An Investigation into Koopman Operator Theory

Linear Subspace Methods for Koopman Mode Decomposition, and Parameter Estimation for Nonlinear Discrete Time Dynamical Systems using the Koopman Operator Theory

Pier Tabak, 6654312
Mathematics, Utrecht University

June 16, 2021

Abstract

This thesis is about Koopman operator theory. The first, main, chapter of this thesis concerns the connection between classical linear subspace methods and Koopman operator theory. We show that these are very much related. The Univariate Principal Components (UPC) algorithm can be used to construct approximations of Koopman eigenfunctions, lying within the span of delay coordinates. In general these are poor eigenfunction approximation, but provide a good Koopman mode decomposition, when compared to Extended Dynamic Mode Decomposition (EDMD) on several low-dimensional examples.

The short second chapter concerns an investigation into the usage and limitations of parameter estimation of discrete-time dynamical systems from partial measurements using Koopman operator theory. We argue that including delay coordinates allows for the estimation of all parameters, but that doing so gives rise to a computational issue in computing Galerkin inner products. We propose to overcome this by continuous approximation of the empirical distribution. We are not able to substantiate these ideas with numerical results. We find that the computational costs are too high, and propose to tackle this by resorting to polynomial approximation, for which integration and optimization are easier.

Contents

Preface

My study in mathematics started with a desire to study complex systems, in particular the economy. For my thesis I finally saw the opportunity to delve into the topic. Originally, I formulated my research interested as follows: *system identification from incomplete measurements for complex adaptive systems with forcing, with application to the economy.* Koopman operator theory had caught my interest somewhere earlier during the master, and has also lead me to ask professor Jason Frank to be my supervisor.

At the beginning of my thesis period I started looking into Koopman operator theory and its connections to parameter estimation. This drove me into various directions. I encountered an article [4] which hinted at connecting linear subspace identification methods with Koopman identification methods. Attempts to connect these did not bring me much. However, I did notice a strong connection between subspace identification methods and Koopman operator theory, but without an application to parameter estimation.

Finally, I distinguished two seperate research interests. Firstly my original interest in parameter estimation for economics using Koopman operator theory, and secondly the connection between linear subspace methods and Koopman operator theory. This culminated in a split thesis, with a short chapter on parameter estimation for discrete-time dynamical systems from partial measurements, and a large chapter on the latter topic.

The outcomes of my research into Koopman identification are limited. I argue that including delay coordinates as observables in the Koopman operator approximation allows for the estimation of all parameters. However, it brings with it a practical challenge in the computation of inner products. I suggest that this challenge might be met by resorting to polynomial approximation. Unfortunately, I am not able to provide numerical results.

The reseach into the connection between linear subspace methods and Koopman operator theory has delivered. Both theoretically and numerically I show that linear subspace methods can be used to obtain an approximate Koopman mode decomposition. I call the algorithm *Subspace Koopman Algorithm*. In comparison to EDMD, the mode decomposition as provided by SKA offers a better description as measured in terms of predictable power, as tested in various examples.

During my thesis Jason Frank has supported me by providing on a continuous basis a listening ear, constructive feedback, suggestions and a lot of good questions. I sometimes had a tendency to concentrate my efforts on a certain topic, without first analysing my motives for working on the topic. Jason Frank then asked the questions which made me realize this. I would like to thank Jason Frank for being my supervisor.

1 Introduction

The topic of this thesis is the approximation of the Koopman operator. Koopman operator theory [3] states that autonomous dynamical systems have linear dynamics in the space of observables even when the dynamics in the state space are non-linear. In this work, we consider the discrete case of dynamics given by a map $f : \mathbb{X} \to \mathbb{X}$, on state space $\mathbb{X} \subseteq \mathbb{R}^n$. Observables are functions on the state space $g : \mathbb{X} \to \mathbb{C}$. Here $\mathbb{C}$ is the set of complex numbers. Given a function space $\mathbb{G}$ of observables that is closed under composition with f, the Koopman operator is the linear map $\mathcal{K} : \mathbb{G} \to \mathbb{G}$ defined as $\mathcal{K}(g) := g \circ f$.

Often one is interested in obtaining a triplet of Koopman eigenfunctions ϕ_i, eigenvalues λ_i, and modes v_i. Denote by $\mathrm{id} : \mathbb{X} \to \mathbb{X} \subseteq \mathbb{R}^n$ the vector valued observable, called the identity observable, defined by $\mathrm{id}(x) = x$ for all $x \in \mathbb{X}$. The Koopman modes relate the eigenfunctions to the state space dynamics by $\mathrm{id} = \sum_{i=1}^{\infty} v_i \phi_i$.

One of the useful characteristics of Koopman operator theory is the availability of data-based methods that compute approximations of some of the aforementioned quantities. In [10] (introduction) and [15] an overview of these algorithms is given.

The most frequently used algorithm is Extended Dynamic Mode Decomposition (EDMD) [9, 10]. To use EDMD one needs to specifiy a library of observable functions, $\mathcal{G}_N = \{\psi_i\}_{i=1}^N$. EDMD then computes an approximation of the Koopman operator and related quantities. Per definition the approximate eigenfunctions lie in the space spanned by the library of observables, $\mathbb{G}_N = \mathrm{span}(\mathcal{G}_N)$. Such approximations only make sense when $\mathbb{G}_N$ contains an (almost) Koopman invariant subspace. In most situations one is not only interested in obtaining eigenfunctions and values, but in obtaining the modes too. Because the Koopman modes are the expansion coefficients of expressing the identity observable in terms of the eigenfunctions, getting an approximation of the modes requires that the identity observable lies in or close to the Koopman invariant subspace contained in $\mathbb{G}$. One of the difficulties [10, 11, 16, 17, 18, 19] in using EDMD is in choosing a library of observables which contains a Koopman invariant subspace that includes the identity function.

This thesis is divided in two parts. In the first part we investigate how the UPC algorithm, a linear subspace algorithm, can be used to obtain an approximate Koopman mode decomposition. In the second part we will investigate how Koopman operator theory can be used to simulation-free estimate parameters of models from partial measurements. We conclude this introduction by providing a basic introduction in to Koopman operator theory and EDMD.

Koopman Operator Theory

Throughout this work we will adopt ideas and notation from [9] in discussing Koopman operator theory and EDMD. Let $\mathbb{X} \subseteq \mathbb{R}^n$ be some state space, and f be

some map providing the dynamics:

$$x_{t+1} = f(x_t).$$

We call $g : \mathbb{X} \to \mathbb{C}$ an observable function. Let $\mathbb{G}$ be some function space of observables which is closed under composition with f:

$$g \in \mathbb{G} \Rightarrow g \circ f \in \mathbb{G}.$$

The Koopman operator $\mathcal{K} : \mathbb{G} \to \mathbb{G}$ is defined as

$$\mathcal{K}(g) := g \circ f.$$

A Koopman eigenfunction ϕ, is a function that satisfies $\mathcal{K}(\phi) = \lambda\phi$, for some eigenvalue $\lambda \in \mathbb{C}$. Denote by $\{\phi_i\}_{i=1}^{\infty}$ the Koopman eigenfunctions. The Koopman modes $\{v_i\}_{i=1}^{\infty}$, $v_i \in \mathbb{C}^n$ are defined by $\mathrm{id} = \sum_{i=1}^{\infty} v_i\phi_i$. The modes can be interpreted as the expansion coefficients of expanding id in the eigenfunction basis.

The Koopman operator has gained a lot of attention recently because of the fact that it is a linear operator, irrespective of whether the map f is linear or non-linear. This allows for both an easy approximation of the Koopman operator from data, as well as a Galerkin method to construct a projected Koopman operator given a model, as discussed later. The Koopman operator is an infinite dimensional operator; in order to be able to work with the Koopman operator we look at projections onto a finite dimensional space. Assuming that $\mathbb{G}$ is a Hilbert space with linearly independent basis $\{\psi_i\}_{i=1}^{\infty}$, we denote by $\mathbb{G}_N$ a finite dimensional subspace of $\mathbb{G}$, given by $\mathbb{G}_N := \mathrm{span}(\psi_1, \psi_2, ..., \psi_N)$. Furthermore, assume that $\mathbb{X}$ is measurable, and μ is some measure on X, we define by $\mathcal{P}_N^{\mu} : \mathbb{G} \to \mathbb{G}$ the projection operator:

$$\mathcal{P}_N^{\mu} g = \mathrm{argmin}_{\tilde{g} \in \mathbb{G}_N} \|\tilde{g} - g\|_{\mu}.$$

Where the norm is defined with respect to the inner product $\langle f, g \rangle_{\mu} := \int_{\mathbb{X}} f\bar{g}\,\mathrm{d}\mu$, where $\bar{\cdot}$ denotes the complex conjugate. We introduce the shorthand notation $\mathcal{K}_N : \mathbb{G}_N \to \mathbb{G}_N$ to be the Koopman operator projected onto a finite dimensional subspace by $\mathcal{K}_N := \mathcal{P}_N^{\mu}\mathcal{K}$. Such finite dimensional linear operators can be represented in matrix form by $K_N^{\mu} \in \mathbb{R}^{N \times N}$, as defined by

$$\mathcal{K}_N^{\mu}(g) \equiv a_g K_N^{\mu} \Psi,$$

for $g = a_g\Psi$, where $\Psi : \mathbb{X} \to \mathbb{C}^N$ is the vector function $\Psi = [\psi_1, \psi_2, ..., \psi_N]^{\top}$, $a_g \in \mathbb{C}^N$ is a row vector, and $\top$ denotes the matrix transpose.

Let $\hat{\phi}$ be an eigenfunction of $\mathcal{K}_N$, with eigenvalue $\hat{\lambda}$. Since $\mathcal{K}_N$ is an operator on $\mathbb{G}_N$, we have $\hat{\phi} \in \mathbb{G}_N$, and hence $\hat{\phi} = a_{\hat{\phi}}\Psi$ for some $= a_{\hat{\phi}} \in \mathbb{C}^N$ By definition of an eigenfunction we have

$$\hat{\lambda} a_{\hat{\phi}} \Psi = \hat{\lambda}\hat{\phi} = \mathcal{K}_N(\hat{\phi}) = a_{\hat{\phi}} K_N \Psi.$$

It follows that if $a_{\hat{\phi}}$ is a left eigenvector of K_N, then $\hat{\phi} = a_{\hat{\phi}}\Psi$ is an eigenfunction of $\mathcal{K}_N$.

Let $WK_N = K_N\Lambda$ be an eigenvector decompositon of K_N, with W a matrix of left eigenvectors and Λ a diagonal matrix with eigenvalues. Using the notation $\hat{\Phi} = [\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_N]^\top$, a vector function of Koopman eigenfunctions, we arrive at

$$\hat{\Phi} = W\Psi.$$

Let B be the matrix of expansion coefficients, of expanding id in terms of ψ, that is B is defined by $\mathcal{P}_N^\mu \mathrm{id} = B\Psi$. Finally, let $\hat{v}_i$ be the approximate Koopman modes, such that $\mathcal{P}_N^\mu \mathrm{id} = \sum_{i=1}^N \hat{v}_i\hat{\phi}_i$, or in matrix form $\mathcal{P}_N^\mu \mathrm{id} = V\hat{\Phi}$. By some simple manipulation, the approximate Koopman modes are given by

$$V = BW^{-1}.$$

EDMD

For simplicity, we assume for now that full state measurements are available, so h is the identity function. The EDMD algorithm is a method to approximate K_N^μ from data. For $x_0, x_1, ..., x_M$ and $x_0^\#, x_1^\#, ..., x_M^\#$ data points satisfying $x_t^\# = f(x_t)$, define the matrices

$$\Psi_X = [\Psi(x_0)|\Psi(x_1)|...|\Psi(x_M)] \text{ and } \Psi_{X^\#} = [\Psi(x_0^\#)|\Psi(x_1^\#)|...|\Psi(x_M^\#)].$$

Furthermore, throughout this thesis we denote by ρ the empirical measure:

$$\rho := \frac{1}{M}\sum_{t=0}^{M} \delta_{x_t}, \tag{1}$$

where δ_{x_t} is the Dirac delta function at x_t.

Then, the EDMD approximation $\hat{K}_N^\rho$ is given by

$$\hat{K}_N^\rho = \Psi_{X^\#}\Psi_X^+,$$

where $+$ denotes the Moore-Penrose pseudoinverse. See [10] and [9] for more details. From here on we will regard N to be fixed, and write $\mathcal{K}_\mathrm{D}^\mu$ for the EDMD approximation of the Koopman operator and $K_\mathrm{D}^\mu = \hat{K}_N^\mu$ for its matrix representation, where the subscript D refers to Data. In [11] (see also [9]) it is proven that if one assumes x_t to be drawn independently from distribution μ,[1] then the operator $\mathcal{K}_\mathrm{D}^\mu$ converges to $\mathcal{K}_N^\mu$ as the number of samples M goes to infinity. Furthermore, irrespective of which sampling is used $\mathcal{K}_\mathrm{D}^\rho = \mathcal{K}_N^\rho$.

[1] Or equivalently, the data is a time-series of an ergodic system with equilibrium measure μ.

Part I
Linear Subspace Methods for Koopman Mode Decomposition

2 Introduction

One of the useful characteristics of Koopman operator theory is the availability of data-based methods that compute approximations of some of the aforementioned quantities. In [10] (introduction) and [15] an overview of these algorithms is given. In this thesis we will focus on the problem of obtaining an approximation of the entire triplet of Koopman eigenfunctions ϕ_i, eigenvalues λ_i, and modes v_i, to which we will refer as the *eigenfunction-mode problem*. The most frequently used algorithm is Extended Dynamic Mode Decomposition (EDMD) [9, 10]. One of the difficulties [10, 11, 16, 17, 18, 19] in using EDMD is in choosing a library of observables which contains a Koopman invariant subspace that includes the identity function.

On the one hand, there is some literature [17, 20] that proposes to tackle this challenge by starting with a large library of observables, and then require some kind of sparsity during the identification of the Koopman operator. Alternatively, in [18] an algorithm is presented which iteratively finds a maximal Koopman invariant subspace given a large set of observables. Both approaches still rely on being initialized with a meaningful library of observables.

On the other hand, Ulam's method provides a non-parametric approach, as used in [21]. It does not rely on observable functions, but instead discretizes the statespace into boxes and computes an approximation of f by counting the box-to-box rate. Left eigenvectors of the matrix representation of this discretization of f are approximate Koopman eigenfunctions. The drawback of this method is clear: the discretization introduces an error. Also, obtaining an approximation of the modes is not a trivial exercise.

In this thesis we will show that well established linear subspace methods [22] can be used to compute non-parametrically approximations of Koopman eigenfunction, eigenvalues and modes. This investigation was originally inspired by the hint towards the usefulness of subspace methods in [4].

Linear subspace methods were developed to identify linear state space models from time series data. Given a time series of output measurements $\{x_t\}_{t=0}^{M}$ of a linear dynamical system of the form

$$\begin{aligned} z_{t+1} &= Az_t \\ x_t &= Cz_t, \end{aligned} \tag{2}$$

the Unweighted Principal Component (UPC) algorithm [22, 23] determines first the trajectory $\{z_t\}_{t=0}^{M}$, and then the matrices A and C.

The connections between subspace methods and the Koopman operator and related algorithms have been investigated in some literature before. In [23] a theoretical connection is made between UPC and the dynamic mode decomposition (DMD) algorithm. They show that the subspace identification problem and the DMD problem are equivalent. DMD precedes the development of EDMD, but can be seen as a special case of EDMD, where the dictionary of observables is just the identity observable. Furthermore, in [24] a modification to DMD is made by including delay coordinates and computing the projection of future measurements on past measurements, as is also done in traditional subspace algorithms. The latter method is developed to handle noise, and is proven to converge to the Koopman operator in case the underlying observables span a Koopman invariant subspace. Other work on the relation between subspace methods and the Koopman operator are not known to us.

