construct optimal initial data functions on a transverse co-dimension one set. Then the eigenfunctions on a subdomain follows the method of characteristics. The underlying geometry of Koopman eigenfunctions involves an extreme multiplicity whereby infinitely many eigenfunctions correspond to each eigenvalue that we resolved by our new concept as a quotient set of functions, in terms of matched level sets. We call this equivalence class of functions a "primary eigenfunction" to further help us to resolve the relationship between the large number of eigenfunctions in perhaps an otherwise low dimensional phase space. This construction allows us to understand the geometric relationships between the numerous eigenfunctions in a useful way. Aspects are discussed how the underlying spectral decomposition as the point spectrum and continuous spectrum fundamentally rely on the domain.

Keywords: *Koopman operator, spectral analysis, Reduced order model, dynamical system, good dictionary, DMD, EDMD*

Mathematics Subject Classification: *37M25, 37C10, 34M45, 47E05, 42-04, 68T99, 65L99*

1 Introduction

The data-oriented Koopman eigenfunction analysis perspection for analyzing dynamical systems has become extensively popular and relevant lately in science and engineering [6, 19, 20], as a way to interpret even a nonlinear dynamical system, as a much simpler linear system. However in turn, a possibly finite dimensional nonlinear system, may well require infinitely many dimensions to interpret as a linear system. This is perhaps not a bad trade-off, infinite dimensions for linearity, and from there, computational schemes proceed to estimate the representation in terms of finite dimensional truncation of the infinite dimensional embedding. In this paper we will introduce a new perspective on the nature of eigenfunctions, relating algebraic and geometric aspects by introducing a notion of "primary eigenfunctions" as a quotient over all those eigenfunctions agree up to the same algebraic structure, that we reduce to a statement regarding same

sets of level sets. We believe that this is an important step in understanding what will be our main goal to construct a good and distinct set of eigenfunctions, for the purpose of representation. This leads us to present a new data driven approach to understanding optimal empirical eigenfunctions by as it turns out they can be simply constructed on a lower dimensional data set transverse to the flow. As an aside, also we will discuss how the nature of the point and continuous spectra are domain dependent.

First, we review the underlying theory. Consider, a differential equation in $M \subset \mathbb{R}^d$,

$$\dot{x} = F(x), \tag{1}$$

with a vector field,

$$F : M \to M. \tag{2}$$

As usual, the nonautonomous scenario, $f(x,t) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ can be written in $d+1$ dimensions as an autonomous problem by augmenting with a time variable. State the flow for each $t \in \mathbb{R}$ (or semi-flow for $t \geq 0$) as a function, $x(t) := \varrho_t(x_0)$ for a trajectory starting at $x(0) = x_0 \in M$.

The associated Koopman operator, often called a composition operator, describes the evolution of "observables", or "measurements" along the flow, [6, 25]. Rather than analyzing individual trajectories in its phase space, we analyze observations measured in the space, and observations are functions. These "observation functions",

$$g : M \to \mathbb{C}, \tag{3}$$

are elements of a space of observation functions $\mathcal{F}$. For example,

$$\mathcal{F} = L^2(M) = \{g : \int_M |g(s)|^2 ds < \infty\}, \tag{4}$$

is commonly used since it is particularly convenient for numerical applications that utilize the inner product associated with the Hilbert space structure [6, 19, 24, 20, 24, 27]. For now we discuss scalar observation functions, but multiple scalar observation functions can be considered together, "stacked" as a composite vector valued observation function. The dynamics of how these observations change over time along orbits is what the Koopman operator defines.

Definition 1 Koopman Operator. *(Composition Operator), [16, 25, 7, 9]. The operator, $\mathscr{K}_t$, associated with ϱ_t, is a (semi-) flow, stated as the following composition,*

$$\begin{aligned} \mathscr{K}_t : \mathcal{F} &\to \mathcal{F}, \\ g &\mapsto \mathscr{K}_t[g](x) = g \circ \varrho_t, \end{aligned} \tag{5}$$

on the function space $\mathcal{F}$, for each $t \in \mathbb{R}$ (or as a semi-flow if the relation only holds for $t \geq 0$). That is, for each x, we observe the value of an observable g not at x, but "downstream" by time t, at $\varrho_t(x)$. See Fig. 1.

An interesting and important feature of the Koopman operator is that it is linear on $\mathcal{F}$, but at the cost of possibly being infinite dimensional, even though it may be associated with a flow ϱ_t that evolves on a finite dimensional space, and indeed even due to a nonlinear vector field.

The spectral theory of Koopman operators [9, 6, 25] concerns eigenfunctions and eigenvalues of the operator K_t, which may be stated in terms that the pair must satisfy the equation,

$$\mathscr{K}_t[\phi_\lambda](x) = b^t \phi_\lambda(x) = e^{\lambda t}\phi_\lambda(x). \tag{6}$$

See Fig. 1. We will write an eigenvalue—eigenfunction pair of the Koopman operator as $(\lambda, \phi_\lambda(x))$, and call this pairing a "KEIG", and from here forward we will emphasize the association between such an eigenfunction and its eigenvalue, by the subscript, $\phi_\lambda(x)$. For convenience, we will say "KEIGs" when referring to a Koopman eigenvalue and eigenfunction pair, and write, $(\lambda, \phi_\lambda(x))$. The multiplying factor follows the eigenvalue, $b = e^\lambda$. A major point that we plan to make herein will be that for each λ, not only is the eigenfunction ϕ_λ not unique, there are *at least uncountably infinitely many* functions associated with

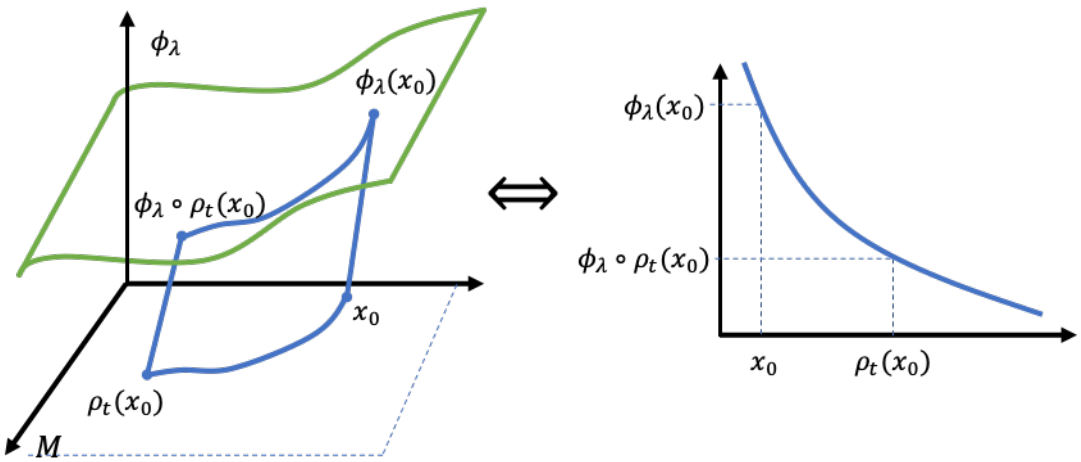

Figure 1: The action of a Koopman operator (composition operator) is to extract the value of a measurement function downstream, (Left) Eq. (5), but (Right) for eigenvalue and eigenfunction pair (KEIGS) $(\lambda, \phi_\lambda(x))$ that function has the property Eq. (6) effectively interpreting the change of ϕ_λ *along* an orbit as if the dynamics is linear even if the flow may be nonlinear in its phase space M.

each λ, and we state this even while allowing only unit normalized eigenfunctions, to remove the trivial idea that constant multiples of eigenfunctions are eigenfunctions.

A trending concept in the empirical study of dynamical systems has come to be the spectral decomposition of observables into eigenfunctions of the Koopman operator. Let a vector valued set of observables $\mathbf{g}(x) = [g_1(x), ..., g_D(x)] : M \to \mathbb{C}^D \in \mathcal{F}^D$, be written as a linear combination of eigenfunctions,

$$\mathbf{g}(x) = \sum_{j=1}^{\infty} \phi_{\lambda_i}(x)\mathbf{v}_j \tag{7}$$

where the vectors $\mathbf{v}_j \in \mathbb{C}^D$ are called Koopman modes. Further, the power of this concept lies in the following expression that describes the dynamics of observations in terms that remind us of the linear Fourier analysis, but now of the Koopman modes. That is if, $\mathbf{y}(t) := \mathbf{g} \circ \mathbf{x}(t) \equiv \mathbf{g} \circ \varrho_t(\mathbf{x}_0) := K_t[\mathbf{g}](t)$, then,

$$\mathbf{g}(t) = \sum_{j=1}^{\infty} e^{\lambda_j t}\phi_{\lambda_j}(\mathbf{x}_0)\mathbf{v}_j. \tag{8}$$

What is not generally discussed is the uniqueness of this decomposition, and furthermore the nature and even the cardinality of these eigenfunctions. Furthermore, given nonuniqueness of the representation we are particularly interested in *an efficient* representation. Thus we see the nonuniqueness not as a problem, but rather as an opportunity to leverage toward better efficiency. We will interpret this efficiency phrase in similar manner as is commonly used in the data driven concept leading to the POD-KL (Karhunen-Loeve) decomposition, which is that a basis set should allow as-fast-as-possible convergence, when considered against all other allowable basis sets, [14, 23, 32, 11, 31]. Since this construction toward efficient representation by eigenfunctions will be a main outcome of this work, we define more carefully.

Definition 2 *A finite set of unit eigenfunctions* $\{\varphi_{\lambda_i}(x)\}_{i=1}^k$, $\phi_{\lambda_i} : U \to \mathbb{C}$, $\|\varphi_{\lambda_i}\|_{L^2(U)} = 1$ *for an open domain* $U \in M$ *designed empirically for a given target observation function,* $q : U \to \mathbb{C}$ *is* **an efficient set of Koopman eigenfunctions and corresponding eigenvalues** *(KEIGS) if it yields minimal error for a given finite partial sum:*

$$\min_{\mathbf{a}} \|\sum_{i=1}^{k} a_i\varphi_{\lambda_i} - q\|_{L^2(U)} \leq \min_{\mathbf{b}} \|\sum_{i=1}^{k} b_i\phi_{\tilde{\lambda}_i} - q\|_{L^2(U)} \tag{9}$$

as contrasted against any other set of k-unit eigenfunctions, $\{\phi_{\tilde{\lambda}_i}(x)\}_{i=1}^k$, allowing that even the eigenvalues may not necessarily match.

Here we will address these issues, in the setting of the underlying dynamics being a low-dimensional differential equation.

2 The Eigenfunction PDE

How does one construct an eigenfunction of a given flow? Much has been been written lately about the data-driven approach given data as trajectories, how to approximate eigenfunctions computed through numerical methods, namely DMD, EDMD and variants [30, 33, 35, 34]. There exists however an analytical description in terms of a PDE that follows the infinitesimal generator of the Koopman operator. Corresponding to the statement of $\mathscr{K}_t$ as a semi-group of compositions, follows the action of the infinitesimal generator, [15, 27, 21, 5],

$$\mathcal{L} = F \cdot \nabla, \tag{10}$$

and so we recall,

Theorem 1 *[27, 4] A smooth exact eigenfunction of the Koopman operator of a given flow, for a given eigenvalue $\lambda \in \mathbb{C}$ must satisfy the following PDE,*

$$F(x) \cdot \nabla \phi_\lambda(x) = \lambda \phi_\lambda(x), \tag{11}$$

if M is compact, and $\phi_\lambda : M \to \mathbb{C}$, is in $C^1(M)$, or alternatively, if $\phi_\lambda \in C^2(M)$.

The idea behind this theorem recalls the chain rule, $\dot{g}(x)$ for an orbit $x(t) = \varrho_t(x)$, so, (using the component notation, $x_i = [x]_i$ for the i^{th} component of x),

$$\frac{d}{dt}\phi(x(t)) = \sum_i \frac{\partial \phi}{\partial x_i}\frac{dx_i}{dt} = \sum_i \frac{\partial \phi}{\partial x_i} F_i(x) = \nabla \phi \cdot F(x). \tag{12}$$

However, the proof refers to the infinitesimal generator. This PDE, Eq. (11) is classified as a quasi-linear type, and the solution thereof is well handled by the method of characteristics as discussed in many classical textbooks on PDEs, [12]. We review this method of characteristics solution approach in this context in Appendix 8.

In characterizing the set of eigenfunctions, it is useful to recall some of the standard spectral theory of linear operators.