We note that the quest for Koopman modes and eigenfunctions is the quest for a linear state space model. We are looking for a vector of approximate eigenfunctions $\hat{\Phi} : \mathbb{X} \to \mathbb{C}^N$, $\hat{\Phi} = [\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_N]$, such that $\hat{\Phi}$ satisfies the eigen problem

$$\mathcal{K}(\hat{\Phi}) = \hat{\Phi} \circ f \approx \Lambda \hat{\Phi},$$

where Λ is a diagonal matrix with eigenvalues. Simultaneously, we desire for the Koopman modes v_i:

$$\mathrm{id} \approx \sum_{i=1}^{N} v_i \hat{\phi}_i = V \hat{\Phi}.$$

This implies that we are looking for a linear system approximation of f:

$$\begin{aligned} z_t &= \hat{\Phi}(x_t) \\ z_{t+1} &\approx \Lambda z_t \\ x_t &\approx V z_t. \end{aligned} \tag{3}$$

Based on this representation of the Koopman eigenfunction-mode problem, we propose to use UPC to non-parametrically obtain approximate eigenfunctions $\hat{\Phi}$, eigenvalues Λ, and modes V. We hereby make use of the UPC property that it first estimate the trajectory z_t, which provides information on the approximate eigenfunction Φ.

In presenting our ideas we will assume the following, which we will refer to as the *perfect mode* assumption.

Assumption 1. *The discrete-time dynamical system f has a finite dimensional Koopman invariant subspace $\mathcal{S}$, which includes the identity observable,* $\mathrm{id} \in \mathcal{S}$.

The rest of this work is organized as follows. First we will cover some background material on the Koopman operator and subspace algorithms in more detail. In section 4 we will present our algorithm, which we name *Subspace Koopman Approximation* (SKA). Subsequently, in section 5 we will prove that if f admits the perfect mode assumption (assumption 1), the algorithm produces exact results. Finally, in 6 we will apply our algorithm to both an example which does admit and one which does not admit the perfect mode assumption.

3 Background

In this section we will discuss elementary theory behind linear subspace methods and the link with Koopman operator theory.

Finite Dimensional Koopman Invariant Subspace

With the discussed Koopman theory we are able to express a direct implication of assumption 1, the perfect mode assumption.

Theorem 3.1. *Consider a discrete-time dynamical system f that satisfies assumption 1. That is, f has a Koopman invariant subspace $\mathcal{S} = \text{span}(\{\phi_i\}_{i=1}^N)$, spanned by the eigenfunctions $\{\phi_i\}_{i=1}^N$, such that $\text{id} \in \mathcal{S}$.*

Then, there exists a triple $\{\mathbb{G}^*, K, V\}$, where $\mathcal{G}^* = \{\psi_i\}_{i=1}^N$, $\psi_i : \mathbb{X} \to \mathbb{R}$ is a set of observables, $K \in \mathbb{R}^{N\times N}$ a matrix , and $V \in \mathbb{R}^{n\times N}$ a matrix, such that

$$\Psi \circ f = K\Psi, \text{ and } \text{id} = V\Psi \text{ for } \Psi = [\psi_1, \psi_2, ..., \psi_N]^\top.$$

Furthermore the state space system

$$\begin{aligned} z_{t+1} &= Kz_t \\ \hat{x}_t &= Vz_t, \end{aligned} \tag{4}$$

is equivalent to f in the sense that if $x_0 = \hat{x}_0$, and $z_0 = \Psi(\hat{x}_0)$, the trajectories $\{\hat{x}_t\}_{t=0}^M$ and $\{x_t\}_{t=0}^M$, with $x_{t+1} = f(x_t)$, satisfy $x_t = \hat{x}_t$, and furthermore $z_t = \Psi(\hat{x}_t)$, for all $t = 0, 1, 2, 3, ...,$

Proof. To see that the first claim is true, just pick for $\mathcal{G}^*$ the eigenfunctions: $\psi_i = \phi_i$, and let $K = \Lambda$ with Λ the diagonal matrix of eigenvalues, then $\Psi \circ f = \Phi \circ f = \Lambda\Phi = K\Psi$. Since $\text{id} \in \mathcal{S}$ we can write the identity function in terms of the eigenfunctions, hence $\text{id} = V\Phi = V\Psi$ for some matrix V.

The second claim follows by induction. We prove that if $x_t = \hat{x}_t$ and $z_t = \Psi(\hat{x}_t)$, then $x_{t+1} = \hat{x}_{t+1}$ and $z_{t+1} = \Psi(\hat{x}_{t+1})$.

Assume $x_t = \hat{x}_t$ and $z_t = \Psi(\hat{x}_t)$. Since $\Psi \circ f = K\Psi$ and $z_{t+1} = Kz_t$, it follows directly that $z_{t+1} = (\Psi \circ f)(\hat{x}_t) = \Psi(\hat{x}_{t+1})$. Furthermore, since $\text{id} = V\Psi$, $\hat{x}_t = Vz_t$, and $x_t = \hat{x}_t$, we have $\hat{x}_{t+1} = Vz_{t+1} = V\Psi(\hat{x}_{t+1}) = V\Psi(x_{t+1}) = x_{t+1}$.

It follows that if $x_0 = \hat{x}_0$ and $z_0 = \Psi(\hat{x}_0)$, then $x_t = \hat{x}_t$, and $z_t = \Psi(\hat{x}_t)$ for all t. □

This little theorem thereby firmly establishes the connection between Koopman operator theory and linear system theory.

Linear Subspace Methods

Let us now look at subspace methods. Originally, subspace methods have been developed for linear systems. The linear subspace identification problem, as formulated in the seminal work of Van Overschee and De Moor [22] is stated in figure

1. Assuming a time series of measurement originates from a linear system, these algorithms are able to identify a corresponding linear system in state space representation, together with the trajectory in the state space, from one single time series.

Stochastic identification problem:

Given: s measurements of the output $y_k \in \mathbb{R}^l$ generated by the unknown stochastic system of order n:

$$x_{k+1}^s = Ax_k^s + w_k \,, \tag{3.1}$$
$$y_k = Cx_k^s + v_k \,, \tag{3.2}$$

with w_k and v_k zero mean, white vector sequences with covariance matrix:

$$\mathbf{E}[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^T & v_q^T \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \,. \tag{3.3}$$

Determine:

- The order n of the unknown system
- The system matrices $A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{l\times n}$ up to within a similarity transformation and $Q \in \mathbb{R}^{n\times n}, S \in \mathbb{R}^{n\times l}, R \in \mathbb{R}^{l\times l}$ so that the second order statistics of the output of the model *and* of the given output are equal.

Figure 1: The subspace identification problem. Copied from [22].

Multiple variants of subspace algorithms are available. We will focus on the UPC algorithm. We will not discuss how this algorithm actually works, instead we refer the reader to [22].

Let us introduce some matrices. X, X', Z, Z' are block Hankel matrices as follows:

$$X = [\bar{x}^0, \bar{x}^1, ..., \bar{x}^i],\ X' = [\bar{x}^1, \bar{x}^2, ..., \bar{x}^i],\ Z = [z_1, ..., z_{i-1}, z_i],\ Z' = [z_1, ..., z_{i-1}], \tag{5}$$

with $\bar{x}^j = [x_j, x_{j+1}, ..., x_{j+4N}]^\top$[2].

In [23] it is shown that the subspace algorithm UPC computes a solution of

[2]In this thesis we maintain $4N$ as the length of the columns of X, with N the dimension of the linear system SKA tries to fit. There is some freedom in the choice of this length. We refer the interested reader to [22].

the following optimization problem:

$$\begin{aligned} &\min_{\Gamma,Z} \|X' - \Gamma Z'\|_F \\ &\text{s.t. } \operatorname{row}(Z) \subseteq \operatorname{row}(X), \end{aligned} \tag{6}$$

where Γ is the observability matrix, with structure $\Gamma = [C, CA, ..., CA^{N-1}]^\top$, and $\operatorname{row}(\cdot)$ is the row space of a matrix.

We will now restate the linear subspace problem and the optimization formulation of [23] in a way more suitable for the purpose of this thesis. A state space representation of a discrete-time linear system has the form,

$$\begin{aligned} z_{t+1} &= Az_t \\ x_t &= Cz_t \end{aligned} \tag{7}$$

here $z_t \in \mathbb{Z}$, with $\mathbb{Z} \subseteq \mathbb{R}^N$ the state space and $x_t \in \mathbb{X}$, with $\mathbb{X} \subseteq \mathbb{R}^n$ the output space.

Theorem 3.2. *Consider a time series of output measurements $\{x_t\}_{t=0}^M$ originating from a discrete-time linear system as in equation 7, and assume knowledge of the parameter N. Then the UPC algorithm computes a solution of the optimization problem in equation 6, for matrices as in equation 5.*

Proof. See [23]. □

4 Method

Motivated by theorem 3.1 we can apply linear subspace methods to nonlinear systems. This theorem seems to imply that the recovered trajectory $\{z_t\}_{t=0}^M$ as found by the subspace algorithm are observations of some observable functions. Apart from some minor issues, as discussed later in remark 1, the relation between the measurements and the recovered linear trajectory $x_t \to z_t$ defines a map $S : \mathbb{X}_S \to \mathbb{Z}$. This map can be interpreted to be a discretization of some vector observable $\Psi : \mathbb{X} \to \mathbb{R}^N$. If the input $\{x_t\}_{t=0}^M$ would approach $\mathbb{X}$ in the $M \to \infty$ limit, we get that S approaches Ψ. Therefore we can interpret the subspace identification equivalently as constructing the Koopman operator and the observables simultaneously. Because the linear subspace problem is ill-posed, up to a similarity transform, we can always obtain a form in which K is diagonal. In that case the corresponding limiting observables Ψ are approximate eigenfunctions of the Koopman operator, with eigenvalues as given in the diagonal K matrix. For finite M we interpolate between all points in order to obtain an approximation $\hat{\Psi}$ of Ψ.

In this section we will develop these ideas further and propose a concrete algorithm. In section 5 we will dive into the theoretical aspects.

We assume Ψ is smooth. In order to obtain reasonable approximation for finite M we require that the sample density is approximately constant over $\mathbb{X}$. This ensures that there is no neighbourhood of $\mathbb{X}$ where the error introduced by interpolation is relatively large.

Originally, subspace methods are designed to work with a single time series. This would therefore restrict the usage of our approach to ergodic systems. For non-ergodic systems a single time-series will not be well distributed over the state space. Since this excludes many dynamical systems and would require very long time series, we make a slight adaptation to the original subspace method.

Consider the definition of X and Z as given previously together with optimization problem 6. Note that for each column $\bar{x}^j$ of X' a single $z_j \in \mathbb{Z} \subseteq \mathbb{R}^N$ instance is estimated, such that $\bar{x}^j \approx \Gamma z_j$. This suggests that instead of using a single time series, we can pick for each column a short time series, of which the initial point is independent of the other columns. In the following part we will use superscript notation to distinguish sequences. Suppose we have M short time series $\{x_t^j\}_{t=0}^N$, $j = 1, 2, .., M$, and $x_{t+1}^j = f(x_t^j)$, we redefine X and Z as

$$X = [\bar{x}^0, \bar{x}^1, ..., \bar{x}^{M-1}], X' = [\bar{x}^1, \bar{x}^2, ..., \bar{x}^{M-1}], \qquad Z = [z_0^1, z_0^2, ..., z_0^M], Z' = [z_0^1, z_0^2, ..., z_0^{M-1}] \tag{8}$$

with

$$\bar{x}^j = [x_0^j, x_1^j, ..., x_N^j]^\top, \qquad x_{t+1}^j = f(x_t^j). \tag{9}$$

In the upcoming theory section we will prove that the UPC algorithm still computes a solution of the optimization problem 6 for the redefined matrices. Let $\Gamma_{\mathrm{a}} = [V, VK, VK^{N-1}]^\top$, and let Z as just defined be the result of applying UPC. The subscript a indicates that it concerns the Γ as provided by the algorithm. We define $\mathbb{X}_{\mathrm{S}} := \{x_t^j \in \mathbb{X} |$ for $j = 1, 2, ..., M,$ and $t = 0, 1, ..., N\}$ to be the collection of state space samples. The map $S : \mathbb{X}_{\mathrm{S}} \to \mathbb{Z}$ can be constructed, in practise, as follows:

$$S(x) = \begin{cases} z_0^j \text{ for } x = x_0^j \text{ for some } j \\ K^k z_0^j \text{ for } x = x_k^j \text{ for } j \text{ and } k = 1, 2, ..., N \end{cases}$$

We assume $\mathbb{X}$ is bounded. If we generate these short time series using initial conditions which are taken at random uniformly, or regularly, over the state space, we have that in the limit of $M \to \infty$ we recover a set of observables. For finite M we can interpolate between all points in order to obtain an approximation $\hat{\Psi}$ of Ψ. If we furthermore require that K is diagonal we obtain approximations of Koopman eigenfunctions.

Remark 1. If there is no Koopman invariant subspace which includes the identity observable of dimension N or less, then the computed z_t^j will be approximations of $\Phi(x_t^j)$. If furthermore, two bursts $\bar{x}^i$ and $\bar{x}^j$ partially overlap such that there is an x that $x \in \bar{x}^i$ and $x \in \bar{x}^j$, then the algorithm will associate two different values to the same x. In such a case one can not directly construct a function S as above, but this can be resolved by taking the average of the different values associated. Alternatively, one can skip the construction of S and work with the collection of tuples of points $(x_t^j, K^t z_0^j)$ and in the interpolation step use a regression technique which can handle non-unique points.

This brings us to our proposed method for estimating the Koopman operator, eigenfunction, eigenvalues, and modes, the Subspace Koopman Approximation (SKA) algorithm:

Input: M short time series $\bar{x}^j$ of length $4N$ as generated by f, according to equation 9, and a parameter d the amount of eigenfunctions the algorithm approximates.

Output: Intermediate: a tuple (Γ_a, S), with observability matrix Γ_a and function $S : \mathbb{X}_S \to \mathbb{Z}$. Final: a triple $(\hat{\Psi}, K, V)$, with $\hat{\Psi}$ approximate Koopman eigenfunctions of f, K a diagonal matrix with approximate Koopman eigenvalues, and V a matrix whose columns are approximate Koopman modes.

1 Construct matrix X as in equation 8.
2 Using the UPC algorithm as in equation 6 calculate Γ_a and Z, such that K is in diagonal form.
3 Construct the map $S : \mathbb{X}_S \to \mathbb{Z}$.
4 Output intermediate result: (Γ_a, S).
5 Interpolate S to obtain the vector valued approximate Koopman eigenfunction $\hat{\Phi} : \mathbb{X} \to \mathbb{C}^N$, with approximate Koopman eigenfunctions $\hat{\phi}_i : \mathbb{X} \to \mathbb{C}$ and $\hat{\Phi} = [\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_d]^\top$.
6 Extract K and V from Γ_a.
7 Output final result: $(\hat{\Psi}, K, V)$.

Algorithm 1: The Subspace Koopman Approximation (SKA) algorithm.

5 Theory

First we will prove that the UPC algorithm functions as desired when applied to non-Hankel matrices, then the bulk of this section will be devoted to the relation between SKA and the Koopman operator.

UPC for non-Hankel Matrices

In this section we will prove a modified version of theorem 3.2. We will show that the UPC algorithm computes a solution of 6 irrespective of the structure of matrix X. We will follow closely, the proof provided in [23], but pay particular attention to the fact that it does not rely on the Hankel structure of X. Before stating the theorem we need to introduce some notation. We will follow the conventional terminology to make the link with existing work more transparent.

Let $X \in \mathbb{R}^{i \times M}$, $X_f = X' = X[:, 2 : \text{end}]$, and $X_p = X[:, 1 : \text{end} - 1]$, again using Matlab index notation. That is, X_f contains all but the first column of X and X_p all but the last. Also, $Z \in \mathbb{R}^{N \times M}$ and $Z_p = Z' = Z[:, 1 : \text{end} - 1]$ contain all but the first column of Z. At this moment it is already important to note that we can define X_f, X_p just as easy if X has no Hankel structure. The only thing that changes is the meaning. Normally, f, p indicate future and past, but if there is no Hankel structure, the columns do not relate to each other chronologically. We use the notation $A/B = AB^+B$ and $A/B^\perp = A(I - B^+B)$.

Furthermore, some more background material is required on the workings of

the UPC algorithm. The UPC algorithm identifies $\Gamma = US^{1/2}$, where $(X_f/X_p)_N = USV^\top$ is an N-truncated singular value decomposition [22, 23]. Now we are ready to state the theorem and proof.

Theorem 5.1. *The UPC algorithm computes a solution of the optimization problem in equation 6 irrespective of the structure of X.*

Proof. We need to prove that there exists a matrix Z such that Γ, Z solves 6, that is minimizes $\|X_f - \Gamma Z_p\|_F$ under condition $\mathrm{row}(Z) \subseteq \mathrm{row}(X)$. Following the same argumentation as in [23], we have that the objective function is equivalent to minimizing $\|X_f/X_p - \Gamma Z_p\|$. This follows from

$$\|X_f - \Gamma Z_p\|_F = \|X_f/X_p - \Gamma Z_p\|_F + \|X_f/X_p^\perp\|,$$

where we use that the rows of Z_p are orthogonal to $X_f/X_p^\perp$. To see the latter note that $\mathrm{row}(Z_p) \subseteq \mathrm{row}(X_p)$, while $\mathrm{row}(X_f/X_p^\perp) \cap \mathrm{row}(X_p) = \emptyset$. This does not depend on X being Hankel, it only depends on $\mathrm{row}(X_f/X_p^\perp)$. Note that the second term, $\|X_f/X_p^\perp\|$ is constant. The rest of the proof is also identical.

The lower bound of $\|X_f/X_p - \Gamma Z_p\|_F$ is attained at $\Gamma Z_p = (X_f/X_p)_N$, the truncated singular value decomposition. If we choose $Z = S^{-1/2}U^\top X_f X_p^\top X$, then $Z_p = S^{-1/2}U^\top X_f X_p^\top X_p$, and clearly $\mathrm{row}(Z) \subseteq \mathrm{row}(X)$. Furthermore, $\Gamma Z_p = (X_f X_p^\top X_p)_N = (X_f/X_p)_N$. □

Where initially the intuition behind UPC and subspace methods was to project past trajectories on future trajectories, as in $\mathbb{X}_f/\mathbb{X}_p$, in turn out that even though this intuition behind UPC when applied to non-Hankel matrices is missing, it still solves the optimization problem 6.

SKA and the Koopman Operator

We will now discuss properties of the approximate Koopman operator and eigenfunctions as provided by the SKA algorithm. Consider the discrete-time dynamical system $f : \mathbb{X} \to \mathbb{X}$ on state space $\mathbb{X}$, and a generic function space $\mathbb{G}$ that contains functions of the form $g : \mathbb{X} \to \mathbb{C}$, and is closed under composition with f. By definition $\mathbb{G}$ is Koopman invariant. We call a subspace $\mathcal{S} \subseteq \mathbb{G}$ Koopman invariant if for each $g \in \mathcal{S}$ we have $g \circ f \in \mathcal{S}$. As is apparent from problem statement 6, each row of Z is a linear combination of the rows of X. Since X contains the state measurement and delays of the state measurements, this directly implies that the approximate eigenfunctions lie in the span of the dynamics and its delays, $\hat{\psi}_i \in \mathcal{F}_N$, with

$$\mathcal{F}_N := \mathrm{span}(\{\mathrm{id}_1, \mathrm{id}_2, ..., \mathrm{id}_n, \mathrm{id}_1 \circ f, \mathrm{id}_2 \circ f, ..., \mathrm{id}_n \circ f, ..., \mathrm{id}_1 \circ f^N, \mathrm{id}_2 \circ f^N, ..., \mathrm{id}_n \circ f^N\}).$$

Where $\mathrm{id}_j : \mathbb{X} \to \mathbb{R}$ is the j'th element of $\mathrm{id} : X \to \mathbb{X} \subseteq \mathbb{R}^n$.

Using this insight we can actually see optimization problem 6 as a specification of the more general Koopman mode optimization problem

$$\min_{v_i \in \mathbb{R}^N, \lambda_i \in \mathbb{C}, \phi_i \in \mathcal{F}_N} \left\| \sqrt{\sum_{k=1}^{\tau} \left(f^k - \sum_{i=1}^{N} v_i \lambda_i^k \phi_i \right)^2} \right\|_\mu , \tag{10}$$

where $\|\cdot\|_\mu$ is the norm $\|p\|_\mu := \sqrt{\int_{\mathbb{X}} p^2 \, \mathrm{d}\mu}$. Choosing for μ the emperical measure $\rho := \frac{1}{M} \sum_{j=1}^{M} \delta_{x_0^j}$, the above optimization problem closely resembles optimization problem 6, with the only minor difference that the latter optimization problem excludes the first data point in the objective function.

At first instance the restriction $\phi_i \in \mathcal{F}_N$ may appear to be a major restriction of this algorithm. We will show that in fact there is no restriction. We distinguish two cases, the case where $\mathcal{F}_N$ for $N \to \infty$ is finite dimensional and the case where it is infinite dimensional.

Let us start with the first case. If $\mathcal{F}_N$ for $N \to \infty$ is finite dimensional, then there exists N^* for which $\dim(\mathcal{F}_{N_1}) < \dim(\mathcal{F}_{N^*}) = \dim(\mathcal{F}_{N_2})$ for all $N_1 < N^*$ and $N_2 > N^*$. It follows directly that $\mathcal{F}_{N^*}$ is a minimal finite dimensional Koopman invariant subspace. In the next section we will prove that in such a situation the SKA algorithm yields exact results.

In the case $\mathcal{F}_N$ for $N \to \infty$ is infinite dimensional, we will prove in two stages that the Koopman operator defined by the SKA algorithm converges to the actual Koopman operator. We show $\mathcal{F}_N \to \mathbb{G}$ where we consider $\mathbb{G} = L_2(\mu)$, as in [10], and show that for $M \to \infty$ $\mathcal{K}_{N,M} \to \mathcal{K}_N$. This is further discussed in the subsection 5.

Exact Results

Let us first consider an example of a system for which $\mathcal{F}_N$ for $N \to \infty$ is finite dimensional. This is the case for each linear system. Consider a linear system of the form $x_t = A^t x_0$, with $x_t \in \mathbb{R}_N, A \in \mathbb{R}^{N \times N}$. We have $\mathrm{id}_j(x) = e_j x$, with e_j the j'th unit row vector. Furthermore, $(\mathrm{id}_j \circ f^k)(x) = e_j A^k x$. It is well known that $e_j A^N$ can be written as a linear combination of $e_j A^0, e_j A^1, ..., e_j A^{N-1}$. Then we see directly that the corresponding space $\mathcal{F}_N$ cannot have dimension greater than N^2 (dimension of state space times number of compositions with f).

We will now prove that for a specific class of dynamical systems, the SKA algorithm yields exact results. Let us now provide a definition and a small lemma which we need later on.

Definition 5.1. A finite dimensional Koopman invariant subspace $\mathcal{S}$ of f is a *minimal* Koopman invariant subspace, which *includes the identity* if $\mathrm{id}_j \in \mathcal{S}$ for $j = 1, 2, .., n$ and for every Koopman invariant subspace $\mathcal{S}^*$ of f, with $\mathrm{id}_j \in \mathcal{S}^*$, we have $\dim(\mathcal{S}) \leq \dim(\mathcal{S}^*)$. Here $\dim(\cdot)$ is the Hamel dimension, and $\mathrm{id} = [\mathrm{id}_1, \mathrm{id}_2, ..., \mathrm{id}_n]^\top$.

Just as in the above definition, we will abuse language and notation by speaking about the inclusion of a vector valued function in $\mathcal{S}$, a space of scalar valued

functions. Let us formalize what we mean. For a_i the components of some vector function $a : \mathbb{X} \to \mathbb{C}^s$, such that $a = [a_1, a_2, ..., a_s]^\top$, a subspace $\mathcal{S}$ of $\mathbb{G}$ includes a if and only if $a_i \in \mathcal{S}$ for $i = 1, 2, ..., s$. Furthermore, we will abuse notation by writing $a \in \mathcal{S}$ if and only if $\mathcal{S}$ includes a.

The following lemma is elementary, but makes the proofs in coming theorems easier.

Lemma 5.2. *If the dynamical system f admits the perfect mode assumption (that is assumption 1), then there exists a unique and minimal finite dimensional Koopman invariant subspace which includes the identity function.*

Before proceeding to a proof, we introduce the notion of a Krylov space, which for this text we define as follows. The k-th Krylov subspace Kr_k of a tuple

$$(a : \mathbb{X} \to \mathbb{C}, T : \mathbb{X} \to \mathbb{X})$$

is given by

$$\mathrm{Kr}_k(a, T) := \mathrm{span}(\{a, a \circ T, a \circ T^2, ..., a \circ T^{k-1}\}).$$

We extend this definition by allowing for vector valued functions a. For $a : \mathbb{X} \to \mathbb{C}^s$ we define

$$\mathrm{Kr}_k(a, T) := \mathrm{span}(\{a_1, a_1 \circ T, ..., a_1 \circ T^{k-1}, a_2, a_2 \circ T, ..., a_2 \circ T^{k-1}, ..., a_s, a_s \circ T, ..., a_s \circ T^{k-1}\}).$$

Now we are ready to state the proof of the just mentioned lemma.

Proof. The existence of a minimal Koopman invariant subspace which includes the identity is trivial. It remains to prove that it is unique. Assume the dynamical system f has a finite dimensional minimal Koopman invariant subspace $\mathcal{S}$ which includes the identity function. Proof of uniqueness by contradiction. Assume, furthermore, that there are two distinct minimal Koopman invariant subspaces which contain the identity function, $\mathcal{S}_1 \neq \mathcal{S}_2$.

$\mathcal{S}$ is finite dimensional and Koopman invariant, so there exists minimal $k \in \mathbb{N}$ such that $\mathrm{Kr}_k(\mathrm{id}, f) = \mathrm{Kr}_j(\mathrm{id}, f) =$ for all $j \geq k$. Since $\mathrm{id} \in \mathcal{S}_1$ and $\mathcal{S}_1$ is Koopman invariant, we have $\mathrm{id} \circ f^k \in \mathcal{S}_1$ for each $k \in \mathbb{N}$, therefore $\mathrm{Kr}_i(\mathrm{id}, f) \subseteq \mathcal{S}_1$ for all $i \in \mathbb{N}$. Because $\mathcal{S}_1$ is minimal we have $\mathcal{S}_1 = \mathrm{Kr}_k(\mathrm{id}, f)$.

The same argumentation is true for $\mathcal{S}_2$. So $\mathcal{S}_1 = \mathrm{Kr}_k(\mathrm{id}, f) = \mathcal{S}_2$, and we have a contradiction. We conclude that the minimal Koopman invariant subspace is unique. □

In order to be able to say something about the uniqueness of the recovered approximation of the eigenfunctions we need the notion of observability and associated ideas from systems theory.

Definition 5.2. A dynamical system is called observable if a finite sequence of output measurements uniquely determines the state. Moreover, if the dynamical system is defined by a linear state space system with system matrices (A, C) as in equation 7, the tuple (A, C) is called observable if and only if the associated dynamical system is observable.

We need the following well known [25] theorem on observability.

Theorem 5.3. *For $A \in \mathbb{C}^{N\times N}, C \in \mathbb{C}^{n\times N}$, and Γ the observability matrix, the following statements are equivalent:*

1. *The tuple (A, C) is observable,*
2. $\operatorname{rank}\begin{pmatrix} A-\lambda I \\ C \end{pmatrix} = N$ *for each eigenvalue λ of A,*
3. $\operatorname{rank}\Gamma = N$.

Let us state an intermediate lemma, before going over to the main theorem.

Lemma 5.4. *Consider a discrete-time dynamical system f, with a minimal Koopman invariant subspace $\mathcal{S} = \operatorname{span}(\{\psi_i\}_{i=1}^N)$, spanned by the eigenfunctions $\{\psi_i\}_{i=1}^N$, such that $\mathrm{id} \in \mathcal{S}$. Let K be the diagonal matrix containing the Koopman eigenvalues, and V be the matrix containing the modes, as in equation 3.*

Then the tuple (K, V) is observable.

Proof. We introduce the shorthand notation

$$E_\lambda = \begin{pmatrix} K-\lambda I \\ V \end{pmatrix}.$$

Proof by contradiction. Assume (K, V) is not observable. By theorem 5.3 there is a Koopman eigenvalue λ^* for which

$$\operatorname{rank} E_{\lambda^*} < N.$$

This implies there is a column of E_λ that can be written as a linear combination of the other columns. Furthermore, because K is a diagonal matrix containing the eigenvalues, the matrix $K-\lambda^* I$ contains as many zero columns as the multiplicity of λ^*.

Consider the case that this eigenvalue λ^* has multiplicity 1. The j'th column of matrix $K-\lambda^* I$ is the zero vector, for some j corresponding to the location of λ^*. All other columns are linearly independent because K is diagonal. So, necessarily the j'th column of E_λ should be a linear combination of the other columns. It is easy to see that this implies that the j'th column of E_λ is the zero vector. This implies C has a zero column. Therefore, there is a mode v_j equal to zero and the space $\{\psi_j\}_{j\neq i^*}$ also contains the identity function. This space is smaller than $\mathcal{S}$, so $\mathcal{S}$ is not a minimal Koopman invariant subspace.

Consider the case that the eigenvalue has multiplicity greater than 1. Several columns of $K-\lambda I$ will be zero. Let $\mathcal{I}$ be the set of indices of these columns. All other columns are linearly indepedent. Necessarily, the columns of E_λ with index $j \in \mathcal{I}$ are linearly dependent. Recall that V contains the Koopman modes. This implies there are linearly dependent Koopman modes with the same associated eigenvalue. So the Koopman modes satisfy

$$v_{i^*} = \sum_{j\neq i^*, j\in\mathcal{I}} \beta_j v_j,$$

for some i^*. Because $\mathcal{S}$ includes the identity function we have:

$$\begin{aligned}\mathrm{id} &= \sum_i v_i\psi_i = \sum_{j\neq i^*} v_j\psi_j + \sum_{j\neq i^*, j\in\mathcal{I}} \beta_j v_j \psi_{i^*} \\ &= \sum_{j\neq i^*, j\notin\mathcal{I}} v_j\psi_j + \sum_{j\neq i^*, j\in\mathcal{I}} v_j(\psi_j + \beta_j\psi_{i^*}).\end{aligned}$$

Since the eigenvalues of the eigenfunction with index $j \in \mathcal{I}$ are the same, the functions $\psi_j + \beta_j\psi_{i^*}$, for $j \in \mathcal{I}$ are eigenfunctions as well. It follows that we can formulate a Koopman invariant subspace which contains the identity function that is smaller than $\mathcal{S}$, hence $\mathcal{S}$ is not a minimal Koopman invariant subspace.

Since the cases are exhaustive, the assumption that (K, V) is not observable, implies that the space $\mathcal{S}$ is not minimal. This is a contradiction. We conclude that (K, V) is observable. □

Now we are ready to state and prove the main theorem.

Theorem 5.5. *Consider a discrete-time dynamical system f that admits the perfect mode assumption. Let $\mathcal{S} = \mathrm{span}(\{\phi_i\}_{i=1}^N)$, with $\mathrm{id} \in \mathcal{S}$ be the minimal Koopman invariant subspace which includes the identity. Suppose $\dim(\mathcal{S}) = N$. Here ϕ_i are again eigenfunctions. Denote by V the matrix of modes, by K the diagonal matrix of Koopman eigenvalues, and let $\Gamma_f = [V, VK, ..., VK^{N-1}]^\top$ be the observability matrix associated to f.*

Assume one has access to $M > 4N$ bursts of state measurements $\{x_t^j\}_{t=0}^{4N}$ of length $4N$. The SKA algorithm, with parameter $d = N$, intermediate result (Γ_a, S) satisfies:

- $\Gamma_\mathrm{a} = \Gamma_f$
- *$S(x) = \Phi(x)$ for all $x \in \mathbb{X}_\mathrm{S}$, and $\Phi : \mathbb{X} \to \mathbb{R}^N$ some concatenation of the eigenfunctions $\{\phi_i\}_{i=1}^N$.*

Proof. Since $\mathcal{S}$ is a Koopman invariant subspace including the identity function, we have by theorem 3.1, that there exist matrices K, and V satisfying equation 4. It follows directly that

$$\bar{x}^j = \Gamma_f \Psi(x_0^j).$$

Therefore (Γ_f, Z) with $z_j = \Psi(x_0^j) = \Gamma_f^{-1}\bar{x}^j$, is a solution of problem 6.