Remark 1 *From spectral theory we review, [36, 1, 13], the resolvent set of a linear operator $\mathcal{L} : \mathcal{F} \to \mathcal{F}$ consists of those functions in a Hilbert space $\mathcal{F}$ for which the resolvent exists. That is, when the operator,*

$$L_\lambda(\mathcal{L}) = (\mathcal{L} - \lambda I), \text{ as a function of } \lambda \in \mathbb{C}, \tag{13}$$

has an inverse, and when that operator exists call it, $\mathcal{R}_\lambda(\mathcal{L})$. The Koopman PDE, Eq. (11), may be stated in terms of the resolvent operator, as

$$L_\lambda(\mathcal{L})\phi_\lambda(x) = (F \cdot \nabla - \lambda I)\phi_\lambda(x) = 0. \tag{14}$$

Definition 3 *The set of all complex numbers λ such that the resolvent operator $\mathcal{R}_\lambda(\mathcal{L})$ 1) exists, 2) is bounded, and 3) is densely defined is called the resolvent set, $\rho(\mathcal{L}) \subset \mathbb{C}$. The complement is called the spectrum of $\mathcal{L}$, denoted $\sigma(\mathcal{L})$. By complimentary definition, the type of failure mode according to each of each of these properties classifies the type of spectrum:*

1. *If $\mathcal{R}_\lambda(\mathcal{L})$ fails to exist, as there is no inverse for those λ, then declare the point spectrum, $\lambda \in P_\sigma(\mathcal{L}) \subset \mathbb{C}$.*

2. *If $\mathcal{R}_\lambda(\mathcal{L})$ exists and it is densely defined, but is not bounded, then declare the corresponding $\lambda \in C_\sigma(\mathcal{L})$, the continuous spectrum.*

3. *If $\mathcal{R}_\lambda(\mathcal{L})$ fails to be densely defined for a given λ, then declare the residual spectrum, $\lambda \in R_\sigma(\mathcal{L})$.*

Clearly, $\sigma(\mathcal{L}) = P_\sigma(\mathcal{L}) \cup C_\sigma(\mathcal{L}) \cup R_\sigma(\mathcal{L})$. Now we come to the point of this remark: the names of these spectral sets are poorly chosen in the sense that they may be described as misleading, not indicating the cardinality of the sets. Specifically, there are problems where the continuous spectrum is not itself "continuous" as in some problems it might be uncountable, but in other problems, it may well be countable, or even finite. And likewise, there are examples of operators where the point spectrum may be finite, countable, or even uncountable, [1]. So as not to be distracted by the names as alluding to cardinality, in particular we point out that the point spectrum of the Koopman operator is generally uncountable, and the next two sections reveal this fact. More will be said in Appendix 9 regarding the role of the domain of the dynanamical system and the consequences to the spectral decomposition.

Now we state solutions of the Koopman PDE, as corollary to Theorem 1, the details of which are proved in the Appendix 8, as a straight forward application of the method of characteristics for this quasi-linear PDE. The solution by method of characteristics [12] in this Koopman PDE setting follows from discussion in [4], and allowing larger domains at least as large as nonrecurrence where the following formula was also noted [17] to prevent the complication of defining multi-valuedness. We define that a set P is nonrecurrent in a time epoch time interval $[t_1, t_2]$ that contains zero, if for every $z \in P$, if $t_1 \leq 0 \leq t_2$, then $orbit(z, t_1, t_2) \cap P$ is a connected subset (curve) in P, (let $orbit(z, t_1, t_2) = \cup_{t \in [t_1, t_2]} \varrho_t(z)$).

Theorem 2 *Let the Koopman eigenfunction PDE, Eq. (11) be defined for an ODE, $\dot{x} = F(x)$, with a flow $x(r) = \varrho_r(x_0) : M \times \mathbb{R} \to M$. Assume a closed initial data manifold (curve) $\Lambda \subset M$ that is nonrecurrent for some time epoch, $r \in [t_1, t_2]$, that contains 0, and let $U = \cup_{t \in [t_1, t_2]} \varrho_t(\Lambda)$ be the resulting nonrecurrent closed domain. Furthermore, let a $C^1(\Lambda)$ initial data function be defined, $h : \Lambda \to \mathbb{C}$. (See Fig. 2). Then an open-KEIGS pair, $(\lambda, \phi_\lambda(x)), \phi_\lambda : U \to \mathbb{C}$ has the form,*

$$\phi_\lambda(x) = h \circ s^*(x) e^{\lambda r^*(x)}, \tag{15}$$

where $r^*(x)$ is the "time"-of-flight such that for a point $x \in U$, there is an intersection in U by pull back to the data surface Λ,

$$r^*(x) = \{r : \varrho_{-r}(x) \cap \Lambda \neq \emptyset\}. \tag{16}$$

For each $x \in U$,

$$s^*(x) = s \circ \varrho_{-r^*(x)}(x), \tag{17}$$

is the parameterization on Λ of that first intersection point.

See Fig. 2 as a descripiton of an "open"-eigenfunction [4, 26] which is easily understood as resulting from propogating the data by the method of characteristics. The domain of the particular eigenfunction extends,forward and backward in time. Notice that by asserting nonrecurrence of Λ, that U is constucted to be nonrecurrent. Each $x \in U$, $\{r : \varrho_{-r}(x) \cap \Lambda \neq \emptyset\}$ will consist of a single value, which may be positive or negative depending on if it is upstream or downstream of Λ. Eq. (16) is written as a set, but it has a single value for each x, by assumption of nonrecurrence.

Proof 1 *The proof is in Appendix 8, Eqs. (78)-(83), as this is a straightforward application of the method of characteristics, [12]. See Fig. 2, where we illustrate that a general solution of the eigenfunction PDE is simply a pull back along the flow through x, to read the data on Λ, and then scale it according to the linear action of $(e^\lambda)^r$, for the "time" it takes to pull back the point $r = r^*(x)$.*

from $x \in U$, "pull"-back along the flow from x to Λ, where initial data h is defined. Therefore, the value of ϕ_λ at x is h but scaled as a function of the time of the pullback: $e^{\lambda r}$.

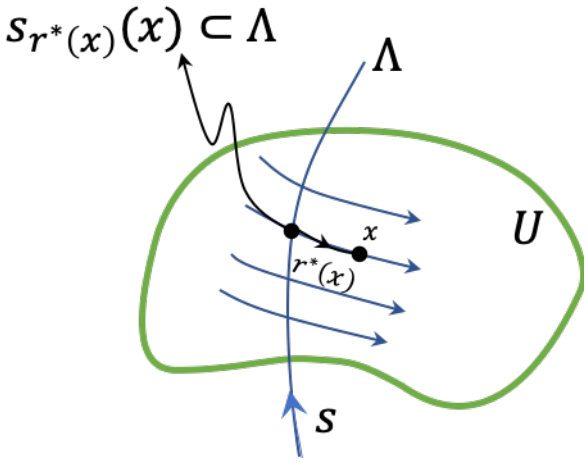

Figure 2: A general eigenfunction, Eq. (15) solution of Eq. (11) is defined in terms of measuring the initial data $h(s)$ on the data surface Λ, and for each x there is a unique point where that measurement is taken. Then the solution ϕ_λ is a linear scaling of that measurement, by the time of the pull back.

Remark 2 *Given a point $x \in U$, finding $r^*(x)$ is a task that is computationally comparable to a shooting problem, pulling x back to the data curve Λ along the flow, and recording the "time"-of-flight. For some problems where the flow is known in closed form, this may be solvable analytically. On the other hand, a "Poincare'" section type solver can be developed to numerically continue from initial condition x by the backward flow, $\dot{x} = -F(x)$. From this follows $s^*(x)$ by Eq. (17).*

The simple form of the eigenfunction solutions, $\phi_\lambda(x)$, Eq. (15), has an equally simple geometric interpretation. Notice that since $s^*(x)$ is constant along the flow, then $h \circ s^*(x)$ is also constant along the flow. Therefore,

$$\phi_\lambda(x) e^{-\lambda r^*(x)} = h \circ s^*(x). \tag{18}$$

Therefore these functions have **level sets along integral curves** (orbits) which can said another way:

Corollary 1 *If x and $\tilde{x}$ are on the same orbit in U, so that $x = \varrho_r(\tilde{x})$, then,*

$$\phi_\lambda(x) = \phi_\lambda(\tilde{x}) e^{\lambda r}, \tag{19}$$

where r is the "time" to flow from $\tilde{x}$ to x.

This statement reduces to the general form Eq. (15), on the data set $\tilde{x} \in \Lambda$. The dynamics evolves the eigenfunction as simple linear scaling of initial data, allowing when real valued monotone increasing or decreasing, or when complex valued eigenvalues, this linear scaling includes oscillations. In this sense, the observable dynamics of an eigenfunction, are especially simple as defined by the level sets of $\phi_\lambda(x) e^{-\lambda r^*(x)}$.

An interesting consequence is to note that not all seemingly complicated initial data functions, $h : \Lambda \to \mathbb{C}$ are "really" complicated even if they seem so, meaning they may not have started that way. That is, there may well be another transverse data curve, $\tilde{\Lambda}$ such that the data on Λ propagates and may seem less (or more) complicated on $\tilde{\Lambda}$. Specifically,

Corollary 2 *If $h(s) : \Lambda \to \mathbb{C}$ data is defined on Lambda $\in U$, then the corresponding equivalent data on $\tilde{\Lambda} \in U$ is a function $\tilde{h}(\tilde{s}) : \tilde{\Lambda} \to \mathbb{C}$ if corresponding points $\tilde{x} \in \tilde{\Lambda}$ flow to $x \in \Lambda$, $x = \varrho_{r^*(x)}(\tilde{x})$. Then with the same smoothness as the flow, there exists a transformation, $\alpha : \tilde{\Lambda} \to \Lambda$ such that $\tilde{h}(\tilde{s}) = h \circ \alpha(s)$.*

This is just a simple reinterpretation of Eq. (19). Consider that for each $x \in U$, then $(r^*(x), s^*(x))$ has already been defined as the pull back time and parameterized position on Λ, and h can be evaluated there. Or likewise, $(\tilde{r}(x), \tilde{s}(x))$ are the pull back onto $\tilde{\Lambda}$.

Consider the linear example in the Appendix, Eqs. (70), (73) based on the linear system whose eigenfunctions are derived from two different transverse surfaces. Λ is a horizontal line, and then $\tilde{\Lambda}$ is a vertical line. In one case, we got the observer function (a_2, x_2) by choosing $\lambda_2 = a_2$ and a constant data function, but (a_1, x_1) in the second case with $\lambda_1 = a_1$ but again the constant function. See Fig. 3.

$$\tilde{s} = s^{-\frac{a_2}{a_1}}, \tag{20}$$

and so,

$$\tilde{h}(\tilde{s}) = h(s^{-\frac{a_2}{a_1}})s^{\frac{\lambda}{a_1}}. \tag{21}$$

defines the data on a vertical data curve Λ from data originally stated on a horizontal curve Λ. This is a general principle that initial data curves and data can be otherwise stated elsewhere in the domain, and even sometimes more conveniently.

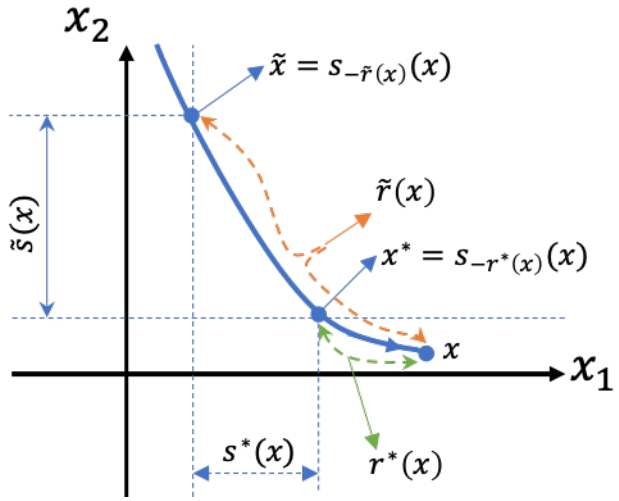

Figure 3: Data h stated on one data curve Λ can be pulled back to restated as $\tilde{h}$ on some other data curve $\tilde{\Lambda}$ here shown as the Example Eqs. (20)-(21) demonstrating the principle of Corollary 2.

3 Cardinality, First Discussion Follows The Algebraic Structure

There are at least two major reasons for the many, the uncountably many eigenfunctions. First there is a simple and well known, algebraic property of eigenvalues and eigenfunctions as follows. But this is not the exclusive source of these many eigenfunctions.

Proposition 1 *Algebraic Property. The KEIGS of a Koopman operator of a flow form a semigroup that is isomorphic to* $(\mathbb{R}^+, +)$ *as follows. For any* $0 < \alpha_1, \alpha_2 < \infty$, *then if* $(\lambda_1, \phi_{\lambda_1}(x))$ *and* $(\lambda_2, \phi_{\lambda_2}(x))$ *are a pair of KEIGS in an open domain* U, *then if* $(\lambda, \phi(x)) = ((\alpha_1\lambda_1 + \alpha_2\lambda_2), (\phi_{\lambda_1}(x))^{\alpha_1}(\phi_{\lambda_2}(x))^{\alpha_2})$ *and* $\phi_\lambda(x)$ *exists in* U *then it is also a KEIGS. [6]*

There are at least two proofs of this proposition in this setting of a continuous flow, which are essentially equivalent but appear different in presentation. The usual proof appeals to the composition property of the Koopman operator to which we refer to the citations. The second argument is based on the infinitesimal operator, meaning it appeals directly to the Koopman eigenfunction PDE which is in fact quasi-linear, this being the property that is essential to the result; our argument is straight forward but since we have not seen this particular direct argument before, we include these details in Appendix 10. For discrete time systems,

including Poincare' maps of continuous flows strobed at discrete times t, then the corresponding semigroup property also holds, with only a slight adjustment of the proof. In fact, an even stronger than the semi-group structure follows, in an open domain where the corresponding inverses are nonsingular that the semigroup is in fact an Abelian group, but such a statement my greatly restrict the domain without assumptions on the details of the function spaces.