It remains to prove that it is the unique solution (up to a permutation of the rows of Z). First we establish that Γ_f is the unique Γ which solves the problem. Suppose there are two different Γ for which there exists Z' such that $X' - \Gamma Z' = 0$, and have A in diagonal form. Since we picked $d = N$ as our algorithm parameter, we have that Γ is full rank. Otherwise there would be a Koopman invariant subspace of dimension smaller than N.

Because the observability matrices are different and full rank the Z matrices for which the optimization problem are solved are also different. Hence, there are two separate sets of eigenfunctions associated.

This means that there are two distinct minimal Koopman invariant subspaces, which contradict with theorem 5.2.

Finally we need to establish that S necessarily coincides with Ψ. This follows from theorem 5.4 and 5.3, because Γ_f is full rank, Z is unique and therefore S is unique. □

A direct consequence is the following.

Theorem 5.6. *Consider a discrete-time dynamical system f that admits the perfect mode assumption. Let $\mathcal{S} = \text{span}(\{\phi_i\}_{i=1}^N)$, with $\text{id} \in \mathcal{S}$ be the minimal Koopman invariant subspace that includes the identity. Suppose $\dim(\mathcal{S}) = N$. Assume one has access to $M > 4N$ bursts of state measurements $\{x_t^j\}_{t=0}^{4N}$ of length $4N$, for which the initial points x_0^j of each short burst $\bar{x}^j$ are regularly spaced on the state space. Let $\hat{\Phi}_M$ be the final result of the SKA algorithm with linear interpolation, with parameter $d = N$, and number of samples M. Then as the number of burst M tends to infinity we obtain the exact eigenfunctions Φ:*

$$\|\hat{\Phi} - \Phi\| \to 0 \text{ as } M \to \infty.$$

Proof. From theorem 5.5 it follows that for all $x \in \mathbb{X}_S$ $S(x) = \Phi(x)$. Linear interpolation implies $\hat{\Phi}(x) \in \text{affinespan}(\{S(y)|y \in \mathbb{X}_x\})$, where $\mathbb{X}_x$ contains the $2n$ elements in $\mathbb{X}_S$ closest to x. For $M \to \infty$ we have $y \to x$ for all $y \in \mathbb{X}_x$. Finally, since Φ is continuous, $\text{affinespan}(\{S(y)|y \in \mathbb{X}_x\}) = \text{affinespan}(\{\Phi(y)|y \in \mathbb{X}_x\}) \to \Phi(x)$. □

Convergence

We would like to prove that the SKA induced approximation of the Koopman operator converges to the true Koopman operator. Let $\mathcal{K}_{N,M}$ be the SKA approximation of the Koopman operator for M short burst of length $4N$ and a linear system approximation of dimension N. Given the SKA output $K, \hat{\Phi}$, we define $\mathcal{K}_{N,M}$ by

$$\mathcal{K}_{N,M}(g) := a_g K \hat{\Phi},$$

for $a_g \hat{\Phi} = g$. Because SKA does not attempt to find optimal eigenfunctions, but instead an optimal mode decomposition, we cannot directly follow the proof structure of [10]. We have not develop a proof, but we will share our intuition as to why we expect that it can be proven

$$\mathcal{K}_{N,M} \to \mathcal{K} \text{ as } N, M \to \infty.$$

We restrict our attention to the case that $\mathcal{F}_N$ converges to a basis of a, to be specified, infinite dimensional function space $\mathbb{F}$. Consider the true Koopman modes $\{v_i\}_{i=1}^\infty$, infinite matrix representation of the Koopman operator K, and Koopman eigenpairs $\{\phi_i\}_{i=1}^\infty$, $\{\lambda_i\}_{i=1}^\infty$. Allowing for infinite matrices, let $z_0^j = \Phi(x_0^j)$ for some concatenation of the eigenfunctions Φ. Let Γ be the observability matrix defined by K and the infinite matrix of Koopman modes V. Then Γ, Z solve optimization problem 6, because

$$x_t^j = \sum_{i=1}^{\infty} v_i \lambda_i^t \phi_i(x_0^j).$$

Comparison with Hankel DMD

In this short section we will compare SKA with Hankel DMD [26, 27]. Hankel DMD can be seen as a version of EDMD where as the library of observables one takes delay coordinates. Obviously SKA and EDMD are different algorithms, but how do the results of Hankel DMD and SKA compare?

We distinguish three properties. Firstly, Hankel DMD is only capable of providing analytical relations for the eigenfunctions, that is it provides approximate eigenfunctions as a linear combination of the delay coordinates. SKA provides this too, but in addition it provides an actual function that one can evaluate on the state space. This is relevant for complex or real world examples where there is no usable analytical expression of the dynamics.

Secondly, Hankel DMD requires a single time series as input, while SKA does not require this. Therefore, using SKA offers more flexibility, which is especially useful for non-ergodic systems.

Lastly, the optimization problem that Hankel DMD and SKA solve is different. Effectively, EDMD tries to find eigenfunctions, while SKA tries to find a set of functions which are approximately eigenfunctions and are such that the identity is contained approximately in their span. This suggests that Hankel DMD might be better suited for Koopman eigenfunction approximation, while SKA for Koopman mode decomposition.

6 Results

In this section we apply our SKA method to various non-linear dynamical systems: systems for which it is known the perfect-mode assumption 1 is satisfied, systems for which this is unknown, and systems for which it is known that the assumption does not hold.

Method

We will use the accuracy criterion as introduced by [28] to assess the quality of the eigenfunction approximations. The criterion is entirely data driven and computes the out-of-sample performance of the Koopman eigenfunctions. Given a set of points $\{x_k\}_k$ where each point is drawn independently from the uniform distribution, and $\{x_k^\#\}_k$ is the image under f, $x_k^\# := f(x_k)$, the quality α of an approximate Koopman eigenfunction $\hat{\phi}$ can be approximated by:

$$\alpha(\hat{\phi}) := \frac{\sum_k |\hat{\phi}(x_k^\#) - \mu\phi(x_k)|}{\sum_k |\hat{\phi}(x_k)|}. \tag{11}$$

This in itself does not tell anything about the quality of the mode decompostion. Remember that we are interested in obtaining a finite approximate Koopman mode

decomposition $\sum_{i=1}^{N} v_i \lambda_i^t \hat{\psi}_i$. Ideally we wish to find a Koopman mode decomposition such that it yields a reasonable approximation of the dynamics for finite time,

$$f^t \approx \sum_{i=1}^{N} v_i \lambda_i^t \hat{\psi}_i.$$

Motivated by this relation and the just mentioned eigenfunction criterion, we propose to calculate the following mode accuracy criterion:

$$\beta_T := \frac{\sum_k |f^T(x_k) - V \Lambda^T \hat{\Phi}(x_k)|}{\sum_k |f^T(x_k)|}, \tag{12}$$

for V the matrix of modes, Λ the diagonal matrix of eigenvalues, and $\hat{\Phi}$ the vector valued function of approximate eigenfunctions. Recall that f^T denotes the composition of f with itself T times. We assume again that x_k are independently distributed according to the uniform distribution.

When reporting modes and eigenfunctions, we always report normalized values. We normalize such that $\max_{x \in \mathbb{X}}(|\hat{\phi}(x)|) = 1$. Since we rely on a sampling of the state space the implementation of this normalization might introduce some small errors.

Invariant Example

The first example that we will treat is the map $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x) = \begin{pmatrix} \gamma x_1 \\ \delta x_2 + (\gamma^2 - \delta) x_1^2 \end{pmatrix}, \tag{13}$$

with $\gamma = 0.9$ and $\delta = 0.8$. For this map the Koopman eigenfunctions are known analytically, and the performance of EDMD has been documented using the quality criterion [29]. We copy their setup.

We take $M = 300$ bursts, with initial condition uniformly distributed over $[-1, 1] \times [-1, 1]$. We set $N = 5$, and take bursts of length 20. The results are the following. In table 1 the matrix V of modes is displayed.

$$V = \begin{bmatrix} 0.994e+00 & 8.139e-17 & 8.111e-17 & -1.385e-15 & 3.747e-15 \\ -1.085e-16 & -1.698e+00 & -0.989e+00 & -1.639e-18 & 4.634e-17 \end{bmatrix}$$

Table 1: The mode approximation as computed by the SKA algorithm for example 1 (eq. 13). Each column is accociated with an eigenfunction. The fourth and fifth modes are approximately zero.

The last two columns of the matrix, that is the last two modes, are zero to machine precision, which indicates that the first 3 Z coordinates already give a perfect representation of the dynamics. In other words: there exists a Koopman invariant subspace of dimension 3. Assumption 1 is satisfied. It is easy to see

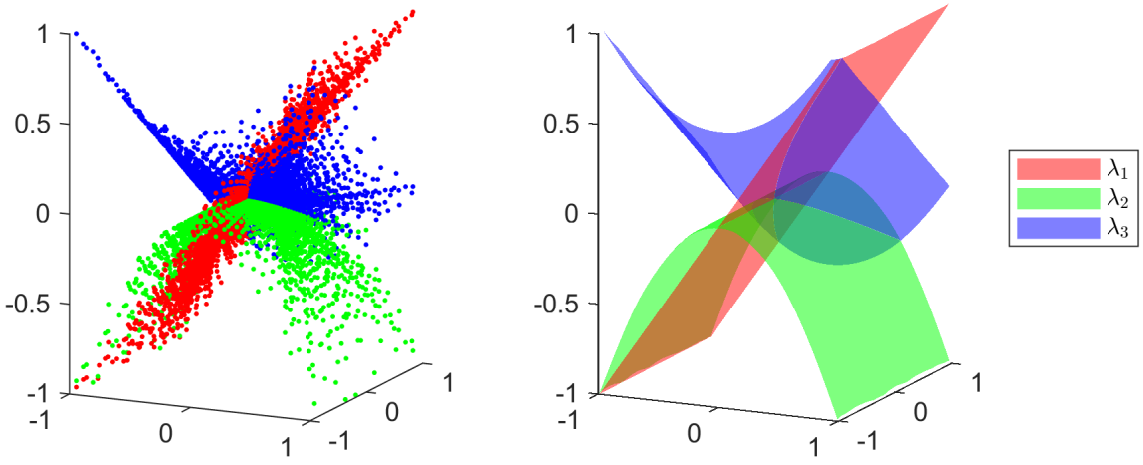

Figure 2: The map $S : \mathbb{X}_S \to \mathbb{Z}$ for example 1 (eq. 13), as computed with SKA using random sampling, restricted to the first three coordinates of Z. The horizontal plane is the state space. Eigenvalue indexes correspond to table 6. On the left are all individual points of map S, while on the right it is interpolated.

analytically that this is indeed true. We can rewrite the dynamical system 13 in a linear form:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} \gamma x_1 \\ \delta x_2 + (\gamma^2 - \delta) x_3 \\ \gamma^2 x_3 \end{pmatrix}$$

If the initial condition satisfies $x_3 = x_1^2$ then the trajectory of the above linear system will be the same as the trajectory of f. The map S, which can be interpreted as the discretization of the first three Koopman eigenfunctions of f, can be observed in figure 2. In figure 3 all five approximate eigenvalues are plotted as heatmaps.

index	eigenvalues	modes	α_{linear}	$\alpha_{natural}$	α_{biharm}	α_{loess2}
1	$9.00e-01$	$1.28e+00$	$3.56e-16$	$3.65e-16$	$8.16e-04$	$4.68e-16$
2	$8.00e-01$	$8.01e-01$	$6.51e-03$	$6.63e-03$	$2.15e-03$	$1.11e-14$
3	$8.10e-01$	$7.77e-01$	$1.18e-02$	$1.20e-02$	$3.49e-03$	$1.99e-14$
4	$-8.86e-02$	$1.76e-08$	$9.88e-01$	$9.90e-01$	$9.94e-01$	$9.80e-01$
5	$-4.41e-01$	$1.88e-08$	$1.32e+00$	$1.32e+00$	$1.33e+00$	$1.34e+00$

Table 2: The out of sample performance of the approximated SKA eigenfunction for example 1 (eq. 13), by interpolation from S, using various interpolation methods: linear interpolation, natural interpolation (C^1), biharmonic interpolation, and a Lowess quadratic non-parametric regression. Under the header of 'modes' we have also reported the norm of the mode.

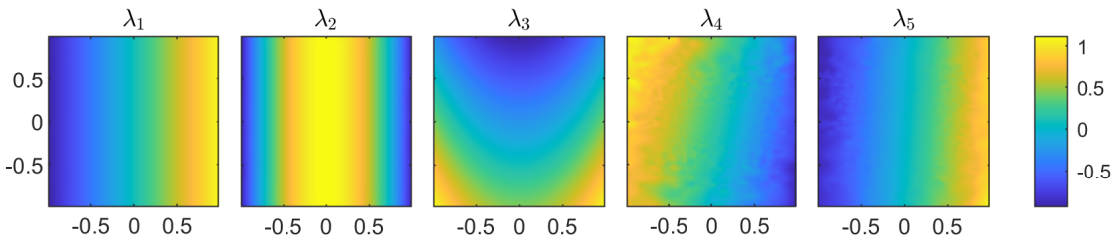

Figure 3: The approximate eigenfunctions $\hat{\phi} : \mathbb{X}_S \to \mathbb{Z}$ for example 1 (eq. 13), as computed with SKA using random sampling. Color indicates function value.

The approximation of the eigenfunctions requires an interpolation. For various interpolation techniques we have calculated the quality of the approximate eigenfunctions. We use 1000 independently and uniformly distributed points. See table 6. We note that the performance heavily depends on choice of interpolation technique. The quadratic Lowess non-parametric yields the best results, of $\alpha \approx 10^{-15}$, this is no surprise since the first three eigenfunction are indeed polynomials of maximal degree 2. EDMD performs comparably, also with $\alpha \approx 10^{-15}$. We therefore conclude that the SKA algorithm is able to retrieve eigenfunctions for this example, if a suitable interpolation method is chosen.

Furthermore, the quality of the fourth and fifth approximate eigenfunction is very poor. This example shows a major drawback of the proposed method. Because it tries to find a minimal Koopman invariant subspace which includes the identity, it is unable to find eigenfunctions that do not lie in this minimal Koopman invariant subspace.

For further study of the performance we use a regular grid of samples instead of uniformly distributed samples. This makes interpolation easier, and is not a restriction for simulation models. We will study β for SKA and EDMD.

We generate in first instance $M = 1024$ short bursts with initial conditions uniformly spaced over $[-1, 1] \times [-1, 1]$, that is, along both directions we have 32 grid points. Secondly, we pick 71 grid points in each direction, to get a total of $M = 5041$ short bursts. For EDMD we supply as the library all monomials $x_1^i x_2^j$ with $i, j = 0, 1, 2, 3, 4, 5$. We set $N = 10$, that is we consider only the first (sorted by mode norm) N eigenfunctions for EDMD, and let SKA compute an N dimensional linear approximation. We use linear interpolation. We know already that there exists a perfect representation for $N = 3$, but we pick $N = 10$ for uniformity with the next example. We use 20000 uniformly random distributed points as our test dataset. Figure 4 shows how β evolves in time for both SKA and EDMD, for both densities of bursts.

The first thing to notice is that EDMD performs better in first instance, which is to be expected because the true eigenfunctions lie in the span of the library, so EDMD is able to obtain the true eigenfunctions, whereas SKA still requires an interpolation step which introduces an error. More interesting though is what

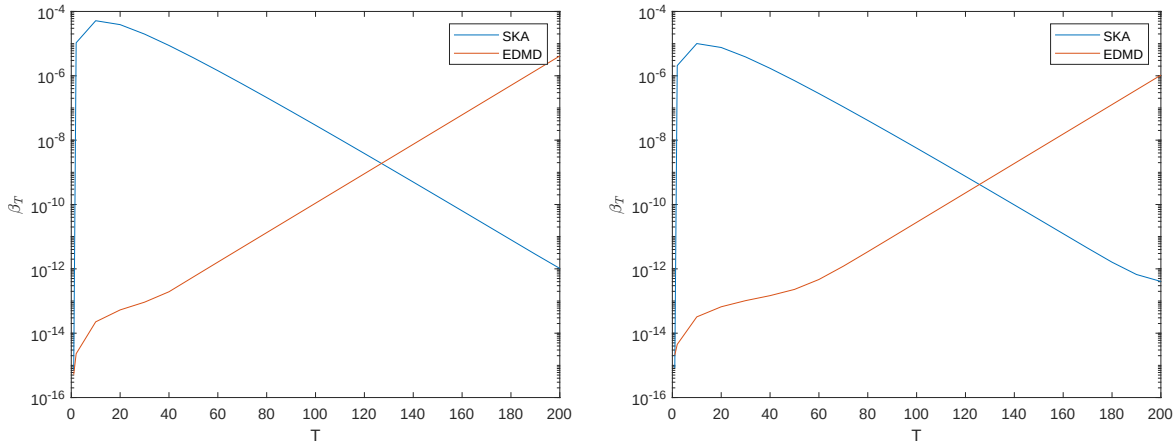

Figure 4: β for example 1 (eq. 13), for both EDMD and SKA. Left uses a training data set of $32 \times 32 = 1024$ points, while the right side is with $71 \times 71 = 5041$ points.

happens later in time. The error for EDMD grows, whereas for SKA it decreases. This is not naively to be expected. Note however that in our definition of β (eq. 12), we are dividing by $\sum_{k=1}^{M} |f^T(x_k)|$. Since the system under study can be regarded as an asymptotically stable linear system, we know that for $T \to \infty$ all trajectories converge to zero. Hence the denominator of β goes to zero. This provides a possible explanation for the observation. For SKA the numerator goes more quickly to zero than the denominator, while for EDMD the denominator goes more quickly to zero. Lastly we note that the results do not change considerably if the sampling resolution increases.

Modified

The first example has a finite dimensional Koopman invariant subspace which includes the identity. Therefore, by the discussed theory, it was to be expected that the algorithm works fine. We will now slightly modify the example such that it probably does not have a finite dimensional Koopman invariant subspace. We consider the dynamical system:

$$f(x) = \begin{pmatrix} \gamma x_1 - \epsilon \exp(-x_1) x_2 \\ \delta x_2 + (\gamma^2 - \delta) x_1^2 \end{pmatrix}, \tag{14}$$

with $\gamma = 0.9$, $\delta = 0.8$ and $\epsilon = 0.1$. This system had one stable equilibrium point at $(0, 0)$. For sake of comparison we first use randomly distributed samples as before, with $M = 300$ bursts. In figure 5 and 6 we display respectively a 3D plot of the first three eigenfunctions and heatmaps of all five approximate eigenfunctions. We observe that all five approximate eigenfunctions now have a smooth shape, as compared to the previous example. In table 3 the eigenvalues, mode norms and accuracy criterion are provided. The accuracy is considerably lower for all interpolation methods, order $\alpha \approx 10^{-3}$. If we should follow the interpretation of the authors of [28], we would need to conclude that these are very poor results.

As a rule of thumb, we consider every approximate eigenfunction with $\alpha < 10^{-10}$ a reasonable approximate eigenfunction. Therefore we conclude that SKA is not able to adequately identify the eigenfunctions. Of course these results depend on the sampling resolution, because we are interpolating.

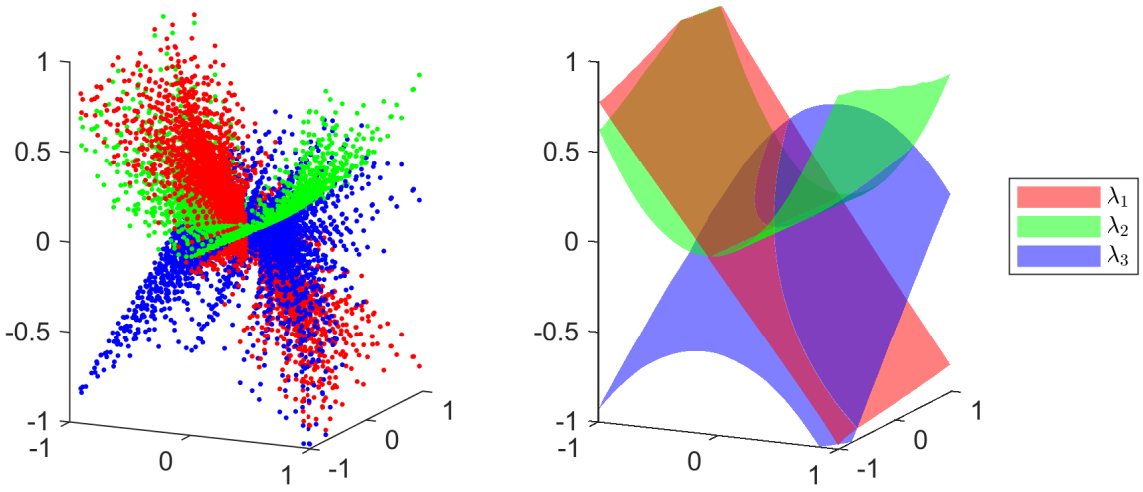

Figure 5: The map $S : \mathbb{X}_S \to \mathbb{Z}$ for example 2 (eq. 14), as computed with SKA using random sampling, restricted to the first three coordinates of Z. The horizontal plane is the state space. Eigenvalue indexes correspond to table 3.

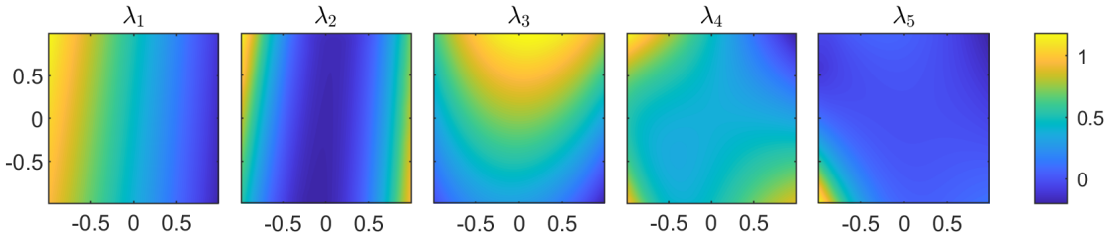

Figure 6: The approximate SKA eigenfunctions $\hat{\phi} : \mathbb{X}_S \to \mathbb{Z}$ for example 2 (eq. 14), as computed with SKA using random sampling. Color indicates function value.

Next, we turn again to regular sampling for computational convenience. We use the same setup as in the previous example. For the previous and this example we discussed shortly the influence of the interpolation method in the case of random sampling, now we will discuss the influence of the the interpolation method in the case of regular sampling. In appendix 14 we show the eigenfunction accuracy criterion for various interpolation techniques (basic Matlab routines). The results suggest two things: the interpolation error is insignificant if the sampling resolution is sufficiently high, and the interpolation method only matters at low

index	eigenvalues	modes	α_{linear}	$\alpha_{natural}$	α_{biharm}	α_{loess2}
1	$9.00e-01$	$1.09e+00$	$6.71e-04$	$6.59e-04$	$6.49e-04$	$1.46e-03$
2	$8.10e-01$	$1.11e+00$	$1.73e-02$	$1.73e-02$	$5.51e-03$	$8.06e-03$
3	$8.00e-01$	$1.83e+00$	$9.89e-03$	$9.89e-03$	$3.19e-03$	$4.87e-03$
4	$7.23e-01$	$7.80e-02$	$6.49e-02$	$6.53e-02$	$5.97e-02$	$1.82e-02$
5	$6.74e-01$	$1.06e-01$	$9.66e-02$	$9.71e-02$	$8.37e-02$	$7.94e-02$

Table 3: The out of sample performance of the approximated SKA eigenfunction for example 2 (eq. 14), by interpolation from S, for random sampling, using various interpolation methods: linear interpolation, natural interpolation (C^1), biharmonic interpolation, and a Lowess quadratic non-parametric regression. Under the header of 'modes' we have also reported the norm of the mode.

sampling resolution. Both statements follow from the observation that the accuracy criterion appears to be independent of sampling resolution and interpolation method, if the resolution is sufficiently high. In the appendix we also provide the performance of EDMD with the library of observables as in the previous example. For EDMD all eigenfunction have quality $\alpha < 10^{-3}$, whereas for SKA the performance is worse, multiple approximate eigenfunctions have $\alpha \approx 10^{-1}$.

In figure 7 we have plotted β as a function of time for both sampling densities, for SKA using linear interpolation. The sampling resolution has only a minor effect. The difference between SKA and EDMD is again significant. Note that the system under study has only one equilibrium point, which is stable, therefore the all trajectories converge to zero. Apparently the error in the SKA mode decomposition goes to zero at the same rate as the trajectories them selves, while for EDMD the error goes less quickly to zero than the trajectories them selves.

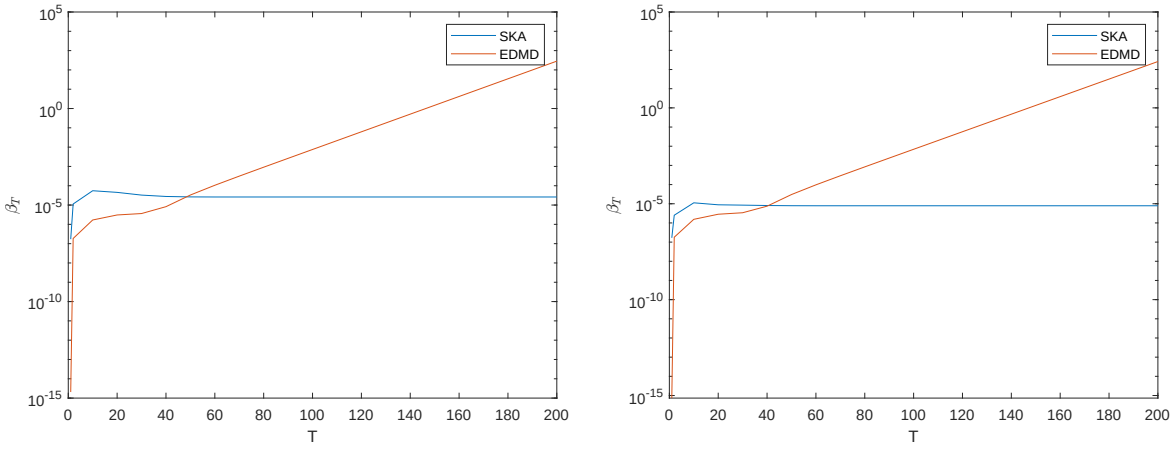

Figure 7: β for example 2 (eq. 14), for both EDMD and SKA. Left uses a training data set of $32 \times 32 = 1024$ points, while the right side is with $71 \times 71 = 5041$ points.

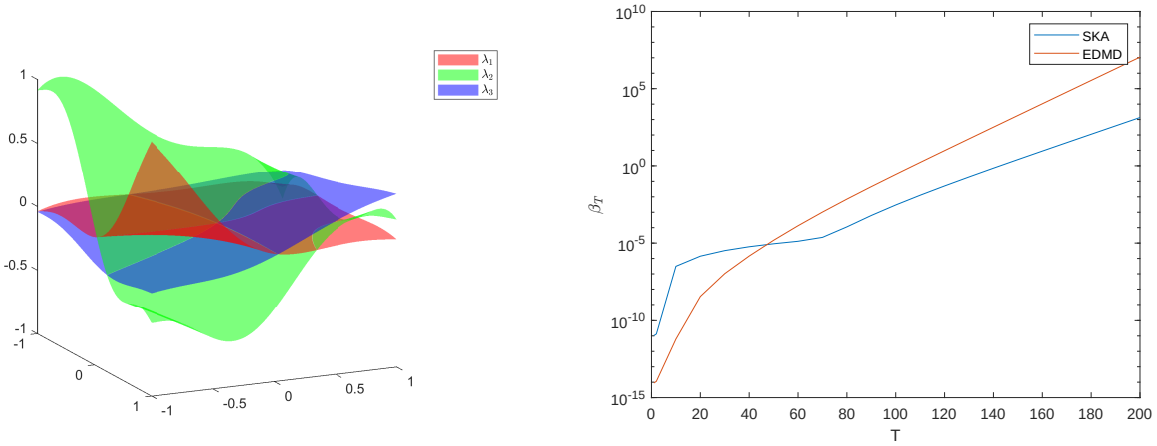

Figure 8: On the left, the map $S : \mathbb{X}_S \to \mathbb{Z}$ for example 3 (eq. 15), as computed with SKA using random sampling, restricted to the first three coordinates of Z. The horizontal plane is the state space. Eigenvalue indexes correspond to table 4. On the right β for example 3, (eq. 15), for both EDMD and SKA.

Population dynamics

Next we consider an example for which we know there exists no finite Koopman invariant subspace which includes the identity function. We look at the well known Susceptible-Infected-(Recovered) model from epidemiology, which has two equilibria, one stable and one unstable, and hence no finite dimensional Koopman invariant subspace which includes the identity function. The recovered class is not modeled since it does not affect the dynamics. Let $f : \mathbb{X} \to \mathbb{X}$, with $\mathbb{X} = [-1, 1] \times [-1, 1]$ be given by

$$f(x) = \begin{pmatrix} x_1 - \mu x_1 x_2 \\ x_2 + \mu x_1 x_2 - \eta x_2 \end{pmatrix}, \tag{15}$$

with $\mu = 0.01$ and $\eta = 0.003$. We again use the same setup, with the exception that we pick $100 \times 100 = 10000$ grid points to be on the safe side. In particular for EDMD we use the same library of observables, a library of monomials. We justify this choice by the fact that f is a polynomial. In figure 8 we show a 3D plot of the first three approximate SKA eigenfunctions and the mode decomposition performance β of both SKA and EDMD. Furthermore, in figure 9 we show heat maps of the first five approximate SKA eigenfunctions. In table 4 and 9 (in appendix 14) we provide the eigenfunction performance of SKA, respectively EDMD. The eigenfunction accuracy of EDMD is again better than that of SKA, by at least one order of magnitude, up to five orders of magnitude. On other hand in figure 8 it can be observed that SKA outperforms EDMD as it comes to mode decompostion accuracy β in the mid- and long-term. Consider for example $T = 80$, the difference between EDMD and SKA is almost 10^2, in favour of SKA. For $T > 100$ the performance of both mode decompositions is $\beta > 10^{-2}$. Therefore for $T > 100$ the mode decompositions become useless. The reason that the errors keep growing is because the linear system is unstable; there is an eigenvalue greater than 1.

index	eigenvalues	modes	α_{linear}	α_{cubic}	α_{spline}
1	$1.02e+00$	$4.65e-01$	$5.80e-03$	$5.80e-03$	$5.80e-03$
2	$1.00e+00$	$2.38e+00$	$3.53e-03$	$3.54e-03$	$3.54e-03$
3	$9.97e-01$	$2.71e+00$	$5.47e-03$	$5.47e-03$	$5.47e-03$
4	$9.97e-01$	$2.71e+00$	$5.47e-03$	$5.47e-03$	$5.47e-03$
5	$9.88e-01$	$1.01e-01$	$1.64e-02$	$1.64e-02$	$1.64e-02$
6	$9.88e-01$	$1.01e-01$	$1.64e-02$	$1.64e-02$	$1.64e-02$
7	$9.81e-01$	$5.84e-01$	$1.18e-02$	$1.18e-02$	$1.18e-02$
8	$9.40e-01$	$3.08e-02$	$1.42e-02$	$1.42e-02$	$1.42e-02$
9	$1.13e+00$	$1.31e-05$	$5.67e-02$	$5.66e-02$	$5.66e-02$
10	$1.06e+00$	$9.50e-03$	$1.16e-02$	$1.15e-02$	$1.15e-02$

Table 4: The out of sample performance of the approximated SKA eigenfunction for example 3 (eq. 15), by interpolation from S for regular sampling with $N = 10000$, using various interpolation methods: linear, cubic, and spline interpolation. Under the header of 'modes' we have also reported the norm of the mode. The eigenfunctions with index 9 and 10 contribute approximately zero to the mode decomposition.