As a direct consequence, powers of a KEIGS are also KEIGS as we see by choosing, $\phi_1 = \phi_2$ in Proposition 1. If $(\lambda, \phi_\lambda(x))$ is a KEIGS, then for any $s > 0$, then $(\lambda^s, (\phi_\lambda(x))^s) = (\lambda^s, \phi_{\lambda^s}(x))$ is also a KEIGS. Consider a special case, where $\phi_1 = \phi_2 = 1$ and we see that $\lambda_1\lambda_2$ is the eigenvalue of the product of eigenfunctions, $\phi_{\lambda_1}(x)\phi_{\lambda_1}(x)$. Even more specialized, with the same eigenfunction, we see that λ_1^2 is an eigenvalue of $\phi_{\lambda_1}(x)^2$. Squares of eigenfunctions, or likewise any other power for that matter, are eigenfunctions.

Remark 3 Example. *In Appendix 8 example 9, follows that $\dot{x} = ax$, has the state observer as KEIGS: $(\lambda, \phi_\lambda(x)) = (a, x)$. Therefore, by Proposition 1, the following are also examples of KEIGS, $(\lambda, \phi_\lambda(x)) = (2a, x^2)$, $(\lambda, \phi_\lambda(x)) = (3a, x^3)$, etc., as well as $(\lambda, \phi_\lambda(x)) = (\frac{1}{2}a, \sqrt{x})$, etc, in appropriate domain for x. All of these follow by exponentiation directly from Proposition 1, but also this is in agreement with the direct derivation of the KEIGS in Eq. (15) that yields, for $h = 1$, $\phi_\lambda(x) = cx^{\frac{\lambda}{a}}$ for any $\lambda \in \mathbb{C}$.*

The following is immediate due to the algebraic property associating KEIGS with exponentiation.

Corollary 3 *The cardinality of KEIGS is at least uncountable.*

However, this algebraic property generating eigenfunctions is not the only source counting "new" KEIGS, and furthermore, in terms of a sensible dimensional definition of new, that is "primary", these are *not really new*. We will define in the next section what we mean by *not really new*. Furthermore, another source of KEIGS due to freedom in choosing data functions on transvers curves generates in fact a cardinality greater than the uncountable generated by the algebraic property just reviewed. We discuss these in the next section.

Interestingly, $P_\sigma(\mathcal{L}) = \mathbb{C}$, meaning the point spectrum is uncountable. Therefore, the continuous spectrum is empty, $C_\sigma(\mathcal{L}) = \emptyset$. In fact this can be more nuanced in that the nature of the spectral decomposition, point, continuous and residual spectra are also balanced on the nature of the domains chosen, particularly if the subdomain is recurrent, or further restricted to nonrecurrent, as we explore much more extensively in Appendix 9. Furthermore, generally in spectral theory, there exist examples where any of these three spectrum may be empty, finite, infinite but countable, or uncountable. So the names themselves may seem misleading, particularly as the name point spectrum suggest countable, and continuous spectrum suggests a continuum, counter to what may be the truth of their cardinal nature.

4 Primary Eigenfunctions Complete the Geometry

While the algebraic property generates at least uncountably many KEIGS, and as we have argued recently [4] and likewise others have pointed out [17], there is geometry encoded in the eigenfunctions, we will now argue that there is no geometry to be generated beyond the underlying Riemannian manifold. So how do we resolve these seemingly conflicting statements, that there are so many eigenfunctions, perhaps infinitely many, but perhaps only a relatively low dimensional manifold? To this end we now define *primary eigenfunctions,* which is related to, but not identical to "principal" eigenfunctions as described in Mohr-Mezic, [28], in a manner meant to distinguish geometry versus algebra and to be described shortly.

Our definition of primary KEIGS, will be in terms of the set of level sets. First recall a definition of level sets. For a general function, $f : U \to V$, we use the notation for a level set, of level c in the range V, from the domain U to be the set,

$$L_c(f) = \{x : f(x) = c, x \in U\}. \tag{22}$$

If $c \notin V$, then $L_c(f) = \emptyset$.

Now we can define:

Definition 4 *(Primary KEIGS). For any one KEIGS pair, $(\lambda, \phi_\lambda(x))$, $\lambda \in \mathbb{C}, x \in M$, we may form an equivalence relationship to any other KEIGS pair $(\tilde{\lambda}, \tilde{\phi}_{\tilde{\lambda}}(x))$, $\lambda \in \mathbb{C}$, $x \in M$, that share the same set of level sets in U. That is, write $(\lambda, \phi_\lambda(x)) \doteq (\tilde{\lambda}, \tilde{\phi}_{\tilde{\lambda}}(x))$ iff in terms of the set of level sets, for each $c \in V$, there is a $\tilde{c} \in V$ such that $L_c(\phi_\lambda) = L_{\tilde{c}}(\phi_\lambda)$. That is the level sets match between the functions, although not necessarily corresponding to the same levels. Then, continuing to refer in terms of any one specific KEIGS, $(\lambda, \phi_\lambda(x))$, we define an equivalence class of KEIGS,*

$$\overline{(\lambda, \phi_\lambda(x))} := \{(\tilde{\lambda}, \tilde{\phi}_{\tilde{\lambda}}(x)) : (\tilde{\lambda}, \tilde{\phi}_{\tilde{\lambda}}(x)) \doteq (\lambda, \phi_\lambda(x))\} \tag{23}$$

That is, the set of KEIGS that are "$\doteq$" equivalent to a specific chosen (convenient) member $(\lambda, \phi_\lambda(x))$ of the equivalence class is a quotient set of functions.

There is a geometric statement that justifies this equivalence class, as follows.

Theorem 3 *Given a KEIGS, $(\lambda, \phi_\lambda(x))$, then any other KEIGS derived simply by exponentiation shares the same set of level sets, in the sense that for each level set of one, there is a level set of the other, although not necessarily representing the same level. Therefore, the algebraic property of eigenfunctions defines the primary KEIGS, at least in terms of inclusion.*

$$\{(\lambda^p, (\phi_\lambda(x))^p, p \in \mathbb{R}\} \subset \overline{(\lambda, \phi_\lambda(x))}. \tag{24}$$

Proof 2 *Let,*

$$L_c(\phi_\lambda) = \{x : \phi_\lambda(x) = c, x \in M\}, \tag{25}$$

be the c is a constant level set for a fixed $c \in \mathbb{C}$ in the range of $\phi_\lambda(x)$. Then,

$$L_{c^s}(\phi_\lambda)^s = \{x : (\phi_\lambda(x))^s = c^s, x \in M\}, \tag{26}$$

is the c^s constant level set of the eigenfunction $(\phi_\lambda(x))^s$ which is also an eigenfunction by the algebraic property. Algebraic manipulation clarifies that $L_c \subseteq L_{c^s}$, with equality at least when c is positive, or s is an even integer.

Remark 4 *For convenience, we use the phrase "primary KEIGS" to refer to the equivalence class of functions (that is the set of functions), but also for simplicity, sometimes we use the same phrase "primary KEIGS" when referring to a single eigenvalue, eigenfunction (KEIGS) pair that can be used to generate the entire set by exponentiation. We do this especially when describing a favorite (simple) such pair.*

For example, consider the linear example, $\dot{x} = ax$, Appendix 8, has in particular the primary KEIGS,

$$\overline{(a, x)} = \{(a^p, x^p) : p \in \mathbb{R}\}. \tag{27}$$

A particularly convenient member of this set follows the choice, $\lambda = a$. In this case, we call the corresponding eigenfunction $\phi_a(x) = x$ as "the state observer" since it is the identity function. For convenience we "overload the phrase," and drop the overline if stating that (a, x) is a primary KEIGS, also referring to the set of equivalent KEIGS. See Fig. 4, where we show several equivalent KEIGS in this primary KEIGS.

Now to the title of this section, that "Primary KEIGS complete the geometry", that we admit is only a necessary but not sufficient condition, as revealed by the following example.

Remark 5 Example 1: Non-Primary KEIGS of a Linear System, Despite Same Eigenvalues. *In the Appendix, Eqs. (70), (73), we showed that the linear system,*

$$\dot{x_1} = a_1 x_1, \dot{x_2} = a_2 x_2, \tag{28}$$

has a KEIGS in general form $\phi_\lambda(x_1, x_2) = x_2^{\frac{\lambda}{a_2}} h\left(\frac{x_1}{x_2^{(\frac{a_1}{a_2})}}\right)$, from an initial data curve Λ as a horizontal line, and this class includes the observer function, KEIGS, (a_2, x_2) when the initial data is chosen to be

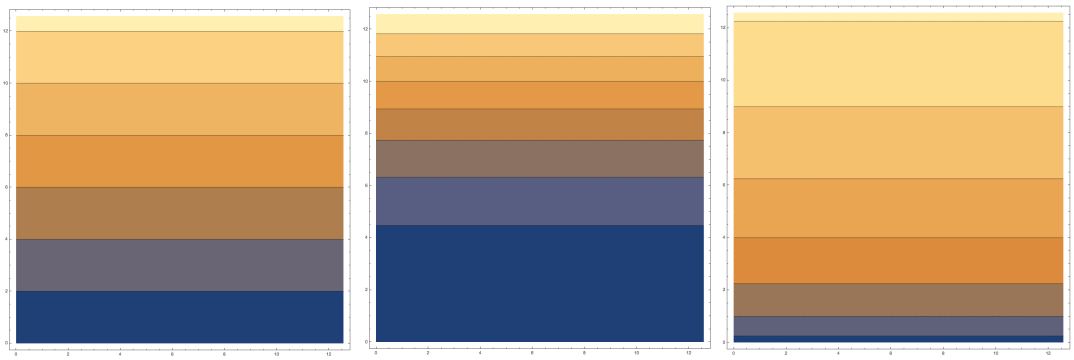

Figure 4: The algebraic property yields a family of KEIGS according to Theorem 3. Here, from the linear system, Eq. (28), results equivalent KEIGS, $(a_2, x_2), (a_2^2 x_2^2), (\sqrt{a_2}, \sqrt{x_2})$, shown left to right, as corresponding contour plots. Notice that the set of level sets are the same and in this sense we define the eigenfunctions equivalence, Eqs. (24)-(27), as defined according Definition 4 of primary KEIGS. Contrast this figure to Fig. 5.

a constant, $h(s) = 1$. However, if the initial data on Λ is not chosen as a constant, say for example, as $h(s) = s$, then

$$\phi_\lambda(x_1, x_2) = x_1 x_2^{\frac{\lambda - a_1}{a_2}}, \tag{29}$$

results. Or using the same eigenvalue, $\lambda_2 = a_2$,

$$\phi_{a_2}(x_1, x_2) = x_1 x_2^{\frac{a_2 - a_1}{a_2}}. \tag{30}$$

To see that these are **not** *part of the same primary KEIGS, despite that the* **same** *eigenvalues and same initial data curve, but different data functions are used, we must check that these result in different sets of level sets. Since the gradient is perpendicular to the level sets, and the perp-gradient points along level sets, we can check that,*

$$\nabla(x_1 x_2^{\frac{a_2 - a_1}{a_2}}) \cdot \nabla_\perp(x_2) =< x_2^{\frac{a_2 - a_2}{a_2}}, \frac{(a_2 - a_1)}{a_2 x_1 x_2^{\frac{a_1}{a_2}}} > \cdot < 1, -0 >= x_2^{\frac{a_2 - a_1}{a_2}} :\neq 0. \tag{31}$$

For example, if the system were such that $a_1 = 1, a_2 = 2$, then this dot product is clearly spacially varying, as $\sqrt{x_2}$. So transversality is generally easy to construct even for simple examples. See Fig. 5.