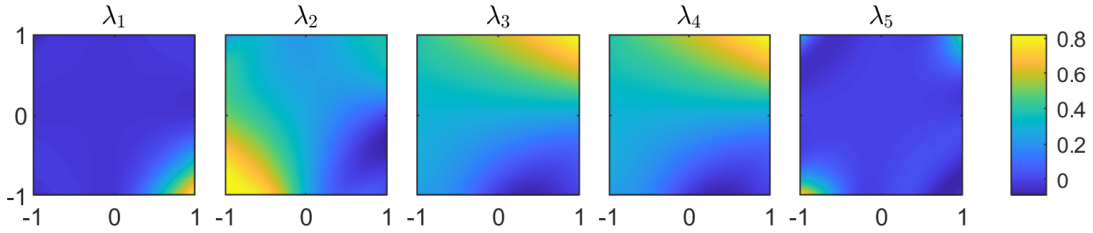

Figure 9: The approximate SKA eigenfunctions $\hat{\phi} : \mathbb{X}_S \to \mathbb{Z}$ for example 3 (eq. 15), as computed with SKA using random sampling. Color indicates function value.

Lorenz system

As a last example we consider the four dimensional Lorenz-96 system given by

$$f(x) = \begin{pmatrix} x_1 + \delta[(x_2 - x_3)x_4 - x_1 + 8] \\ x_2 + \delta[(x_3 - x_4)x_1 - x_2 + 8] \\ x_3 + \delta[(x_4 - x_1)x_2 - x_3 + 8] \\ x_4 + \delta[(x_1 - x_2)x_3 - x_4 + 8] \end{pmatrix}. \tag{16}$$

We set $\delta = 0.005$. We looked at several time series and found that the system almost never leaves the box $[-10, 10]^4$, so we take this as the state space. We apply SKA in various setups: $N = 10$, $M = 10.000$; $N = 20$, $M = 10.000$; $N = 60$, $M = 10.000$; and $N = 20$ $M = 1.000.000$. In all cases all the eigenvalues lie in $(0.95, 1)$, and the eigenfunction accuracy is of order between 10^{-2} and 10^{-1}. Our main interest is the mode accuracy performance β. In 10 we show the performance for all SKA setups

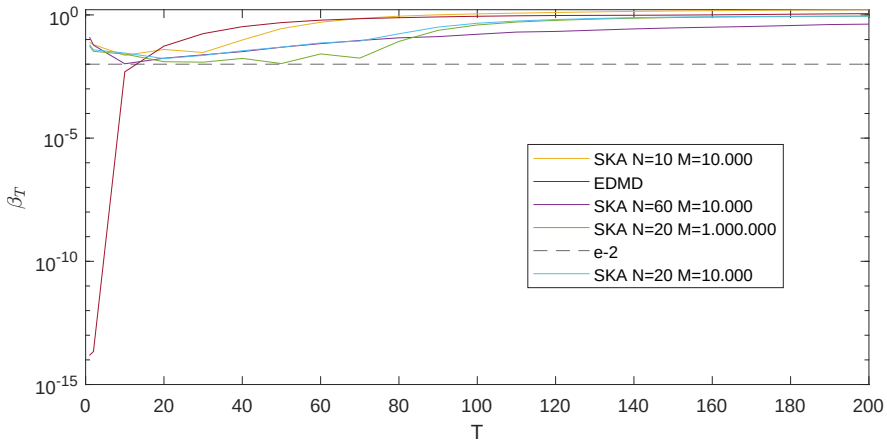

Figure 10: The quality of the mode decomposition for various SKA setups and EDMD, for example 4 (eq. 16).

and an EDMD application. For EDMD we use the library of monomials $x_1^i x_2^j x_3^k x_4^l$, for $i, j, k, l = 0, 1, 2$. We do not include higher order monomials for computational reasons.

We make three observations. First of all for small T SKA performs very poorly relative to EDMD and previous examples. One possible explanation is the following. SKA minimizes the error over the entire length of the burst, so the first time instance in not different then the $4N$'th. The Lorenz system is know to be chaotic, it is hard to predict, hence the errors are all relatively large. Therefore for small T the error is approximately constant and relatively large. Why is it then that in the other examples we did observe a relative low error at the first time step? This might be attributed to interpolation. The interpolation introduced another error, which does grow in time. For the Lorenz example the interpolation error is insignificant in comparison to the error introduces by the linear system approximation, so we do not observe an initial time dependence of the error. On the other hand for the other examples, it might be that the interpolation error is significant. EDMD works only with one time-step, so it is to be expected that for very small time it yields the best results.

Secondly, for $20 < T < 100$ there appears to be a dependence on the sample resolution. For $n = 20$, tripling the resolution in each direction, that is increasing the amount of samples from 10.000 to $1.000.000$ reduces the error from from approximately 10^{-1} to maximally 10^{-2}. Average improvement is less than one order of magnitude.

Thirdly, for $T > 100$ we see no dependency on the sample resolution, but instead on n. However at this point all errors are greater than 10^{-1} and hence the mode decompositions are too poor to be of use.

7 Discussion

In this thesis we have investigated how linear subspace methods might be used as an alternative algorithm for computing a Koopman mode decomposition. The algorithm is dubbed Subspace Koopman Algorithm (SKA), and attempts to find an almost Koopman invariant subspace which includes the identity, in the span of delay coordinate functions.

In the theory section we proved that if a dynamical system f has a finite dimensional Koopman invariant subspace which includes the identity, then SKA is able to find all eigenfunctions and modes exactly. Furthermore, we gave the outlines of a possible proof of convergence of SKA to the actual Koopman operator in case f has no finite dimensional Koopman invariant subspace which includes the identity.

In the results section we showed that SKA is indeed capable of finding the eigenfunction and modes for an example with a finite dimensional Koopman invariant subspace which includes the identity. We then considered a bunch of examples for which there is no finite dimensional Koopman invariant subspace which includes the identity. We found that SKA performs worse than EDMD if it comes to approximating eigenfunctions. However, in most cases SKA outperforms EDMD if it comes to mode decomposition accuracy, for which we introduced a new criterion (eq. 12).

We conclude that SKA is a promising algorithm for Koopman mode decomposition, especially for systems where the dynamics are not known explicitly. SKA is able to reconstruct approximate eigenfunctions in the span of delay coordinates, without relying on explicit equations. For dynamical systems with a finite dimensional Koopman invariant subspace, SKA is able to identify all eigenfunctions and modes.

In the future the authors wish to investigate whether it is possible to extend SKA with a library of non-observables, and whether it is possible to abuse the optimization solving nature of SKA to find sparse almost Koopman invariant subspaces.

Also, future research should address how the ideas developed here for discrete-time dynamical systems arch over to continuous-time, differential systems.

Part II
Parameter Estimation for Nonlinear Discrete Time Dynamical Systems using the Koopman Operator Theory

8 Introduction

The estimation of model parameters from data, also called system identification, for sophisticated models is an essential step in the use of models in practise. The usage of agent based models (ABMs) to forecast the economy by actual economic actors, such as central banks, is a prime example in which the author is interested. These models can be categorized as stochastic discrete-time dynamical systems, of which only partial state measurements are available. This class of models will be the focus of this study.

The existing methods for parameter estimation for stochastic differential equations can be distinguished in two categories: maximum likelihood based, or objective-function based [1]. This distinction also appears to hold up in the case of estimation for ABMs [2]. This thesis focusses on objective-function based methods. An important feature of all standard estimation methods for large models is the fact that they are simulation based. This is a consequence of it being a hard job to arrive at analytical expression for solutions of complex models. However, simulation based estimation has two drawbacks: computing trajectories of large models can be computationally expensive, and furthermore, using a finite number of simulations/trajectories always introduces a sampling error.

In light of this we are motivated to find objective functions that can be calculated directly, both from data as well as from an analytical description of the model, without the need of simulation of entire trajectories.

In the last twenty years there has been enormous interest in Koopman operator theory. Having been developed in the 30's by Koopman [3], it has gained traction only recently. Koopman operator theory establishes that for a general discrete-time dynamical system $f : \mathbb{X} \to \mathbb{X}$, and an observable $g : \mathbb{X} \to \mathbb{C}$ on the state space, the map $g \to g \circ f$, is a *linear* map on a function space of observables. Because the mathematical toolbox for linear systems is considerably larger than the non-linear mathematical toolbox, Koopman operator theory opens a promising pathway to studying non-linear dynamical systems.

Other interesting properties of the Koopman operator include: the fact that, up to a homeomorphism on the state space, there is a bijection between the Koopman operator and the dynamics f; the possibility of writing a finite dimensional projection of the Koopman operator in matrix form; the availability of algorithms to estimate the Koopman operator directly from data; and the possibility of relating an expression of the functional f to the Koopman operator.

The presence of all these properties makes the Koopman operator theory framework interesting for parameter estimation. The promise is that we can estimate the Koopman operator directly from data, and use the bijective relation between the Koopman operator and f to determine the dynamics f, and thereby estimate the parameters.

Currently, several authors are investigating these possibilities. Known to the author of this thesis are the following works. In [4] and [5] Koopman theory is used to estimate parameters of continous time dynamical system, which are linear in the parameters, and for which full state measurements are available. The possibilities of using Koopman theory for parameter estimation in stochastic differential equations are developed in [6]. Both methods that are developed in these works rely on the connection between the finite time Koopman operator and the infinitesimal generator of the dynamics. This limits these methods to differential equations, and prevents their application to discrete-time systems. Furthermore, both methods assume the availability of full state measurements.

For discrete-time system identification in presence of full state measurements, the SINDy method has been developed in [7]. This method is, however, limited to usage with full state measurements and requires linearity in the parameters.

This leaves open the quest for a Koopman based identification method for discrete-time systems with partial measurements. Inspired by Takens embedding theorem [8], in this thesis we will, first of all, investigate the possibilities of including delay coordinates into the finite dimensional Koopman projection in order to retrieve information about the entire system and enable the estimation of all system parameters from partial measurements, for deterministic systems. We hope that in the future our contribution, combined with the work in [6] on estimation for stochastic systems, will lead to a simulation-free parameter estimation method for discrete-time stochastic dynamical systems from partial measurements.

We show that including delay coordinates introduces a problem with the computation of the Galerkin approximation. To overcome this difficulty, we propose to approximate the discrete empirical measure with a continous kernel distribution. This constitutes the first contribution of this thesis.

The second contribution made in this thesis is the proposal of polynomial approximation of the dynamics as a means to reduce the computational cost of the optimization problem.

We will now formally state the partial-measurement parameter estimation problem we are studying. Let $f : \mathbb{X} \to \mathbb{X}$, and $h : \mathbb{X} \to \mathbb{O}$, with $\mathbb{X} \subseteq \mathbb{R}^n$ the state space and $\mathbb{O} \subseteq \mathbb{R}^m$ the output space, define a discrete-time dynamical system:

$$\begin{aligned} x_{t+1} &= f(x_t) \\ y_t &= h(x_t). \end{aligned} \tag{17}$$

We consider the case in which h is known and f is known up to a set of parameters Θ, such that $f = f_{\tilde{\Theta}}$ for known f_Θ and unknown $\tilde{\Theta} \in \mathbb{Q}$, for some box $\mathbb{Q} \subset \mathbb{R}^q$. We assume Θ determines f_Θ uniquely, so there exist no Θ_1, Θ_2 such that $\Theta_1 \neq \Theta_2$ and $f_{\Theta_1} = f_{\Theta_2}$. Our goal is to develop a method which, given a *time-series* of output measurements $\{y_t\}_{t=0}^M$, provides an estimate $\hat{\Theta}$ such that $\hat{\Theta} \approx \tilde{\Theta}$.

The rest of this thesis is organized as follows. We will first discuss in section 9 a Galerkin approximation approach of the Koopman operator; Koopman based identification; and delay coordinates. Subsequently, in section 10 we will discuss the difficulties arising in computing the inner products when using delay coordinates. In section 11 we describe our proposed method and in 12 we discuss an example. In section 13 we will propose polynomial approximation in order to reduce computational complexity.

9 Fundamentals

Galerkin Approximation

As discussed in [10] and [12] the operator $\mathcal{K}_N^\mu$ is in fact a Galerkin approximation of $\mathcal{K}$: it is the orthogonal projection of $\mathcal{K}$ onto a finite subset of elements, or put differently $\mathcal{K}_N(g) = \text{argmin}_{g^* \in \mathbb{G}_N} \|g^* - \mathcal{K}(g)\|$. The matrix K_N^μ then consists of inner products in the following way:

$$K_N^\mu = (G^\mu)^{-1} A^\mu,$$

with elements $A_{ij}^\mu := \langle \psi_i, \mathcal{K}\psi_j \rangle_\mu$ and $G_{ij}^\mu := \langle \psi_i, \psi_j \rangle_\mu$. Remember that the inner product is given by $\langle f, g \rangle_\mu = \int_{\mathbb{X}} f \bar{g} \, \mathrm{d}\mu$. Given a model of a dynamical system and a measure, one can compute the inner products analytically or by numerical integration. We will refer to this approach as the Galerkin approach. Where EDMD relies on data, this approach relies on a model.

Identification

In line with [6] we will focus on identification in the projected Koopman space. We define $K^\mu(\Theta) := (G^\mu)^{-1} A^\mu(\Theta)$ with $A^\mu(\Theta)_{ij} = \langle \psi_i, \psi_j \circ f_\Theta \rangle_\mu$. Recall that we fixed N. The identification then reduces to finding an optimal estimate $\hat{\Theta}$ by computing

$$\hat{\Theta} = \text{argmin}_{\Theta \in \mathbb{Q}} \|K^\mu(\Theta) - K_{\mathrm{D}}^\rho\|_F. \tag{18}$$

Here F denotes the Frobenius norm. We use the Frobenius norm instead of the 2-norm because in [6] it is shown that it makes little difference in practise, but is computationally beneficial. Since inverting G^μ is numerically unstable, we instead propose to look at:

$$\hat{\Theta} = \text{argmin}_{\Theta \in \mathbb{Q}} \|A^\mu(\Theta) - G^\mu K_{\mathrm{D}}^\rho\|_F, \tag{19}$$

for μ and ρ corresponding to each other. What this *correspondence* entails will be discussed in section 10. The local minima of the objective function of equation 18 and 19 are the same. However, the global optimum may be different. So depending on the optimization requirements, one still needs to compute $\|K^\mu(\Theta) - K_{\mathrm{D}}^\rho\|_F$ for all local minima of $\|A^\mu(\Theta) - G^\mu K_{\mathrm{D}}^\rho\|_F$.

Delay Embedding

This section and the next sections contain the main contributions of this thesis. If only partial measurements are available, and one does not use delay coordinates, we can only calculate the EDMD with respect to observables of the form $\psi = \phi \circ h$. For many combination of f_Θ and h, the composition $h \circ f_\Theta$ will be independent of some of the parameters Θ. Hence, $A^\mu(\Theta)_{ij} = \langle \phi_i \circ h, \phi_j \circ h \circ f_\Theta \rangle$ will be independent of some parameters Θ, and it will not be possible to estimate these parameters. A simple example which shows this is the following two dimensional map $f_\Theta : \mathbb{R}^2 \to \mathbb{R}^2$ with $h : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$f_\Theta : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 + x_2 \\ \theta x_1 x_2 \end{pmatrix}, \qquad h : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to x_1.$$

It is easy to see that $(h \circ f_\Theta)(x) = x_1 + x_2$ is independent of Θ, and hence $A^\mu(\Theta)$ is independent of Θ. Surely, any system which is not observable for some unobserved states will also suffer from this problem.

For system which are observable, we can circumvent this issue by considering a delay embedding. Let the Takens map $H : \mathbb{X} \to \mathbb{R}^{n+1}$ be given by

$$H_\Theta := [h, h \circ f_\Theta, h \circ f_\Theta^2, ..., h \circ f_\Theta^{2n}],$$

where $f^k := f \circ f \circ f ...$, k times. If the dependency on Θ is irrelevant, or if Θ is fixed, we will write just H. We define the delay space $\mathbb{D}$ by $\mathbb{D} = H(\mathbb{X})$. Takens theorem [8] states that for a large class of dynamical systems, for each dynamical system f, there exists a homeomorphic dynamical system $f^* : \mathbb{D} \to \mathbb{D}$ such that $H^{-1} \circ f^* \circ H = f$ and H is a bijection. As is commonly done, and motivated by Takens theorem, we propose to append our dataset with delay coordinates in order to capture all dynamics.