Remark 6 Example 2: Non-Primary KEIGS of a Nonlinear System

Consider the following ODE in the plane,

$$\begin{aligned} \dot{x_1} &= -x_2 + x_1(\mu - x_1^2 - x_2^2) \\ \dot{x_2} &= x_1 + x_2(\mu - x_1^2 - x_2^2), \end{aligned} \tag{32}$$

which is a well known ODE as it is often used as a normal form to present the unfolding of a Hopf bifurcation, [29]. Polar coordinates conveniently follows the change of variables, $x_1 = r cos(\theta), x_2 = r sin(\theta)$,

$$\begin{aligned} \dot{r} &= r(\mu - r^2), \\ \dot{\theta} &= 1. \end{aligned} \tag{33}$$

The supercritical Hopf bifurcation occurs at $\mu = \mu_c = 0$, whereafter, $\mu > \mu_c$ there is a stable limit cycle. In Fig. 6 we show a stream plot for many initial conditions in and around the limit cycle, choosing $\mu = 1$. This makes a good case to study the numerics of using the Koopman PDE to find eigenfunctions by a numerical

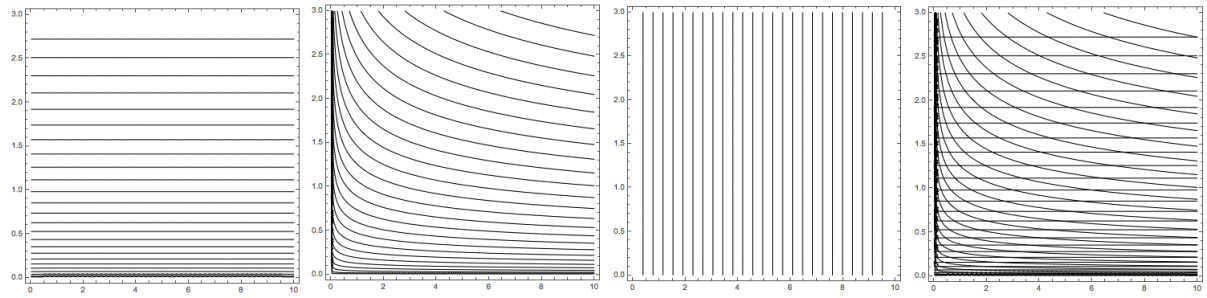

Figure 5: Non-equivalent KEIGS. Following Example 1, KEIGS due to same eigenvalue and same initial data set Λ but different initial data functions, $h, \tilde{h} : \Lambda \to \mathbb{C}$ result in KEIGS whose level sets are not equivalent, and therefore they generate different primary KEIGS. Here clearly non-equivalent KEIGS result from using the simple linear system, Eq. (28), and the same Λ is the horizontal line at x_2=1, and the same eigenvalue, $\lambda = a_2$, *but differing data functions.* Left uses $h(s) = 1$ yielding $\phi_{a_2}(x_1, x_2) = x_2$ (that is the eigenfunction usually discussed, and it is usually called the "observer" function), but next left to right result from data $h(s) = s^2$, $h(s) = s$ yielding $\phi_{a_2}(x_1, x_2) = x_1 x_2^{\frac{a_2 - a_1}{a_2}}$, and $\phi_{a_2}(x_1, x_2) = x_1^2 x_2^{\frac{a_2 - 2a_1}{a_2}}$, (and if for example $a_2 = 2, a_1 = 1$, then specifically $\phi_{a_2}(x_1, x_2) = x_1\sqrt{x_2}$, and $\phi_{a_2}(x_1, x_2) = x_1^2$). Observe the differing level sets, as visually we see clearly by this geometry that $\nabla(x_1 x_2^{\frac{1}{2}}) \cdot \nabla_\perp(x_2) \neq 0$ in Eq. (31). The right figure shows the transverse intersections of the first two eigenfunctions on the left that were due to $h(s) = 1$ and $h(s) = s^2$.

integration method. We can contrast the numerical solution to closed form computation as this nonlinear problem is still analytically solvable. By separation of variables, using initial condition $r(t_0) = r_0, \theta(t_0) = \theta_0$,

$$t - t_0 = \frac{2 \ln r(\ln r_0^2 - \mu)}{\ln r_0(r^2 - \mu)}, \tag{34}$$

which if choosing, $t_0 = 0$, *gives,*

$$r(t) = \frac{e^{\mu t}\sqrt{\mu} r_0}{\sqrt{\mu - r_0^2 + e^{2\mu t} r_0^2}}. \tag{35}$$

To state an eigenfunction by Theorem 2, we must choose an initial data curve Λ *and there are many possible choices leading to different domains with of course specific KEIGS depend on the domain. Choosing for example,* Λ *to be a circle of fixed radius centered on the origin, but* $R > \sqrt{\mu}$, *so larger than the radius of the limit cycle, then the eigenfunction is,*

$$\phi_\lambda(r_0, \theta_0) = h(\theta_0 + t) e^{\frac{2\lambda \ln R(\ln r_0^2 - \mu)}{\ln r_0(r^2 - \mu)}}. \tag{36}$$

Given a cartesian point initial point, (x, y), *we compute the corresponding initial* (r_0, θ_0) *as usual,* $r_0 = \sqrt{x^2 + y^2}$, $\theta_0 = \tilde{tan}^{-1}(x, y)$ *(using the often used 4-quadrant aware arctan2 function which we write as* $\tilde{tan}^{-1}$ *rather than the "common"* tan^{-1} *function), and so,*

$$\phi_\lambda(x_1, x_2) = h(\tilde{tan}^{-1}(x_1, x_2) + t) e^{\frac{2\lambda \ln R(\ln(x_1^2 + x_2^2) - \mu)}{\ln \sqrt{x_1^2 + x_2^2}(x_1^2 + x_2^2 - \mu)}}. \tag{37}$$

See Fig. 6, where it is clear that the level sets of these eigenfunctions, even due to same eigenvalues, but different initial data h, are not coincident. So by definition, the primary KEIGS they generate are not equivalent. The generality of this observation is straightforward as summarized by the following statement:

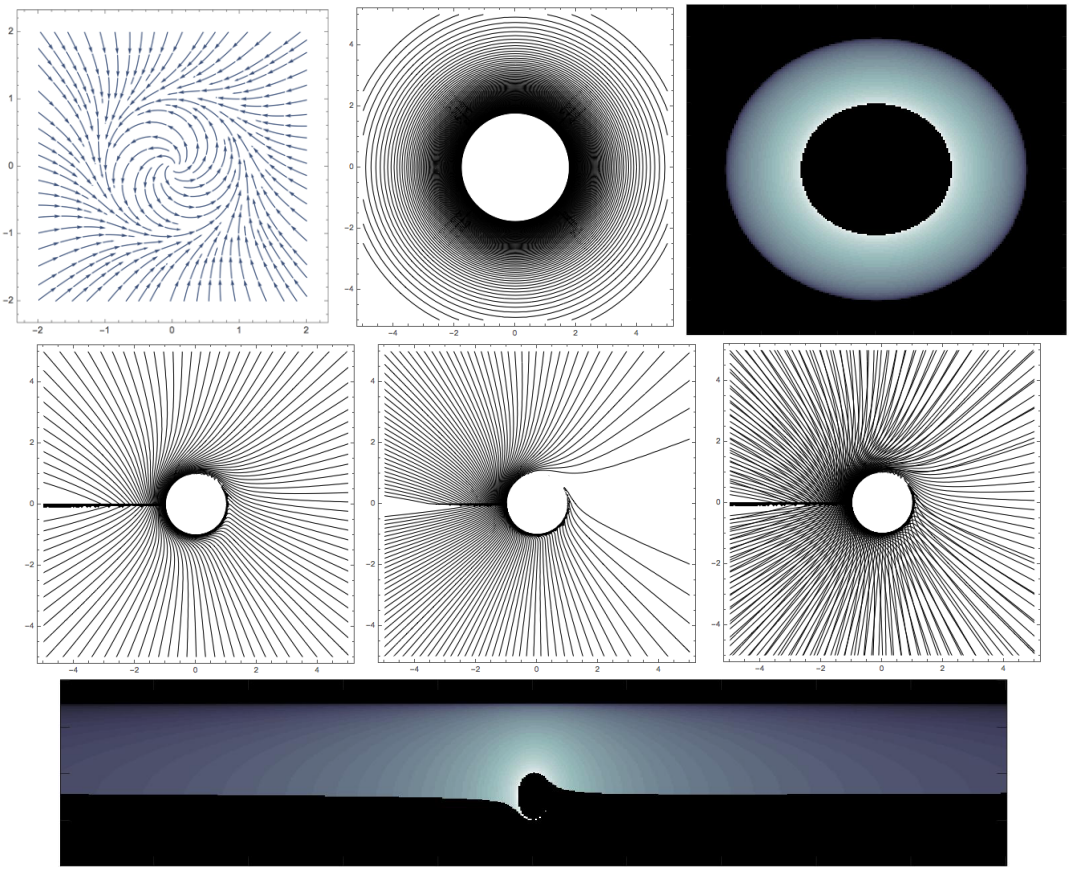

Figure 6: Considering the vector field of the normal form of a Hopf bifurcation, Eq. (32), (Upper left) A stem plot for $\mu = 1$ we see a limit cycle at radius $r = 1$, (Upper middle) A data curve Λ is a circle of radius $R > 1$ then an eigenfunction ϕ_λ follows, here shown in an open domain U that is the annulus, $r < \sqrt{x^2 + y^2} < R$, and for nitial data $h(s) = 1$. (Upper right) The same eigenfunction computed by numerical shooting backward to Λ for $(x, y) \in U$. (Middle left) for data function $h(s) = s$ on the same circular data set Λ and notice the level sets are clearly transverse to those of upper middle. (Middle) Level sets of ϕ_λ resulting from data $h(s) = s^2$ on the same circle $R > r$, (Middle right) notice that the level sets of the eigenfunction ϕ_λ from $h(s) = s$ and $h(s) = s^2$ clearly differ as hey cross transversally. (Bottom) Now the data set is chosen $\Lambda =$ a horizontal line at $y = 1$, and $h(s) = 1$ shown, compute again by numerical shooting.

Theorem 4 *Given KEIGS with the same eigenvalue λ, and different data functions $h, \tilde{h}$ on an otherwise same initial data set, $h, \tilde{h} : \Lambda \to \mathbb{C}$, we expand the KEIGS notation, $(\lambda, \phi_{\lambda,h,\Lambda}(x))$, and $(\lambda, \phi_{\lambda,\tilde{h},\Lambda}(x))$. These KEIGS are generally not equivalent unless a special circumstance of a data compatibility Eq. (39) and (40) is satisfied.*

Summarizing, eigenfunctions depend also on the data curve and also the specific data function on the data curve. Our full notation is expanded to include these, $\phi_{\lambda,h,\Lambda}(x)$, noting that it is possible that likely there is no equivalence, $\phi_{\lambda,h,\Lambda}(x) \dot{\neq} \phi_{\lambda,\tilde{h},\Lambda}(x)$, even with the same eigenvalue λ and even if the same data curve Λ, but differing data functions, $h \neq \tilde{h}$. It is however true that some data h on a given data curve Λ pulls back to be stated elsewhere in the domain by other data curves $\tilde{\Lambda}$ pulled back accordingly to a new $\tilde{h}$, already noted at Eq. (21). Nonequivalence is not certain, and we give the condition below, Eqs. (38)-(40). Nonetheless, when convenient and clear we suppress the fuller notation of subindices and write simply, ϕ_λ.

Remark 7 *We have already noted that the cardinality of KEIGS is at least uncountable due to the algebraic property (first discussion), but now (second discussion) we notice that even with the quotient, the cardinality of primary KEIGS may be greater than uncountable due to the cardinality of initial data functions h, and the possibility that these may produce distinct primary KEIGS.*

Consider if $\nabla\phi_{\lambda,h} \cdot \nabla_\perp \phi_{\lambda,\tilde{h}} \neq 0$. Referring to the general solution of a KEIGS for a specifically chosen initial data function on a data surface defined, $h(s)$ on Λ, then by Eq. (15), with notation extended to include these relevant information, $\phi_{\lambda,h,\Lambda}(x) = h \circ s^*(x) e^{\lambda r^*(x)}$. Likewise, for a different data function $\tilde{h}(s)$, but on the same data surface, $\phi_{\lambda,\tilde{h},\Lambda}(x) = h \circ s^*(x) e^{\lambda r^*(x)}$, observe that for a given point x these share the same point $s^*(x)$, and time of flight back along the flow to Λ, $r^*(x)$ (as defined in Theorem 2). Then consider the orientation of level sets of $\phi_{\lambda,h,\Lambda}(x)$ relative to those of $\phi_{\lambda,\tilde{h},\Lambda}(x)$ at a point $x = (x_1, x_2)$. We have specialized this discussion to two dimensions for simplicity, but the outcome is general and similar to show. To decide if level sets of $\phi_{\lambda,h,\Lambda}$ are tangent to level sets of $\phi_{\lambda,\tilde{h},\Lambda}$ is described by,

$$\nabla\phi_{\lambda,h,\Lambda} \cdot \nabla_\perp \phi_{\lambda,\tilde{h},\Lambda} = \lambda e^{2\lambda r^*(x)}(-\tilde{h} \circ s^*(x) h' \circ s^*(x) + h \circ s^*(x) \tilde{h}' \circ s^*(x))(\frac{\partial s^*(x)}{\partial x_2}\frac{\partial r^*(x)}{\partial x_1} - \frac{\partial s^*(x)}{\partial x_1}\frac{\partial r^*(x)}{\partial x_2}) =$$

$$= \lambda e^{2\lambda r^*(x)}(< h(y), \tilde{h}(y) > \cdot < \tilde{h}'(y), -h'(y) >)|_{y=s^*(x)}(< \frac{\partial r^*(x)}{\partial x_1}, \frac{\partial r^*(x)}{\partial x_2} > \cdot < \frac{\partial s^*(x)}{\partial x_2}, -\frac{\partial s^*(x)}{\partial x_1} >) = 0. \quad (38)$$

Thus at $y = s^*(x)$, for $x \in U$ but $y \in \Lambda$, the following data compatibility statements relate to the question of tangency,

$$< h(y), \tilde{h}(y) > \cdot < \tilde{h}'(y), -h'(y) >= 0 \quad (39)$$

or the following statement describing how Λ is oriented with respect to the flow,

$$< \frac{\partial r^*(x)}{\partial x_1}, \frac{\partial r^*(x)}{\partial x_2} > \cdot < \frac{\partial s^*(x)}{\partial x_2}, -\frac{\partial s^*(x)}{\partial x_1} >= 0, \quad (40)$$

that should be prevented if Λ is chosen to be transverse to the flow, then $\phi_{\lambda,h,\Lambda}$ and $\phi_{\lambda,\tilde{h},\Lambda}$ have tangential level sets, despite different data h and $\tilde{h}$. These computations serve to prove Theorem 4. Again, for intuition we refer to Figs. 5-6.

Remark 8 *It is clear that if a set of functions share the same set of level sets, then any linear combination of those functions also shares the same set of level sets; Stated formally, if,*

$$g \in span(\overline{(\lambda, \phi_\lambda(x))}, \quad (41)$$

then,

$$g \in \overline{(\lambda, \phi_\lambda(x))}. \quad (42)$$

Said conversely, it is not possible to build variation in a direction transverse to a linear combination of a set of equivalent KEIGS. Likely it is inefficient to build such variation using KEIGS that are almost equivalent in the sense that their level sets are not significantly transveral but instead perhaps almost tangent. Thus we are motivated to identify pairwise primary KEIGS such that not only their level sets are transverse to each other (definitive of sets of functions forming two different primary KEIGS) but "decisively" so, as a pre-requisite to building efficient representation of general functions by eigenfunctions, as Definition 2. In the next section we will define an optimization based method in the spirit of POD to generate such functions.

5 Numerical Methods: Optimal Representation by KEIGS

We have described above how there are so many eigenfunctions, generally at least uncountably many. However, above we also describe about how many pairings of eigenfunctions present either the same, or almost the same set of level sets, that we define as an equivalence called primary KEIGS. In [17], authors speak of having enough eigenfunctions for complete representation of arbitrary functions as linear combinations of eigenfunctions, following a phrase "richness", as we also use a similar phrasing in [22]. But in our work we described how to develop an efficient basis set in terms of a machine learning based "dictionary" concept that we called EDMD-DL, [22].

In [8], the authors have developed a variant of EDMD where the observation functions are themselves Koopman eigenfunctions, that they called Koopman Eigenfunction Extended Dynamic Mode Decomposition (KEEDMD). In broad sense this is what we do here, but with the extra technology so as to make for an optimally efficient representation. Therefore in the spirit of naming DMD methods with acronyms, the method we present here might be called, oKEEDMD, the extra "o" for optimal. Also let us point out that our construction of eigenfunctions is very different than any other approach in that we restrict out construction, sufficiently, to building a (an optimal) data function h on a surface Λ that is transverse to the flow.

Now in this paper we are positioned to approach a simple problem for practical applications, which is to ask how few eigenfunctions, optimally how few, can we use for a quality of estimate. Here we present a method to construct empirical eigenfunctions to solve a problem of efficient representation, and in some sense this problem reminds us of the celebrated POD-KL (Principal Orthogonal Mode - Karhunen-Loeve) concept that is popular in spatiotemporal data driven methods for constructing time-averaged optimal modes often from PDE data, [14, 23, 32, 11, 31]. There are distinct differences, most notably our representation is in terms of Koopman eigenfunctions. Nonetheless, what is key in our construction is that while there may well be at least uncountably many eigenfunctions, we have freedom to choose those we like. So we assert an efficiency principle to choose good ones, and thus follows the title of the paper regarding a good dictionary. See Definition 2.

In [17], the authors present a theorem that asserts that an arbitrary continuous functions can be represented as linear combinations of eigenfunctions. At the heart of the proof of their theorem is the idea (we now describe using our notation here) that eigenfunctions in a nonrecurrent domain Ω may be constructed on a transverse data surface $\Lambda \subset \Omega$, and a general (not necessarily eigen) function $q : \Omega \to \mathbb{C}$, that pulls back to Λ where careful choice of the data h allows us optimal efficiency.

Now we will address efficiency of a Koopman eigenfunction representation. In the context of the geometric aspects described in earlier sections, many pairs of eigenfunctions may not be significantly distinguished from each other in the sense that their level sets may either coincide or may be almost not transverse, so general representations may not be efficient. So we pose the efficient optimal representation by eigenfunctions problem as follows. Given $q : U \to \mathbb{C}$, let,

$$\varphi = arg \min_{\lambda,h} \ \|\phi_{\lambda,h} - q\|^2_{L^2(U)} \tag{43}$$

be the eigenfunction $\varphi : U \to \mathbb{C}$ with eigenvalue λ that most closely estimates $q \in L^2(U)$. The domain stated $U = \{\varrho_t(x), x \in \Lambda, t \in [t_1, t_2]\}$, and for presentation here, let $t_1 = 0$, but as long as $0 \in [t_1, t_2]$ the construction proceeds similarly. So that we may refer to Theorem 2, we will describe this problem in the

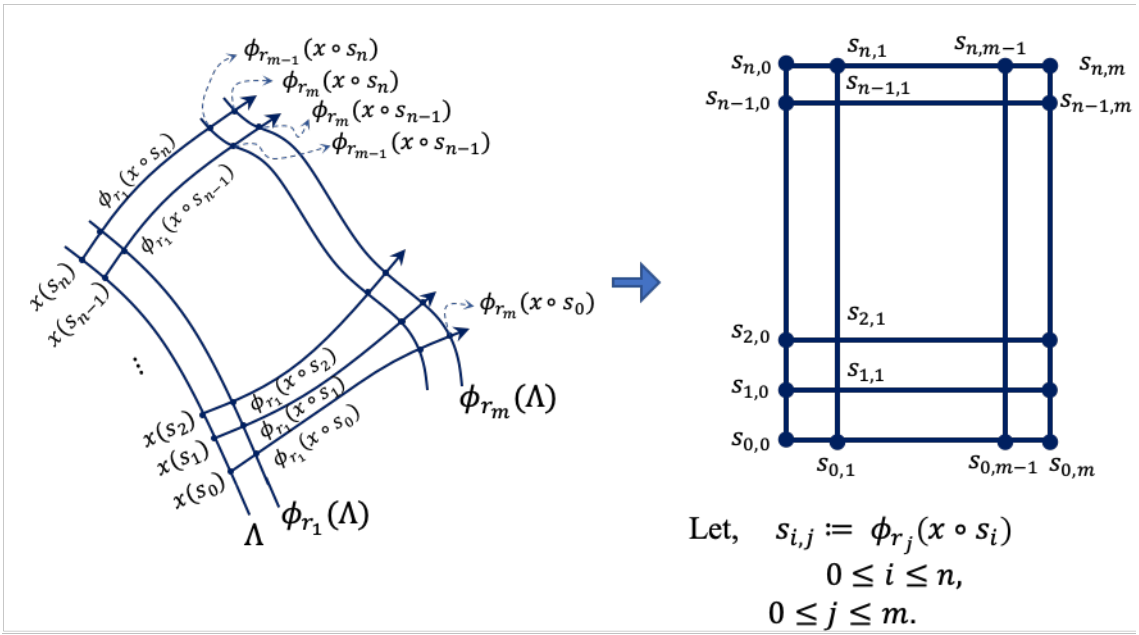

Figure 7: Numerical optimization on a gridded sample of the optimal data function $h : \Lambda \to \mathbb{C}$ in terms of Eq. (43) proceeds on a grid, propagated by the flow to sweep out the domain of fitting, $U = \{\varrho_t(x), x \in \Lambda, t \in [t_1, t_2]\}$, which by Eqs. (46)-(50) leads to a least squares problem for h on an equal time grid (right), interpreted in the phase space (left). Repeated application of the least squares problem Eq. (55) leads to a basis of successively optimal eigenfunctions, Eq. (57).

Sobolov space $\phi_{\lambda,h} \in H^2(U)$. Notice that the optimization is over eigenvalues λ and data functions h, on some specific data curve Λ, so this is quite different from a typical eigenfunction optimization problem that may directly attempt to construct the eigenfunction $\phi_{\lambda,h}$. The argmin will produce sufficiently just a data function h (and corresponding λ) and so that is co-dimension-one (lower dimensional) optimization problem as $\Lambda \in U$.

Recall that this minimization problem coincides with maximal optimal projection as follows,

$$arg\max_{\lambda,h}\ (q,\phi_{\lambda,h}) = arg\min_{\lambda,h}\ \|\phi_{\lambda,h}-q\|^2_{L^2(U)}, \tag{44}$$

using norm and inner product notation as usual,

$$\|f_1\|_{L^2(U)} = \int_U f_1(x)dx,\ \text{and},\ (f_1,f_2) = \int_U f_1(x)f_2(x)dx,\ \text{for any}\ f_1,f_2 \in L^2(U). \tag{45}$$

See Fig. 7.

We proceed to interpret and approximate this optimization problem to develop optimal KEIGS (oKEIGS) based on interpreting Theorem 2, Eq. (15), beginning with a grid on the data surface Λ propagating through the domain U. This leads to a least squares problem as follows. Let $s_0 < s_1 < ... < s_n$ be a uniform partition of a data curve Λ. (In a multivariate setting a general grid will comparably lead to a linear inverse problem, but we proceed to describe the idea in this simpler case that Λ is one-dimensional). Data $h : \Lambda \to \mathbb{C}$ is likewise partitioned and indexed,

$$h_i := h(s_i). \tag{46}$$

Over a uniform grid in time, $r_0 = 0 < r_1 < ... < r_m$, let,

$$S_{i,j} := \varrho_{r_j} \circ x(s_i),\ 0 \le i \le n,\ 0 \le j \le m, \tag{47}$$

in terms of the coordinate function into the domain $x(s) : \Lambda \to U \subset M$. This propagates by the flow $\varrho_r(\Lambda)$ for the gridded sample in space, and time, as shown in Fig. 7 indexed uniformly on time level of $s^*(x)$ starting from Λ, (defined in Eq. (17)), that are in turn level sets in h.

In terms of the grid, the optimization problem Eq. (43), for the function $\varphi(x)$ represented on the grid $\varphi \circ x(s_{i,j})$ is approximated by solving the finite rank least squares problem,

$$\tilde{\varphi} = arg\min_h \|\phi_{\lambda,h} \circ x(s_{i,j}) - q \circ x(s_{i,j})\|^2_F = arg\min_{h_i} \sum_{j=1}^{m}\sum_{i=1}^{n} |e^{\lambda r_j}h_i - q_{i,j}|. \tag{48}$$

The tilde, "˜" describes that the vector is an array representing the function on the grid $S_{i,j}$ at points Eq. (47) shown in Fig. 7, and

$$\tilde{\varphi}_{\lambda,h} \approx \varphi_{\lambda,h} \tag{49}$$

on that grid. Stating both functions q and $\tilde{\varphi}_{\lambda,h}$ are discretely sampled at

$$\varrho_{r_j} \circ x(s_i),\ 0 \le i \le n, 0 \le m,\ \ q_{i,j} := q(S_{i,j}). \tag{50}$$

The usual Frobenius norm for arrays, $\|v\|^2_F = \sum_{i,j} v^2_{i,j}$ is used. The problem becomes more convenient when the arrays are reshaped. To this end, we write the optimal initial data as an $n \times 1$ vector $h^o(\lambda) \in \mathbb{C}^n$ that solves a classical least squares problem,

$$h^o(\lambda) = arg\min_h \|A(\lambda)h - b\|_2, \tag{51}$$

where,

$$A(\lambda) = E(\lambda) \otimes I_n,\ \text{where},\ b = reshape(q, mn, 1), \tag{52}$$

and $E(\lambda)$ is an $m \times 1$ vector,

$$E(\lambda) = [e^{\lambda r_0}e^{\lambda r_1}...e^{\lambda r_m}]^t \in \mathbb{C}^m, \tag{53}$$

and I_n is the $n \times n$ identity matrix. With $\otimes$ as the Kronecker product, this makes $A(\lambda)$ an $mn \times n$ matrix consisting of the E vector repeated n times. Further, b is an $mn \times 1$ data vector describing the $n \times m$ array for the grid sample of the data function $q(S_{i,j})$. The eigenfunction approximation, but reshaped as an $mn \times 1$ vector, follows the the least squares solution, $h^o(\lambda)$,

$$p^o(\lambda) = A(\lambda)h^o(\lambda), \tag{54}$$

to be reshaped to the domain, as an $m \times n$ array of estimated function values, $\tilde{\varphi}^o(\lambda)$. Notice the arbitrary flexibility in choosing λ. As such we have written each of the expressions in Eqs. (55)-(53) as functions of λ. Any λ will lead to an optimal vector $h^o(\lambda)$ and thus $\tilde{\varphi}^o(\lambda)$. This reflects the same flexibility already noted, but now in other words. In fact, for a given q.

Finally, we summarize,

$$\begin{aligned} (\lambda_1^o, h_1^o) &= arg\min_{\lambda,h} \|A(\lambda)h - b\|_2, \\ p_1^o &= A(\lambda_1^o)h_1^o, \\ \tilde{\varphi}_1^o &= reshape(p_1^o, n, m). \end{aligned} \tag{55}$$

Notice the subindex "1" of $\tilde{\varphi}_1^o$ is placed to describe that further optimization may proceed as below to successively reduce the residuals.

See Fig. 8 showing example optimal solutions of a specific example problem. Let,

$$q(x) = 3e^{\frac{-(x_1^2+x_2^2)}{10}}, \tag{56}$$

$q : \mathbb{R}^2 \to \mathbb{R}$, chosen to challenge the method since it is not an eigenfunction, and also for an interesting shape. We will construct an approximation of this general function in terms of eigenfunctions from the Hopf-normal form problem already studied above, Eq. (32). In Fig. 8a we show a line segment in red, that denotes a segment of an initial data curve Λ, and over a time interval $0 \le t \le 2$ this curve sweeps out the fitting domain U shown outlined in blue, and we also showed a curve through the middle for reference, $U \subset M = \mathbb{R}^2$. The specific domain, as well as the specific function q to be fitted, and the specific dynamical system are all relevant to the specific optimal solution.