Let τ be the number of delays we include. We redefine H to be a map $H : \mathbb{X} \to \mathbb{D} \subseteq \mathbb{R}^\tau$ given by $H := [h, h \circ f, h \circ f^2, ..., h \circ f^\tau]$ and write the delay coordinates as:

$$\tilde{y}_t = \begin{pmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+\tau} \end{pmatrix} = H(y_t), \tag{20}$$

for $\{y_t\}_{t=0}^M$ a time-series of output measurements.

For τ sufficiently large $H \circ f_\Theta$ depends on all Θ, and hence for observables of the form $\psi : \mathbb{X} \to \mathbb{C}$, given by $\psi = \phi \circ H$, for $\phi : \mathbb{D} \to \mathbb{C}$, we have that $A^\mu(\Theta)_{ij}$ depends on all Θ.

For the just-mentioned example we would have

$$H(x) = \begin{pmatrix} x_1 \\ x_1 + x_2 \end{pmatrix}, \qquad (H \circ f_\Theta)(x) = \begin{pmatrix} x_1 + x_2 \\ x_2 + x_2 + \theta x_1 x_2 \end{pmatrix},$$

and we note that for many observables of the form $\psi = \phi \circ H$, $A^\mu(\Theta)_{ij} = \langle \psi_i, \psi_i \circ f_\Theta \rangle_\mu$ depends on Θ.

Using a full Takens embedding, that is using $2n+1$ delays, ensures that the dynamics in the delay space are homeomorphic to the dynamics in the state space. On first thought, it appears that we only require something weaker; that the analytical expression for $K(\Theta)$ depends on all parameters Θ. However, on second thought, we need that H is a bijection such that the inner products with respect to an empirical measure on the delay space are properly defined. This will become apparent later.

10 On Measures in Inner Product Computation

The identification method as defined by equation 19 appears to be pretty straightforward. However, it rests on the ability to compute $K^{\mu}(\Theta)$ for μ corresponding to the empirical measure ρ. In this section we will first discuss what this correspondence entails in the case no delay embedding is used, and then for the interesting case where a delay embedding is used.

No delay embedding

For full state measurements we can just regard h to be the identity map. In this case the output space equals the state space, $\mathbb{O} = \mathbb{X}$. Let μ be some measure on the state space. We consider a time-series of consecutive state measurements $x_1, x_2, ..., x_M$, and $x_1^{\#}, x_2^{\#}, ..., x_M^{\#}$ satisfying $x_i^{\#} = f(x_i)$. So, $x_{t+1} = x_t^{\#}$. Let $\{\psi_k\}_{k=1}^{N}$, $\psi_i : \mathbb{X} \to \mathbb{C}$ be a library of linearly independent polynomials, with $\Psi : \mathbb{X} \to \mathbb{C}^N$ the shorthand notation as defined earlier. In order for equation 19 to make sense $K^{\mu}(\Theta)$ and K_{D}^{ρ} need to be computed with respect to the same, or approximately the same, inner product. There are three options to do so.

- We can use the empirical measure ρ directly, and compute $K^{\rho}(\Theta)$. Using the simple but crucial identity

$$\int_{\mathbb{X}} r \, \mathrm{d}\rho = \sum_{t=1}^{M} r(x_t), \tag{21}$$

  the integrals in the Galerkin inner products reduce to a finite sum, which is computationally very straightforward. This is the most obvious approach and is also followed in [6].

- If we can assume that the data series correspond to a measure with a continuous density function, we can estimate the density from our samples, and use it to scale the observations with the density in calculating the EDMD. This is related to importance sampling in Monte Carlo integration. By compensating for the distribution of the data, we are calculating the EDMD with respect to the uniform distribution. The condition that time-series data converge to a continuous distribution implies that the system needs to be ergodic.

- Instead of using the empirical measure directly, we can use a continuous approximation of the (generalized) density function to approximate the Galerkin inner product with respect to that measure by evaluating $\int_{\mathbb{X}} g\bar{h}\,\mathrm{d}\hat{\rho}$, where $\hat{\rho}$ is the estimated density function.

The first approach is exact and easy, so it is the preferred option. When using a delay embedding this option becomes intractable, as we will argue in the next section, and we need to consider the other options.

Delay embedding

We will now consider the case where one has no full state measurements, and needs to use delay coordinates in order to be able to estimate Θ completely. We will show that it is not possible to use the standard, first approach as listed above, and argue that it necessary to resort to the third option of continuous approximation of the density function.

Given the time-series of output measurements we construct the delay coordinates as defined in equation 20. Let the functions ϕ_i be observables on the delay space, $\phi_i : \mathbb{D} \to \mathbb{C}$, and let π be the empirical measure on the *delay space*. We write $\psi_i = \phi_i \circ H$ for the observables on the state space. The empirical measure on the state space ρ is unknown in general.

Applying EDMD on the delayed coordinates amounts to computing K_{D}^{ρ}. The corresponding Galerkin inner products are

$$\langle \psi_i, \mathcal{K}\psi_j \rangle_\rho = \int_{\mathbb{X}} \psi_i \cdot (\psi_j \circ f)\,\mathrm{d}\rho(x) = \int_{\mathbb{X}} \psi_i \cdot (\psi_j \circ f)\,\mathrm{d}\pi(H(x)). \tag{22}$$

Where we require that H is a bijection. Otherwise the right side would not be bounded.

Since ρ is unknown in general we cannot evaluate $\int_{\mathbb{X}} \psi_i \cdot (\psi_j \circ f)\,\mathrm{d}\rho(x)$, however if we attempt to evaluate $\int_{\mathbb{X}} \psi_i \cdot (\psi_j \circ f)\,\mathrm{d}\pi(H(x))$ we get into troubles as well. A relation such as equation 21 does not exist for the more complicated expression we now have.

Remark 2. If sufficiently many delay coordinates are included, and the conditions of Takens theorem are met, the map H will be invertible and there exists a homeomorphic dynamical system $f^* : \mathbb{D} \to \mathbb{D}$. Suppose we know explicitly $H^{-1} : \mathbb{D} \to \mathbb{X}$. In this case we know ρ and we are in fact able to calculate $\int_{\mathbb{X}} \psi_i \cdot (\psi_j \circ f)\,\mathrm{d}\rho(x)$. However, even though H is invertible, its inverse might not be available explicitly. Consider for example higher order polynomials, for these there is no expression for the inverse. Although no explicit expression can be obtained, it is still possible to calculate $H_\Theta^{-1}(y)$ for given y and Θ by solving $H_\Theta(x) = y$ for x. For many dynamical systems, such exercises are computationally expensive, and we expect that for most combinations of dynamical systems and optimization methods the computational costs will be too high. Therefore in most cases we can not work on the delay space. In the results section we treat one example where it is possible to write down f^* explicitly.

Let us now consider the various options to obtain $K^{\mu}(\Theta)$ and K_D^{ρ}, with corresponding inner products. As we have just seen, we can not directly compute $\langle \psi_i, \mathcal{K}\psi_j \rangle_{\rho}$. Therefore, the first approach as listed in the previous section is not applicable in general in the case of delay coordinates. Furthermore, ABMs are probably not ergodic [2], and neither is the actual economy, hence we can not resort to the second approach.[3]

Assuming the three options we considered are exhaustive, this leaves us with the last option. We propose to approximate the density function π continuously as a way to make the integral in equation 22 tractable. In case the system is non-ergodic the empirical measure of the data does not converge to a specific distribution. Nevertheless, we can still approximate the empirical measure directly using a kernel based approach. Denote by $\hat{\pi}$ some continuous approximation of π, and write $\hat{\rho}$ for the measure such that $\mathrm{d}\hat{\rho}(x) = \mathrm{d}\hat{\pi}(H(x))$.

We propose to estimate the parameters in a delay embedding by

$$\hat{\Theta} = \mathrm{argmin}_{\Theta \in \mathbb{Q}} \| A^{\hat{\rho}}(\Theta) - G^{\hat{\rho}} K_{\mathrm{D}}^{\rho} \|_F, \tag{23}$$

In the next section we will elaborate more on the estimation method, including an approach to kernel density estimation. We finish this section by developing some intuition as to why such an approximation would yield reasonable results.

Consider the simple case of only one sample, then $\pi = \delta_{d_0}$ for some $d_0 \in \mathbb{D}$. Let us approximate π continuously using some kernel function with bounded support, such that $\hat{\pi}(y) = 0$ for $\{y \in \mathbb{D} \mid |y - d_0| > \Delta\} =: \mathbb{D}_0$ for some small Δ. We abuse notation slightly by treating $\hat{\pi}$ as a probability density function on the delay space, while we define it as a probability measure. This should not lead to confusion, since $\hat{\pi}$ has by definition a continuous probability density function which defines a probability measure in the obvious way.

Assuming some functions $r : \mathbb{X} \to \mathbb{R}$ and $H : \mathbb{X} \to \mathbb{D}$ are sufficiently smooth and Δ is sufficiently small, then $r \circ H^{-1}$ is approximately constant on $\mathbb{D}_0$, and it follows that:

$$\begin{aligned}
&\int_{\mathbb{X}} r(x)\, \mathrm{d}\pi(H(x)) = \int_{\mathbb{D}} r \circ H^{-1}\, \mathrm{d}\pi \\
&= r(H^{-1}(d_0)) = \int_{\mathbb{D}_0} r(H^{-1}(d_0))\hat{\pi}(y)\, \mathrm{d}y \\
&\approx \int_{\mathbb{D}_0} r(H^{-1}(y))\hat{\pi}(y)\, \mathrm{d}y = \int_{\mathbb{D}} r \circ H^{-1}\, \mathrm{d}\hat{\pi} \\
&= \int_{\mathbb{X}} r(x)\, \mathrm{d}\hat{\pi}(H(x)).
\end{aligned}$$

This shows that we can approximate the empirical measure of a single data point with a continuous distribution. Kernel density approximation places a con-

[3] Even in the case the system would be ergodic, the time scale on which time-series data would converge to a space average would be far longer then the time scale on which the economy can be considered to be an autonomous dynamical system.

tinuous distribution at each data point, hence the above intuition generalizes for a multitude of points by summation over all data points.

11 Method

We propose an algorithm for parameter identification from partial measurements using Koopman operator theory. Consider a dynamical system as in equation 17. Assume access to a time-series of output measurements $\{y_t\}_{t=1}^M$. Denote by π the empirical measure on the delay space, associated with this time-series. Let $H : \mathbb{X} \to \mathbb{D} \subseteq \mathbb{R}^\tau$ be the delay map, for minimal τ such that H depends on all Θ. And let $\phi_i : \mathbb{D} \to \mathbb{C}$, $i = 1, 2, ..., N$ be a library of N observables on the delay space. The algorithm is the following.

- Calculate an approximation of the Koopman operator from data using EDMD. Denote the matrix representation by K_D^ρ.
- Continuously approximate π, for example by kernel density approximation. Denote the results as $\hat{\pi}$. Call the associated measure on the state space $\hat{\rho}$.
- Solve the optimization problem as defined in equation 23.

Continuous Approximation of empirical measure

A well develop approach to estimation of probability densities is *kernel density approximation* [13]. A kernel is a function $k : \mathbb{D} \to \mathbb{R}$ that is non-negative and bounded with Lebesgue measure equal to one. Kernel density approximation places a kernel function at each data point of a data set. For the time-series $\{y_t\}_{t=1}^M$ we obtain an approximation of the density π as

$$\hat{\pi}(x) = \frac{1}{hM} \sum_{t=1}^{M} k\left(\frac{x - y_t}{h}\right),$$

for h a smoothing parameter.

12 Results

We leave the investigation into performance of the full algorithm, including the optimization step for future research. That is, we will not investigate the forward error $\hat{\Theta} - \tilde{\Theta}$, instead we will focus on the backward error in the optimization formulation 19. We will study

$$e_{\mu_1,\mu_2}(\Theta) := \|A^{\mu_1}(\Theta) - G^{\mu_1} K_{\mathrm{D}}^{\mu_2}\|_F, \tag{24}$$

especially $e_{\mu_1,\mu_2}(\tilde{\Theta})$, for data generated using parameters $\tilde{\Theta}$.

Illustrative Example

First of all we study the dynamical system $f : \mathbb{R}^2 \to \mathbb{R}^2$ with measurement $h : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f : \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to \begin{pmatrix} x^1 + \theta^1 x^1 x^2 \\ x^2 + \theta^2 x^1 \end{pmatrix} \text{ and } h : \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to x^1. \tag{25}$$

Where we use the superscript to denote elements of a vector, corresponding to $x = [x^1, x^2]^\top$, and reserve the subscript for time indexation. We include one delay coordinate and define $H : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$H : \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to \begin{pmatrix} x^1 \\ x^1 + \theta^1 x^1 x^2 \end{pmatrix}.$$

For this simple example we can obtain an explicit expression for H^{-1}:

$$H^{-1} : \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \to \begin{pmatrix} y^1 \\ -(y^1 - y^2)/(\theta^1 y^1) \end{pmatrix}.$$

Because we know H^{-1} we know the dynamics in the delay space $f^* = H \circ f \circ H^{-1}$, and hence we are able to calculate $\int_{\mathbb{D}} \psi_i \cdot \psi_j \circ f^* \, \mathrm{d}\rho$ directly for the empirical measure and obtain $A^\rho(\Theta)$.

We set $\tilde{\Theta} = [0.001, -0.002]$, and generate time-series data using the true parameters and $x_0 = [1, 2]^\top$ of length 3000. We compute the backward error $e_{\rho,\rho}(\tilde{\Theta})$, as defined in 24, to be $9.4050 \cdot 10^{-13}$. In figure 11 we plot $e_{\rho,\rho}(\Theta)$ for a grid of Θ. We observe that there are multiple points in this parameter space where the error goes to zero. It is easy to show that there are indeed multiple points in the parameter space that yield the same time-series for corresponding initial conditions. We conclude that the method is able to identify the parameters correctly, if the actual empirical measure can be used directly.

The main question is however, whether the method works if we approximate the empirical measure π continuously. Unfortunately, we are not able to report any results that answer this question. For a time-series of length 3000 we attempted to approximate the empirical measure with a kernel distribution using a Gaussian kernel. The computational costs of our implementation were higher than anticipated, and we were not able to compute any integral. In hindsight it is no surprise that computing for each point the summation of 3000 evaluations of a Gaussian function, yields an integral which it computationally (too) costly to numerically integrate.

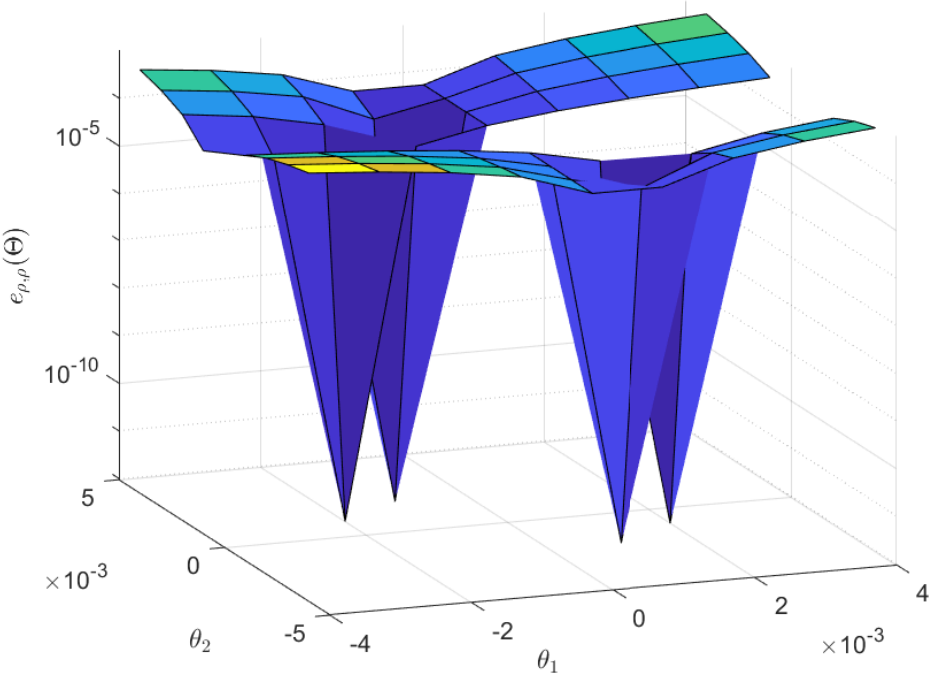

Figure 11: Parameter exploration of dynamical system as defined in 25, for parameters in the box $[-0.004, 0.004] \times [-0.004, 0.004]$.