Continuing in the theme of estimation, we can successively fit the residuals,

$$\begin{aligned} R_k &= b - \sum_{l=1}^{k} p_l^o, \\ (\lambda_{k+1}^o, h_{k+1}^o) &= arg\min_{\lambda,h} \|A(\lambda)h - R_k\|_2, \\ \tilde{\varphi}_{k+1}^o &= reshape(p_k^o, n, m). \end{aligned} \tag{57}$$

It is convenient to normalize each of these eigenfunctions,

$$\overline{\varphi}_k^o = \frac{\tilde{\varphi}_k^o}{c_k}, c_k = \|\tilde{\varphi}_k^o\|_2, \tag{58}$$

but there is no reason to assume these form an orthogonal set of eigenfunctions. Then follows the statenent of optimal estimation by these empirical oKEIGS,

$$q = \sum_{l=1}^{k} c_k \overline{\varphi}_k^o + R_{k+1}. \tag{59}$$

Coefficients c_k describe the importance of the dynamic pattern of each of these oKEIGS.

Continuing with the example, see Fig. 8b, showing the successively optimally decreasing residual error. We see a distinct "elbow" which is a commonly used criterion in reduced order modelling. So we suggest

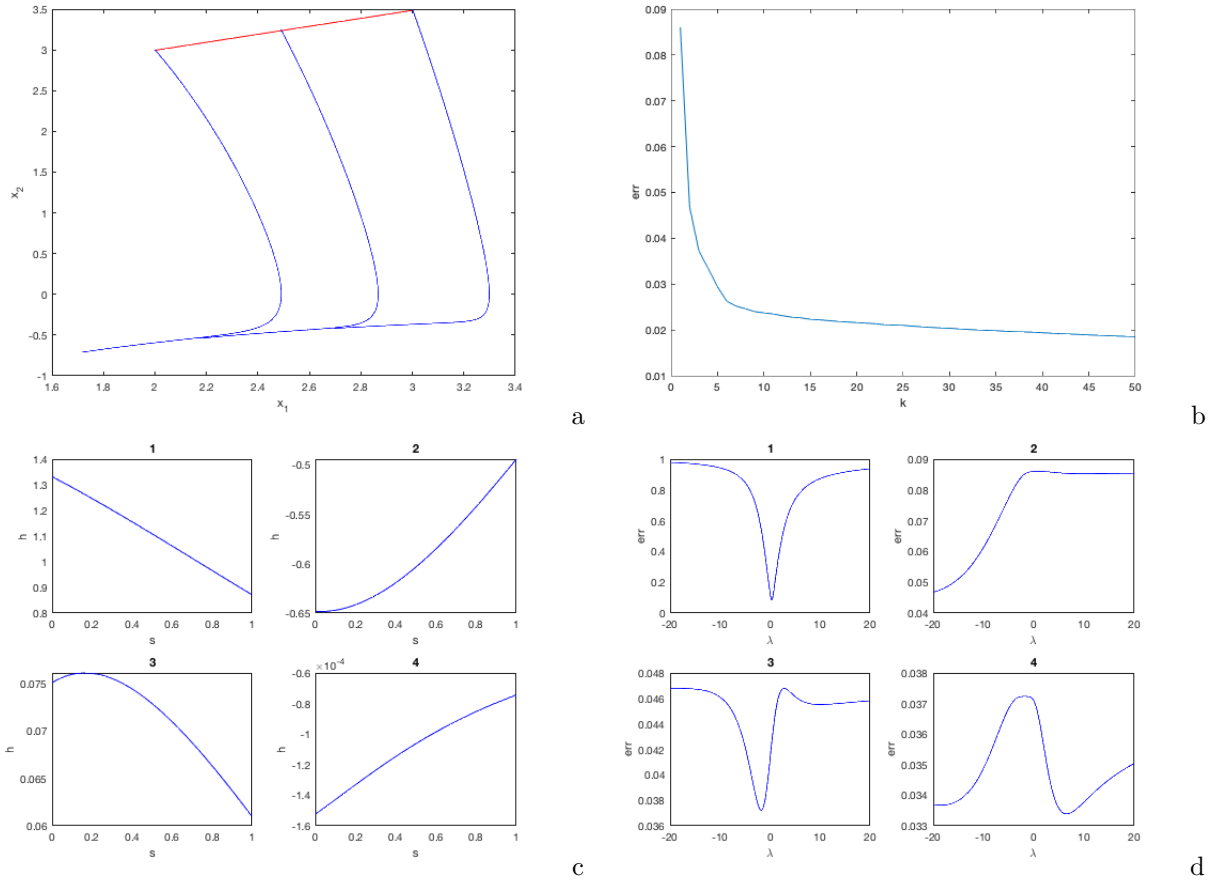

Figure 8: oKEEDMD construction for the Hopf system. (Upper Left) Red data curve Λ evolves by the flow to sweep out domain U shown between the blue curves, (and a blue curve through the middle is shown for reference). (Upper Right) Optimal residual as a function of number of terms fitted, k. (Lower Left) The optimally fitted data functions h fitting the function and its residuals, for the first 4 shown. (Lower Right) The range of errors when fitting the function, and its residuals, as a function of λ, the optimal eigenvalue successively being the minimal value on each curve.

here, that using just 4 terms performs quite well. Figs. 8c-d show the resulting fitted $h_i^o : \Lambda \to \mathbb{R}$ functions which then as above defines the eigenfunctions $\varphi_i^o : U \subset M \to \mathbb{R}$. Finally in Figs. 8d, we show a sweep of real valued eigenvalue candidates for λ_i, $i = 1, 2, 3, 4$. The best fitted eigenvalue is chosen and correspondingly the h_i^0 and φ_i^o. By design, $\|R_k\| \downarrow$ faster for these KEIGS than when using any of the other of the infinitely many available eigenfunctions, and the hope is that $\|R_k\| \to 0$ as $k \to \infty$.

6 Conclusion

Here, we have described how the Koopman operator of a flow in an nonrecurrent domain leads to infinitely many eigenfunctions for each eigenvalue and that all eigenvalues are legitimate. This means that the point spectrum is the full complex plane, but interestingly the spectral decomposition into point and complex spectrum in fact rely on the phase space domain. Furthermore, there at least an uncountable multiplicity of eigenfunctions for each eigenvalue. This all follows the idea that arbitrary functions can be formed on a transverse to the flow data set (curve), and the method of characteristics forms the eigenfunction in the rest of the domain. Even with a simple equivalence class notion that we define here to cope with the algebraic structure of eigenfunctions, whereby we define equivalence of those functions with matched sets of level sets, stating the quotient as a "primary eigenfunction", still there are uncountably many primary eigenfunctions, thus even despite the algebraic structure of eigenfunctions. Then we showed that there are simple geometric considerations to significantly differentiate primary eigenfunctions, when the goal is representation of general functions in terms of linear combinations of primary functions. Thus follows a new kind of empirical scheme to produce efficient functions, that we call a good dictionary. In the spirit of naming algorithms for empirical spectral analysis of the Koopman operator as DMD-like methods, here we call ours oKEEDMD, for "optimal KEEDMD" where EDMD means empirical dynamic mode decomposition for the idea of using basis functions and KE is the new prefix from [8] for Koopman eigenfunction, meaning those basis functions are chosen to be themselves eigenfunctions.

Our construction of a good dictionary is unique from all other constructions since respecting that all eigenfunctions of a nonrecurrent domain are definable entirely on a co-dimension one subdomain that is transverse to the flow, then it is a matter of picking a good data function. So with this in mind we assert a notion of efficient representation that relies just on that basis function but whose argument is only the data function defined only on the co-dimension one set. Thus follows oKEEDMD. We hope that this idea of building eigenbasis functions for reduced order modelling may well find applications beyond the Koopman operator setting where this work was formulated.

7 Acknowledgements

The author received funding from the Army Research Office (N68164-EG) and also DARPA.

8 Appendix: The Method of Characteristics for the Koopman Infinitesimal Generator PDE

Here we recall the classic method of characteristics for quasi-linear PDEs and and this is important to us here since this includes the Koopman PDE, Eq. (11) as a member of this class. For instructive practice, will also specialize the approach to solve the Eq. (11) in an especially simple scenario of a linear ODE. We closely follow the description and notation from the well-regarded textbook by F. John, [12].

A general first order PDE of $u(x, y, ..., z)$ may have a general form, $f(x, y, ..., u, u_x, u_y, .., u_z) = 0$, but if f can be re-arranged algebraically so that the PDE can be written as,

$$a(x, y, u)u_x + b(x, y, u)u_y = c(x, y, u), \ (x, y) \in \Omega \tag{60}$$

then it is called a quasi-linear PDE, and it is amenable to the following solution method. Our presentation is specialized to two spatial variables, but the concept is more general and similar. Allowing a solution surface, $z = u(x, y)$ in an extended (x, y, z) space, then the functions a, b, c may be thought of as a vector field. For brevity, we are skipping the extended discussion of the function domains of enough regularity to carry forward the following. Thus there corresponds a function U, such that $z = u(x, y) \iff U(x, y, z) = 0$, and $\nabla U =< u_x, u_y, -1 >$ is normal to the surface, and the PDE Eq. (60), can be written $\nabla U \cdot < a, b, -c >=< u_x, u_y, -1 > \cdot < a, b, -c >= 0$. Therefore since perpendicular's to the surface (∇U) are also perpendicular to the vector field $< a, b, c >$, then $< a, b, c >$ is a vector field that is tangent to the *characteristic curves.* If we parametrize this surface by r, then we write an ODE for the curves,

$$\begin{aligned} \frac{dx}{dr} &= a(x, y, z), \\ \frac{dy}{dr} &= b(x, y, z), \\ \frac{dz}{dr} &= c(x, y, z). \end{aligned} \tag{61}$$

We assume enough regularity for existence and uniqueness, such as the widely used assumption $< a, b, c > \in Lip(M)$, [29].

Taking Eq. (60) as a Cauchy problem stated with an initial profile on a curve,

$$\Lambda = \{(x, y, z) = (f_1(s), f_2(s), h(s))\}, \tag{62}$$

this reduces a solution to the PDE Eq. (60) as the data function, h on Λ, as,

$$h(s) = u(f_1(s), f_2(s)). \tag{63}$$

This curve Λ should be transverse to the characteristic curves, stated,

$$\frac{d}{ds}(f_1(s), f_2(s), h(s)) \cdot < a(f_1(s), f_2(s), h(s)), b(f_1(s), f_2(s), h(s)), c(f_1(s), f_2(s), h(s)) > \neq 0, \text{ for all } s. \tag{64}$$

Correspondingly, the ODE statement on characteristics, Eq. (61) is stated as an initial value problem for each s.

Remark 9 Example: The Observer Eigenfunction of a Linear System. *To demonstrate the method of characteristics for Koopman eigenfunction analysis, first we recall a now well known scenario, which is a linear ode, that produces the so-called "observer" function as an eigenfunction. Consider again the Koopman PDE, Eq. (11), here specifically for a linear flow. For simplicity of presentation, choose $d = 2$ dimensional domain M, and choose the underlying ODE Eq. (1) to be,*

$$\begin{aligned} \dot{x_1} &= a_1 x_1 \\ \dot{x_2} &= a_2 x_2. \end{aligned} \tag{65}$$

Note that otherwise, the subsequent analysis is largely the same for higher dimensional linear problems, $d > 2$. Then the Koopman PDE is,

$$a_1 \phi_{\lambda, x_1} + a_2 \phi_{\lambda, x_2} = \lambda \phi_\lambda \tag{66}$$

for $\phi_\lambda(x_1, x_2)$ and we have overloaded the notation, that $\phi_{\lambda, x_i} \equiv \frac{\partial \phi_\lambda}{\partial x_i}$. For specificity we assume that the transverse data curve Λ has a specific simple form,

$$\Lambda = \{(x, y, z) = (s, 1, h(s))\}, \tag{67}$$

which is a horizontal line.