13 On Computation

The author has considered one possibility of mitigating the computational problem. The idea is to approximate all functions involved as polynomials. This has two benefits. First, in the calculation of A we can separate the Θ dependency from the state dependency, which allows us to calculate the integrals over the state space only once and obtain an explicit polynomial expression for $A(\Theta)$.

The second benefit arises when tackling the optimization problem. We are interested in finding the global minimum over $\mathbb{Q}$ of a polynomial expression. This requires us to find the points for which the gradient is zero. Within the field of algebraic geometry, dedicated numerical algorithms have been developed to find the (real) roots of systems of polynomials.

Polynomial kernels

In order to obtain a final polynomial expression we also need that the continuous approximation of the empirical measure is polynomial. It is certainly possible to use a polynomial kernel [14], but it has to be investigated whether low degree polynomials yield sufficiently accurate approximations.

High degree polynomials

In case the degree of the final polynomial $A(\Theta)$ is too high to find roots in practise, we wish to approximate it by a lower degree polynomial. Since we assume certain knowledge on the parameters, namely that they lie in the box $\mathbb{Q}$, we know that certain terms will be very small. It has to be investigated what impact this has on computational cost and results to leave out such terms.

14 Discussion

In this short text we have developed the idea of including delay coordinates in Koopman identification, as a means to estimate all parameters of a system from partial measurements. This is motivated by Takens embedding theorem. We hope that in the future this work might be combined with the work of [6] to form a simulation-free Koopman based identification method for stochastic systems with partial measurements, such that it might be deployed for parameter estimation of ABMs.

We argued that a challenge arises in the Galerkin approximation, because the integral becomes intractable, and proposed to solve this by continuous approximation of the empirical measure. Unfortunately, it turned out that this is not computationally tractable. We propose to tackle the computational issue by resorting to polynomials, for which dedicated solvers are available. It also remains to be investigated for which category of models it is tractable in general, for sufficiently simple measures, to compute the inner products numerically or symbolically.

References

[1] A. S. Hurn, J. I. Jeisman, and K. A. Lindsay. "Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations". In: *Journal of Financial Econometrics* 5.3 (2007), pp. 390–455. ISSN: 14798417. DOI: 10.1093/jjfinec/nbm009.

[2] Donovan Platt. "A comparison of economic agent-based model calibration methods". In: *Journal of Economic Dynamics and Control* 113 (2020), p. 103859. ISSN: 01651889. DOI: 10.1016/j.jedc.2020.103859. URL: https://doi.org/10.1016/j.jedc.2020.103859.

[3] B.O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: *Proc Natl Acad Sci U S A.* 17 (1931), pp. 315–318.

[4] Alexandre Mauroy and Jorge Goncalves. "Koopman-based lifting techniques for nonlinear systems identification". In: *IEEE Transactions on Automatic Control* 65.6 (2020), pp. 2550–2565. ISSN: 15582523. DOI: 10.1109/TAC.2019.2941433. arXiv: 1709.02003.

[5] Alexandre Mauroy and Jorge Goncalves. "Linear identification of nonlinear systems: A lifting technique based on the Koopman operator". In: *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (2016), pp. 6500–6505. DOI: 10.1109/CDC.2016.7799269. arXiv: 1605.04457.

[6] Asbjorn N. Riseth and Jake P. Taylor-King. "Operator fitting for parameter estimation of stochastic differential equations". In: *arXiv* (2017), pp. 1–21. ISSN: 23318422. arXiv: 1709.05153.

[7] Steven L. Brunton et al. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.15 (2016), pp. 3932–3937. ISSN: 10916490. DOI: 10.1073/pnas.1517384113. arXiv: 1509.03580.

[8] Lyle Noakes. "The Takens Embedding Theorem". In: *International Journal of Bifurcation and Chaos* 1.4 (1991), pp. 867–872.

[9] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: *Journal of Nonlinear Science* 28.2 (2018), pp. 687–710. ISSN: 14321467. DOI: 10.1007/s00332-017-9423-0. arXiv: 1703.04680.

[10] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. "A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition". In: *Journal of Nonlinear Science* 25.6 (2015), pp. 1307–1346. ISSN: 14321467. DOI: 10.1007/s00332-015-9258-5. arXiv: 1408.4408.

[11] Stefan Klus, Péter Koltai, and Christof Schütte. "On the numerical approximation of the Perron-Frobenius and Koopman operator". In: *Journal of Computational Dynamics* 3.1 (2016), pp. 51–79. ISSN: 21582505. DOI: 10.3934/jcd.2016003. arXiv: 1512.05997.

[12] Adam L. Bruce, Vera M. Zeidan, and Dennis S. Bernstein. "What is the Koopman operator? A simplified treatment for discrete-time systems". In: *Proceedings of the American Control Conference* 2019-July (2019), pp. 1912–1917. ISSN: 07431619. DOI: 10.23919/acc.2019.8814306.

[13] Stanislaw Węglarczyk. "Kernel density estimation and its application". In: *ITM Web Conference, XLVIII Seminar of Applied Mathematics* 23 (2018). DOI: 10.1051/itmconf/20182300037.

[14] Joakim Munkhammar, Lars Mattsson, and Jesper Rydén. "Polynomial probability distribution estimation using the method of moments". In: *PLoS ONE* 12.4 (2017), pp. 1–14. ISSN: 19326203. DOI: 10.1371/journal.pone.0174573.

[15] Stefan Klus et al. "Data-Driven Model Reduction and Transfer Operator Approximation". In: *Journal of Nonlinear Science* 28.3 (2018), pp. 985–1010. ISSN: 1432-1467. DOI: 10.1007/s00332-017-9437-7. URL: https://doi.org/10.1007/s00332-017-9437-7.

[16] Qianxiao Li et al. "Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator". In: *Chaos* 27.10 (2017), pp. 1–25. ISSN: 10541500. DOI: 10.1063/1.4993854. arXiv: 1707.00225.

[17] Steven L. Brunton et al. "Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control". In: *PLoS ONE* 11.2 (2016), pp. 1–19. ISSN: 19326203. DOI: 10.1371/journal.pone.0150171. arXiv: 1510.03007.

[18] Masih Haseli and Jorge Cort. "Fast Identification of Koopman-Invariant Subspaces: Parallel Symmetric Subspace Decomposition". In: *2020 American Control Conference (ACC)*. 2020.

[19] Amit Surana et al. "Linear observer synthesis for nonlinear systems using Koopman Operator framework". In: *IPAC PapersOnLine* 49.18 (2016), pp. 716–723. ISSN: 24058963. DOI: 10.1016/j.ifacol.2016.10.250. URL: http://dx.doi.org/10.1016/j.ifacol.2016.10.250.

[20] Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. "Data-driven discovery of Koopman eigenfunctions for control". In: *arXiv* 98195 (2017), pp. 1–40. ISSN: 23318422. arXiv: 1707.01146.

[21] Gary Froyland, Georg A. Gottwald, and Andy Hammerlindl. "A computational method to extract macroscopic variables and their dynamics in multiscale systems". In: *SIAM Journal on Applied Dynamical Systems* 13.4 (2014), pp. 1816–1846. ISSN: 15360040. DOI: 10.1137/130943637. arXiv: 1310.8001.

[22] Peter Van Overschee and Bart De Moor. *Subspace Identification for Linear Systems*. KLUWER ACADEMIC PUBLISHERS, 1996. ISBN: 9781461304654. DOI: 10.1007/978-1-4613-0465-4.

[23] Sungho Shin, Qiugang Lu, and Victor M Zavala. "Unifying Theorems for Subspace Identification and Dynamic Mode Decomposition". In: *arXiv* March (2020). arXiv: arXiv:2003.07410v1.

[24] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. "Subspace dynamic mode decomposition for stochastic Koopman analysis". In: *Physical Review E* 96.3 (2017), pp. 1–11. ISSN: 24700053. DOI: 10.1103/PhysRevE.96.033310. arXiv: 1705.04908.

[25] Christiaan Heij, André Ran, and Freek van Schagen. "Introduction to Mathematical Systems Theory". In: *Introduction to Mathematical Systems Theory* (2007). DOI: 10.1007/978-3-7643-7549-2.

[26] Hassan Arbabi and Igor Mezić. "Ergodic Theory , Dynamic Mode Decomposition , and Computation of Spectral Properties of the Koopman Operator". In: *SIAM Journal on Applied Dynamical Systems* 16.4 (2017), pp. 2096–2126.

[27] Yoshihiko Susuki and Igor Mezić. "A prony approximation of Koopman Mode Decomposition". In: *Proceedings of the IEEE Conference on Decision and Control*. Vol. 54rd IEEE. Cdc. 2015, pp. 7022–7027. ISBN: 9781479978861. DOI: 10.1109/CDC.2015.7403326.

[28] Hao Zhang et al. "Evaluating the accuracy of the dynamic mode decomposition". In: *Journal of Computational Dynamics* 7.1 (2020), pp. 35–56. ISSN: 21582505. DOI: 10.3934/JCD.2020002. arXiv: 1710.00745.

[29] Jia Zhao and Jarrod Mau. "Discovery of Governing Equations with Recursive Deep Neural Networks". In: (2020), pp. 7–9. arXiv: 2009.11500. URL: http://arxiv.org/abs/2009.11500.

Appendix

SKA Example 2: Eigenfunction Accuracy for Various Grid Densities

index	eigenvalues	modes	α_{linear}	α_{cubic}	α_{spline}
1	$9.00e-01$	$1.21e+00$	$1.04e-05$	$1.74e-06$	$1.64e-06$
2	$8.10e-01$	$1.47e+00$	$1.09e-03$	$1.25e-04$	$1.25e-04$
3	$8.00e-01$	$2.13e+00$	$6.34e-04$	$8.09e-05$	$8.09e-05$
4	$7.29e-01$	$1.20e-01$	$6.95e-03$	$6.05e-03$	$6.05e-03$
5	$7.20e-01$	$1.38e-01$	$1.33e-02$	$1.25e-02$	$1.25e-02$
6	$6.56e-01$	$3.94e-01$	$2.21e-01$	$2.21e-01$	$2.21e-01$
7	$6.47e-01$	$3.53e-01$	$2.08e-01$	$2.09e-01$	$2.09e-01$
8	$6.42e-01$	$5.24e-02$	$1.83e-01$	$1.84e-01$	$1.83e-01$
9	$5.74e-01$	$4.89e-02$	$8.74e-02$	$8.80e-02$	$8.80e-02$
10	$5.86e-01$	$8.05e-03$	$1.38e-01$	$1.38e-01$	$1.38e-01$

Table 5: The out of sample performance of the approximated SKA eigenfunction for example 2, (eq. 14), by interpolation from S for regular sampling with $N = 1024$, using various interpolation methods: linear, cubic, and spline interpolation. Under the header of 'modes' we have also reported the norm of the mode.

index	eigenvalues	modes	α_{linear}	α_{cubic}	α_{spline}
1	$9.00e-01$	$1.21e+00$	$3.09e-06$	$1.68e-06$	$1.68e-06$
2	$8.10e-01$	$1.47e+00$	$3.04e-04$	$1.29e-04$	$1.29e-04$
3	$8.00e-01$	$2.13e+00$	$1.81e-04$	$8.22e-05$	$8.22e-05$
4	$7.29e-01$	$1.20e-01$	$6.27e-03$	$6.15e-03$	$6.15e-03$
5	$7.20e-01$	$1.38e-01$	$1.28e-02$	$1.27e-02$	$1.27e-02$
6	$6.56e-01$	$3.94e-01$	$2.21e-01$	$2.22e-01$	$2.22e-01$
7	$6.47e-01$	$3.51e-01$	$2.09e-01$	$2.09e-01$	$2.09e-01$
8	$6.42e-01$	$5.33e-02$	$1.84e-01$	$1.84e-01$	$1.84e-01$
9	$5.74e-01$	$4.92e-02$	$8.84e-02$	$8.86e-02$	$8.86e-02$
10	$5.86e-01$	$7.78e-03$	$1.39e-01$	$1.39e-01$	$1.39e-01$

Table 6: The out of sample performance of the approximated SKA eigenfunction for example 2, (eq. 14), by interpolation from S for regular sampling with $N = 5041$, using various interpolation methods: linear, cubic, and spline interpolation. Under the header of 'modes' we have also reported the norm of the mode.

index	eigenvalues	modes	α_{linear}	α_{cubic}	α_{spline}
1	$9.00e-01$	$1.21e+00$	$1.86e-06$	$1.65e-06$	$1.65e-06$
2	$8.10e-01$	$1.47e+00$	$1.63e-04$	$1.27e-04$	$1.27e-04$
3	$8.00e-01$	$2.13e+00$	$1.02e-04$	$8.13e-05$	$8.13e-05$
4	$7.29e-01$	$1.20e-01$	$6.15e-03$	$6.12e-03$	$6.12e-03$
5	$7.20e-01$	$1.38e-01$	$1.26e-02$	$1.26e-02$	$1.26e-02$
6	$6.56e-01$	$3.94e-01$	$2.22e-01$	$2.22e-01$	$2.22e-01$
7	$6.47e-01$	$3.50e-01$	$2.09e-01$	$2.09e-01$	$2.09e-01$
8	$6.42e-01$	$5.37e-02$	$1.84e-01$	$1.84e-01$	$1.84e-01$
9	$5.74e-01$	$4.94e-02$	$8.79e-02$	$8.79e-02$	$8.79e-02$
10	$5.86e-01$	$7.67e-03$	$1.39e-01$	$1.39e-01$	$1.39e-01$

Table 7: The out of sample performance of the approximated SKA eigenfunction for example 2, (eq. 14), by interpolation from S for regular sampling with $N = 20164$, using various interpolation methods: linear, cubic, and spline interpolation. Under the header of 'modes' we have also reported the norm of the mode.

index	eigenvalues	modes	α
1	$9.00e-01$	$1.21e+00$	$6.74e-07$
2	$8.10e-01$	$1.47e+00$	$9.13e-06$
3	$8.00e-01$	$2.13e+00$	$5.68e-06$
4	$7.29e-01$	$1.20e-01$	$6.87e-05$
5	$7.20e-01$	$1.37e-01$	$5.64e-05$
6	$6.56e-01$	$4.19e-02$	$3.39e-04$
7	$6.49e-01$	$4.22e-02$	$3.01e-04$
8	$6.40e-01$	$1.51e-02$	$1.27e-04$
9	$5.87e-01$	$1.04e-02$	$9.54e-04$
10	$5.77e-01$	$2.64e-02$	$5.21e-04$

Table 8: The out of sample performance of the approximated EDMD eigenfunction for example 2, (eq. 14). Under the header of 'modes' we have also reported the norm of the mode.

SKA Example 3: Eigenfunction Accuracy for Various Grid Densities

index	eigenvalues	modes	α
1	$1.00e+00$	$6.22e+01$	$2.51e-04$
2	$1.00e+00$	$6.22e+01$	$2.51e-04$
3	$9.98e-01$	$1.47e+01$	$1.25e-03$
4	$9.93e-01$	$1.19e+01$	$1.91e-03$
5	$9.93e-01$	$1.19e+01$	$1.91e-03$
6	$9.90e-01$	$3.08e+01$	$1.55e-03$
7	$9.90e-01$	$3.08e+01$	$1.55e-03$
8	$9.90e-01$	$1.28e+01$	$1.98e-03$
9	$9.90e-01$	$1.28e+01$	$1.98e-03$
10	$9.89e-01$	$4.24e+01$	$2.40e-03$

Table 9: The out of sample performance of the approximated EDMD eigenfunction for example 3, (eq. 15). Under the header of 'modes' we have also reported the norm of the mode.