Then the characteristic curves, Eq. (61) follow,

$$\begin{aligned} \frac{dx}{dr} &= a_1 x, \\ \frac{dy}{dr} &= a_2 y, \\ \frac{dz}{dr} &= \lambda z. \end{aligned} \tag{68}$$

As an initial value problem on Λ with data function $h(s)$, we have a solution,

$$x = se^{a_1 r}, y = e^{a_2 r}, z = e^{\lambda r} h(s), \tag{69}$$

by multiplying factors. Eliminating the variable r (parameterization along a characteristic curve, starting at any given parameterization point s on Λ), $\ln y = a_2 r \Rightarrow r = \frac{\ln y}{a_2} \Rightarrow e^{a_1 r} = e^{\frac{a_1}{a_2} \ln y} = y^{\frac{a_1}{a_2}}$, and substituting the original variables, $(x, y, z) = (x_1, x_2, \phi_\lambda(x_1, x_2))$ yields,

$$\phi_\lambda(x_1, x_2) = x_2^{\frac{\lambda}{a_2}} h\left(\frac{x_1}{x_2^{(\frac{a_1}{a_2})}} \right). \tag{70}$$

*A special simple case follows the choices of 1) initial data is a constant, $h(s) = 1$, on 2) the initial data curve, Λ chosen specifically as a "horizontal line", Eq. (67). Then we get what has been called the "***observer function***," when $\lambda_2 = a_2$ and thus KEIGS,*

$$(\lambda_2, \phi_{\lambda_2}(x_1, x_2)) = (a_2, x_2). \tag{71}$$

This is easily confirmed to be a solution by substitution into the Cauchy problem, Eqs. (66)-(67). An important point is that to get the other "typically" stated eigenfunction, the observer function in the other variable, we must choose a different data curve, and this time not only must it be transverse to the flow, but it must also be transverse to the data curve, Eq. (67). Similarly, this time choosing specifically a vertical line,

$$\Lambda = \{(x, y, z) = (1, s, h(s))\}, \tag{72}$$

yields a general solution for general initial function h,

$$\phi_\lambda(x_1, x_2) = x_1^{\frac{\lambda}{a_1}} h\left(\frac{x_2}{x_1^{(\frac{a_2}{a_1})}} \right). \tag{73}$$

Likewise this yields the observer function of a constant function $h(s) = 1$, and $\lambda_1 = a_1$, on the this time vertical data curve,

$$(\lambda_1, \phi_{\lambda_1}(x_1, x_2)) = (a_1, x_1). \tag{74}$$

If the system is not already diagonal, standard linear theory to diagonalize the system by a similarity transformation of eigenvectors reveals the straightforward and similar results, remembering that conjugate systems share eigenvalues and corresponding eigenfunctions through the homeomorphism as change of variables, [27, 4].

Remark 10 Example: The Observer Eigenfunction of a One-Dimensional Linear Equation. *An even simpler case follows the $d = 1$-dimensional ode,*

$$\dot{x} = ax, \tag{75}$$

where from the Koopman-"pde" is actually the following ode,

$$ax \frac{d\phi_\lambda}{dx}(x) = \lambda \phi_\lambda(x) = \lambda \phi_\lambda(x). \tag{76}$$

In 1-dimension we do not even need to resort to the method of characteristics. The solution follows by the standard method of ODEs called multiplying factors, using initial data ϕ_0,

$$\phi_\lambda(x) = \phi_0 x^{\frac{\lambda}{a}}. \tag{77}$$

Notice there has not as much freedom in one-dimension, other than algebraic powers, there is only a constant distinguishing eigenfunctions, but not a complete function of arbitrary data, and eigenfunctions are usually considered as an equivalence class when they simply differ by a constant factor.

Remark 11 *Notice that not only is there a solution eigenfunction Eq. (70), one for each* $\lambda \in \mathbb{C}$, *but for each* λ, *there is also the freedom to choose the initial data function* $h : \Lambda \to \mathbb{C}$ *arbitrarily. This is the source of a great deal of nonuniqueness that is only in part equivalent to the usual discussion of uniqueness of the Koopman spectrum associated with the algebraic property of eigenfunctions.*

Proof 3 *of Theorem 1 Here we state the problem for a d-dimensional autonomous ODE. Nonautonomous systems in d-dimensions can be similarly handled, first by the standard methodology of augmenting with a new variable to represent time, to* $d+1$*-dimensional autonomous system. Restating that the initial value problem,* $\dot{x} = F(x)$, $x(t_0) = x_0$ *gives a unique flow* $x(t) = \varrho_t(x_0)$ *solution for* $x_0 \in U$, *and* $x \in U \in M$ *is an open set containing the initial value curve,* $\Gamma \subset U$. *Then in* U, $\nabla\phi_\lambda \cdot F = \lambda\phi_\lambda$, *has characteristic curves which are solutions of the ODE,*

$$\begin{aligned} \frac{dx}{dr} &= F(x) \\ \frac{dz}{dr} &= \lambda z, \end{aligned} \tag{78}$$

following closely the formulation of the method of characteristics as above. So then,

$$\begin{aligned} x(r) &= \varrho_r(x_0(s)) \\ z(r) &= e^{\lambda r} h(s). \end{aligned} \tag{79}$$

Here we remind that for initial conditions $x_0 = x_0(s)$ *on the data surface* $\Lambda \in U$, *and since we write* s *as the parameterization on* Λ, *then the initial value is,*

$$\phi_\lambda(x_0) = \phi_\lambda(x_0(s)) = h(s). \tag{80}$$

Therefore, for an arbitrary $x \in U$, *(and* x *is not necessarily on the initial data curve* Λ*), by assumed existence and uniqueness of the flow,* $\varrho_t : U \times \mathbb{R} \to U$, *there exists an* $r^*(x) \in \mathbb{R}$ *such that,*

$$\varrho_{-r^*(x)}(x) \in \Lambda, \tag{81}$$

uniquely by assumption of no recurrence and existence by assumption that Λ *partitions* U, *transversally to the flow, outcome of the assumption,* $U = \cup_{t\in[t_1,t_2]}\varrho_t(\Lambda)$.

Further, for since each $x \in \Lambda$ *has a parametrization,* $s(x)$ *then let,*

$$s^*(x) = s \circ \varrho_{-r^*(x)}(x). \tag{82}$$

Hence,

$$\phi_\lambda(x) = h \circ s^*(x) e^{\lambda r^*(x)}. \tag{83}$$

That is, from $x \in U$, *"pull"-back along the flow from* x *to* $s^*(x) \in \Lambda$, *where initial data* h *"is read", and the value of* $\phi_\lambda(x)$ *at* x *is therefore* $h \circ s^*(x)$, *but scaled linearly by the time of the pullback, forward from* $s^*(x) \in \Lambda$ *back to the original point* x*:* $e^{\lambda r^*(x)}$. *See Fig. 2.*

It is straightforward check to confirm that the example, Eq. (70), that we derived by direct computation specialized to the linear problem above, can also be derived by application of the general formula in Theorem 2, Eqs. (15), (83).

9 Appendix: Example, On Point Spectrum and Continuous Spectrum

Much has been said lately about the importance of including or understanding the continuous spectrum of Koopman operators with many interesting directions including for computation and in terms of the associated measures, [27, 18], including for toral automorphisms, [10]. Yet as far as we know, there is not yet been a study regarding the Koopman operator spectrum, at least in the modern data driven oriented literature, that appeals directly to the classical definition of the spectral theory of operators regarding the decomposition into point, continuous, and residual spectrum, as we reviewed Definition 3 above. In so doing the conclusion of this section is to contrast the significantly different results that follow different chosen domains.

Here we will discuss aspects of the spectrum,

$$\sigma(\mathcal{L}) = P_\sigma(\mathcal{L}) \cup C_\sigma(\mathcal{L}) \cup R_\sigma(\mathcal{L}), \tag{84}$$

of the Koopman infinitesimal operator $\mathcal{L}$, Eq. (10), as well as of the Koopman operator K_t, Eq. (5), Definition 1, and we will do so directly appealing to Definition 3. As it turns out, the specific choice of definition of the domain $U \in M$, of the measurable functions $g : U \to \mathbb{C}$ involved is crucial to distinguish the nature of the spectral decomposition.

We will appeal to an interesting example which was highlighted recently in the work of Mezic, [27], which describes a perfect action-angle system and from this example we will draw our contrasting observations. First we review: let $(I, \theta) \in M = \mathcal{I} \times S^1$, where $\mathcal{I}$ is an annulus in the plane, $\mathcal{I} = [a, b] \in \mathbb{R}^+$ is an interval and S^1 is a circle, and so M is a cylinder or alternatively it can be thought of as an annulus as shown in Fig. 9. As we will see, depending on the domain chosen, there are significant difference, and these are crucially rooted in the concept of nonrecurrence that featured centrally in Theorem 2.

Let,

$$\dot{I} = 0, \ \dot{\theta} = I. \tag{85}$$

From an initial condition, (I_0, θ_0) clearly the solution is $(I(t), \theta(t)) = (I_0, \theta_0 + tI_0)$. We will use this same problem to emphasize the role of domain in contrasting the case of when there is a point spectrum, and/or a continuous spectrum, and also how the cardinality of these belies their names.

9.1 Nonrecurrent subdomain

In a nonrecurrent subdomain U of M, Theorem 2 holds without modification, and therefore solutions of the KEIGS are easily determined and of the form, Eq. (15). Such a subdomain may be chosen as any open set that does not circuit a complete circle, e.g. an open "wedge" $U = (a, b) \times (\alpha_1, \alpha_2) \subsetneq \mathcal{I} \times S_1$, $0 < \alpha_1 < \alpha_2 < 2\pi$. Consider for example,

$$\Lambda = \{(s, \alpha) : \alpha_1 < \alpha < \alpha_2 \text{ is a fixed constant }, a < s < b \text{ is a parameter}\}, \tag{86}$$

is a radial ray in $U \subseteq \mathcal{I}$. See Fig. 9. Further assume a data function is given, $h : \Lambda \to \mathbb{C}$. Then by Eq. (15),

$$\phi_{\lambda,h,\Lambda}(I, \theta) = h(I)e^{\frac{\lambda(\theta-\alpha)}{I}}. \tag{87}$$

Furthermore, by arguments also leading to Corollary 2 and also Remark 7 we know that all $\lambda \in \mathbb{C}$ are eigenvalues (and furthermore all h allow eigenfunctions). Therefore, in terms of spectral decomposition, we conclude that the point spectrum is the full complex plane,

$$P_\sigma(\mathcal{L}) = \mathbb{C}, \text{ but therefore by compliment, the continuous and residual spectra, } C_\sigma(\mathcal{L}) = R_\sigma(\mathcal{L}) = \emptyset, \tag{88}$$

completes the spectrum, $\sigma(\mathcal{L})$ as per Eq. (84).

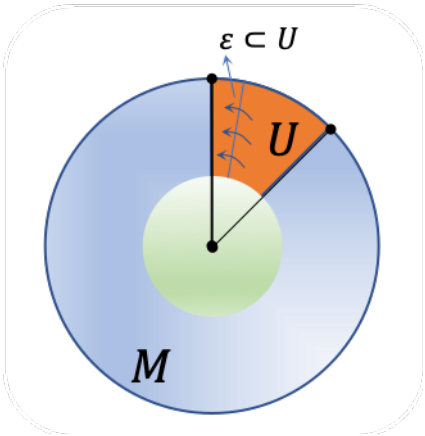

Figure 9: Domain in the annulus $(I\theta) \in M = \mathcal{I} \times S^1$ be chosen as the entire annulus but then the nonrecurrence condition is violated and so Theorem 2 does not apply, but however if a wedge $U = (a, b) \times (\alpha_1, \alpha_2) \subsetneq \mathcal{I} \times S_1$, $0 < \alpha_1 < \alpha_2 < 2\pi$ is chosen, then the theorem does hold, and therefore a different form of the eigenfunctions hold with significant consequences to the nature of the spectral decomposition.

9.2 Recurrent subdomain

Now consider the same dynamical system, the action angle flow Eq. (85), we reconsider the problem, but without the restriction on nonrecurrence. So now we take $U = M = \mathcal{I} \times S^1$ as the full annulus, as did Mezic in [27], so first we review. With nonrecurrence, Theorem 2 does not hold, and therefore we may not resort to the eigenfunction solution Eq. (15) that follows from method of characteristics; that proof breaks down largely because of loss of the concept of a unique solution must be single valued. However, Mezic showed directly that since the Koopman operator $K_t[g](I, \theta) = g(I, \theta + It)$, he proved that eigenfunctions do not exist, stated using the phrase "the eigenfunction in a proper sense do not exist." The argument was, depending directly on the idea of a composition operator rather than the infinitesimal generator, that an eigenfunction would need to satisfy $\phi_\omega(I, \theta) = \phi_\omega(I + \theta + It) = e^{i\omega t}\phi_\omega(I, \theta)$, but allowing constant radius $r = I$ for this system and writing, $\phi_\omega = r^{i\varphi(I,\theta)}$ then also $\phi_\omega(I+\theta+It) = re^{i\varphi(I,\theta)+iIt} = re^{i\varphi(I,\theta)+i\omega t}$, which can only hold if $\omega = I$. The conclusion is that this not possible in the sense of a function, so there is no eigenfunction. But the author goes on to note that,

$$\phi(I, \theta) = e^{i\theta}\delta(I - c), \tag{89}$$

where $\delta(x)$ is the dirac delta-"function", acts like an "eigenfunction" but that is really only true in the distributional sense and not literally as a function. Therefore he notes that this as an "eigenmeasure," but it is not a function. That is, weakly one demands that integrating against test functions, $f(I, \theta)$, the composition operator acts as, $\int_M K_t\phi(I, \theta)f(I, \theta)dId\theta = \int_M \phi(I, \theta+It)f(I, \theta+It)dId\theta = \int_M e^{i(\theta+It)}\delta(I-c)f(I, \theta)dId\theta = e^{ict}\int_M \phi(I, \theta)f(I, \theta)dId\theta$. In fact, more generally it can be shown that an eigenmeasure can be designed by a delta function supported over any solution, as noted in other contexts in terms of the adjoint operator (Frobenius-Perron), [5, 2, 3], but in any case these are not functions.

Now re-examining this same issue leading to the eigenfunction from the first principle of referring to the definitions of spectral decomposition, Definition 3 and Eq. (84), we demand eigen-like objects that are only allowable actual functions. This is leads to the idea of "approximate eigenfunctions" already included as part of spectral theory of operators in the field of functional analysis:

Definition 5 Approximate Eigenvalue. *Under the same assumptions as Definition 3, $b = e^\lambda$ is an approximate eigenvalue of a linear operator K_t if there is no $f \neq 0$ such that $K_t f = b^t v$ (thus b is not an element of the point spectrum) but it satisfies that for any $\epsilon > 0$, there exists a function $f_\epsilon : U \to \mathbb{C}$, $f_\epsilon \in \mathcal{F}$,*

such that

$$\|K_t f_\epsilon - b^t f_\epsilon\| < \epsilon, \tag{90}$$

Furthermore the point is that b must be an element of either the continuous spectrum $C_\sigma(K_t)$ or residual spectrum $R_\sigma(K_t)$, but not the point spectrum. Therefore to show a nonempty $C_\sigma(K_t) \cup R_\sigma(K_t)$, we need only show existence of an approximate eigenfunction. In brief, the reason this is sufficient is that an unbounded linear operator is therefore not a continuous operator [1, 36].

To this end, recall that the action of the "delta function", while not literally a function but rather a generalized function, it is nonetheless described weakly as the limit of the action of a sequence of real functions. One such commonly used sequence of functions are the following specific set of indicator functions, $\delta_n : \mathbb{R} \to \mathbb{R}$, and notice the subindex $n \in \mathbb{Z}^+$,

$$\delta_n(x) = \begin{pmatrix} n & \frac{-1}{2n} < x < \frac{1}{2n} \\ 0 & \text{else} \end{pmatrix} \tag{91}$$

Likewise, a sequence indicator functions can be defined on a complex domain, or otherwise, many other smooth functions are also often used, notably radial basis functions, to build the action of the delta function as a sequence of actual functions. So with δ_n and mirroring Eq. (89), consider a (candidate) approximate eigenfunction,

$$\phi_{\omega,n}(I, \theta) = e^{i\theta} \delta_n(I - \omega). \tag{92}$$

Then,

$$\begin{aligned}
\|K_t \phi_{\omega,n}(I,\theta) - e^{i\omega t}\phi_{\omega,n}(I,\theta)\|_{L^2(M)} &= \|\phi_{\omega,n}(I,\theta + It) - e^{i\omega t}\phi_{\omega,n}(I,\theta)\|_{L^2(M)} \\
&= \|\phi_{\omega,n}(I,\theta + It) - e^{i\omega t}\phi_{\omega,n}(I,\theta)\|_{L^2(M)} \\
&= \|e^{in\theta}(e^{iIt} - e^{i\omega t})\delta_n(I-\omega)\|_{L^2(M)} \\
&= \|e^{in\theta}(e^{iIt} - e^{i\omega t})\mathbf{1}_{(\frac{-1}{2n},\frac{1}{2n})}(I-\omega)\delta_n(I-\omega)\|_{L^2(M)} \\
&\leq \|e^{in\theta}\|_{L^2(M)}\|(e^{iIt} - e^{i\omega t})\mathbf{1}_{(\frac{-1}{2n},\frac{1}{2n})}(I-\omega)\|_{L^2(M)}\|\delta_n(I-\omega)\|_{L^2(M)} \\
&= \|e^{in\theta}\|_{L^2(M)}\|(e^{iIt} - e^{i\omega t})\|_{L^2([\frac{-1}{2n},\frac{1}{2n}]\times S^1)}\|\delta_n(I-\omega)\|_{L^2(M)} \\
&= \|(e^{iIt} - e^{i\omega t})\|_{L^2([\frac{-1}{2n},\frac{1}{2n}]\times S^1)} \\
&\approx \|\mathcal{O}(I-\omega)\|_{L^2([\frac{-1}{2n},\frac{1}{2n}]\times S^1)} \\
&= \epsilon(n)
\end{aligned} \tag{93}$$

The third line follows substitution of the assumed form, Eq. (92). The step to the fourth line of the string of equalities takes note that the chosen $\phi_{\omega,n}$ includes an indicator function, as δ_n in Eq. (91) is zero outside the domain $(\frac{-1}{2n}, \frac{1}{2n})$. The seventh line simply notes that, $\|e^{in\theta}\|_{L^2(M)} = \|\delta_n(I-\omega)\|_{L^2(M)} = 1$. The second to last line follows a Taylor series of the exponential, and then finally the last line follows by defining, $\epsilon(n) \sim \frac{c}{n}$, for a constant c. Thus we conclude that δ_n are approximate eigenfunctions, and since approximate eigenfunctions exist, then $C_\sigma(K_t) \cup R_\sigma(K_t)$ is nonempty. Furthermore this construction is general and leads to $C_\sigma(K_t) \cup R_\sigma(K_t) = \mathbb{C}$, and from the material reviewed above $P_\sigma(K_t) = \emptyset$.

We summarize: Contrasting the result in this subsection to the previous subsection, the same system the Eq. (85) in action-angle form, leads to essentially opposite outcomes depending on the domain chosen. Either a simple system of all point spectrum and empty continuous union residual spectrum if a nonrecurrent subdomain is chosen, or otherwise an empty point spectrum but full continuous union residual spectrum. For other reasons not discussed here we expect the continuous spectrum alone is full and this dovetails with much of the discussion regarding integration against an absolutely continuous measure in [10, 18].

10 Appendix: Proof

Here we state a proof of Proposition 1, a result which is well known, but we include this proof in part because we have not seen this proof and we find it interesting.

Proof 4 *Let (λ_1, ϕ_1) and (λ_2, ϕ_2) each be KEIGS. By Eq. (6), let $b_1 = e^{\lambda_1}$, and $b_2 = e^{\lambda_2}$. Each of these satisfies Eq. (11), $F \cdot \nabla \phi_{\lambda_1} = \lambda_1 \phi_{\lambda_1}$, and $F \cdot \nabla \phi_{\lambda_2} = \lambda_2 \phi_{\lambda_2}$. Now we substitute $(\phi_{\lambda_1}(x))^{\alpha_1}(\phi_{\lambda_2}(x))^{\alpha_2})$ into the Koopman PDE,*

$$\begin{aligned} F \cdot \nabla((\phi_{\lambda_1}(x))^{\alpha_1}(\phi_{\lambda_2}(x))^{\alpha_2})) &= F \cdot (\alpha_1(\phi_{\lambda_1})^{\alpha_1 - 1}\nabla\phi_{\lambda_1}\phi_{\lambda_2}^{\alpha_2} + \alpha_2(\phi_{\lambda_2})^{\alpha_2-1}\nabla\phi_{\lambda_2}\phi_{\lambda_1}^{\alpha_1}) \\ &= F \cdot (\alpha_1(\phi_{\lambda_1})^{\alpha_1 - 1}\phi_{\lambda_2}^{\alpha_2}\nabla\phi_{\lambda_1} + \alpha_2(\phi_{\lambda_2})^{\alpha_2-1}\phi_{\lambda_1}^{\alpha_1}\nabla\phi_{\lambda_2}) \\ &= \alpha_1\phi_{\lambda_1}^{\alpha_1 - 1}\phi_{\lambda_2}^{\alpha_2}(F \cdot \nabla\phi_{\lambda_1}) + \alpha_2\phi_{\lambda_2}^{\alpha_2-1}\phi_{\lambda_1}^{\alpha_1}(F \cdot \nabla\phi_{\lambda_2}) \\ &= \alpha_1\phi_{\lambda_1}^{\alpha_1 - 1}\phi_{\lambda_2}^{\alpha_2}(\lambda_1\phi_{\lambda_1}) + \alpha_2\phi_{\lambda_2}^{\alpha_2-1}\phi_{\lambda_1}^{\alpha_1}(\lambda_2\phi_{\lambda_2}) \\ &= (\alpha_1\lambda_1 + \alpha_2\lambda_2)\phi_{\lambda_1}^{\alpha_1}\phi_{\lambda_2}^{\alpha_2}. \end{aligned} \tag{94}$$

Note also that,

$$e^{\alpha_1\lambda_1 + \alpha_2\lambda_2} = e^{\alpha_1\lambda_1}e^{\alpha_2\lambda_2} = (e^{\lambda_1})_1^{\alpha}(e^{\lambda_2})^{\alpha_2}) = (b_1)^{\alpha_1}(b_2)^{\alpha_2}. \tag{95}$$

This relates the eigenvalues λ_1, λ_2 to the multiplying factors b_1, b_2.

References

[1] Nino Boccara. *Functional Analysis: An Introduction for Physicists*. Elsevier, 1990.

[2] Erik M Bollt. Controlling chaos and the inverse frobenius–perron problem: global stabilization of arbitrary invariant measures. *International Journal of Bifurcation and Chaos*, 10(05):1033–1050, 2000.

[3] Erik M Bollt, Lora Billings, and Ira B Schwartz. A manifold independent approach to understanding transport in stochastic dynamical systems. *Physica D: Nonlinear Phenomena*, 173(3-4):153–177, 2002.

[4] Erik M Bollt, Qianxiao Li, Felix Dietrich, and Ioannis Kevrekidis. On matching, and even rectifying, dynamical systems through koopman operator eigenfunctions. *SIAM Journal on Applied Dynamical Systems*, 17(2):1925–1960, 2018.

[5] Erik M Bollt and Naratip Santitissadeekorn. *Applied and Computational Measurable Dynamics*. SIAM, 2013.

[6] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.

[7] Predrag Cvitanovic, Roberto Artuso, Ronnie Mainieri, Gregor Tanner, Gábor Vattay, Niall Whelan, and Andreas Wirzba. Chaos: classical and quantum. *ChaosBook. org (Niels Bohr Institute, Copenhagen 2005)*, 69, 2005.

[8] Carl Folkestad, Daniel Pastor, Igor Mezic, Ryan Mohr, Maria Fonoberova, and Joel Burdick. Extended dynamic mode decomposition with learned koopman eigenfunctions for prediction and control. *arXiv preprint arXiv:1911.08751*, 2019.

[9] Pierre Gaspard. *Chaos, scattering and statistical mechanics*, volume 9. Cambridge University Press, 2005.

[10] Nithin Govindarajan, Ryan Mohr, Shivkumar Chandrasekaran, and Igor Mezic. On the approximation of koopman spectra for measure preserving transformations. *SIAM Journal on Applied Dynamical Systems*, 18(3):1454–1497, 2019.

[11] Philip Holmes, John L Lumley, Gahl Berkooz, and Clarence W Rowley. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press, 2012.

[12] Fritz John. Partial differential equations. 1975.

[13] R Kadison and John Ringrose. Fundamentals of the operator algebras. *vols. I and II, Academic Press, Orlando*, 1983.

[14] Kari Karhunen. Under lineare methoden in der wahr scheinlichkeitsrechnung. *Annales Academiae Scientiarun Fennicae Series A1: Mathematia Physica*, 47, 1947.

[15] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *arXiv*, 2017.

[16] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315, 1931.

[17] Milan Korda and Igor Mezić. Optimal construction of koopman eigenfunctions for prediction and control. 2019.

[18] Milan Korda, Mihai Putinar, and Igor Mezić. Data-driven spectral analysis of the koopman operator. *Applied and Computational Harmonic Analysis*, 2018.

[19] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[20] Yueheng Lan and Igor Mezić. Linearization in the large of nonlinear systems and koopman operator spectrum. *Physica D: Nonlinear Phenomena*, 242(1):42–53, 2013.

[21] Andrzej Lasota and Michael C Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics*, volume 97. Springer Science & Business Media, 2013.

[22] Qianxiao Li, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

[23] Michel Loeve. Elementary probability theory. In *Probability Theory I*, pages 1–52. Springer, 1977.

[24] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[25] Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[26] Igor Mezic. Koopman operator spectrum and data analysis. *arXiv preprint arXiv:1702.07597*, 2017.

[27] Igor Mezic. Spectrum of the koopman operator, spectral expansions in functional spaces, and state space geometry. 2019.

[28] Ryan Mohr and Igor Mezić. Koopman principle eigenfunctions and linearization of diffeomorphisms. *arXiv preprint arXiv:1611.01209*, 2016.

[29] Lawrence Perko. *Differential equations and dynamical systems*, volume 7. Springer Science & Business Media, 2013.

[30] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[31] Satosi Watanabe. Knowing and guessing a quantitative study of inference and information. 1969.

[32] George Andrew Webber, RA Handler, and L Sirovich. The karhunen–loeve decomposition of minimal channel flow. *Physics of Fluids*, 9(4):1054–1066, 1997.

[33] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *J. Nonlinear Sci.*, 25(6):1307–1346, 6 2015.

[34] Matthew O. Williams, Clarence W. Rowley, and Ioannis G. Kevrekidis. A Kernel-Based Approach to Data-Driven Koopman Spectral Analysis. *arXiv*, v4(1411.2260), 2014.

[35] Matthew O. Williams, Clarence W. Rowley, Igor Mezić, and Ioannis G. Kevrekidis. Data fusion via intrinsic dynamic variables: An application of data-driven Koopman spectral analysis. *EPL (Europhysics Letters)*, 109(4), 2 2015.

[36] Kosaku Yosida. *Functional analysis*. springer, 1968.