UNIVERSITY OF CALIFORNIA

Santa Barbara

Robust Approximation of the Stochastic Koopman Operator

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Science
in Mechanical Engineering

by

Mathias Thomas Wanner

Committee in Charge
Professor Igor Mezić, Chair
Professor Jeffrey Moehlis
Professor Joã Hespanha

December 2020

The thesis of Mathias Thomas Wanner is approved.

Professor Jeffrey Moehlis

Professor João Hespanha

Professor Igor Mezić (Committee Chair)

December 2020

ABSTRACT

Robust Approximation of the Stochastic Koopman Operator

by

Mathias Thomas Wanner

We analyze the performance of DMD-based approximations of the stochastic Koopman operator for random dynamical systems where either the dynamics or observables are affected by noise. Under certain ergodicity assumptions, we show that standard DMD algorithms converge provided the observables do not contain any noise and span an invariant subspace of the stochastic Koopman operator. For observables with noise, we introduce a new, robust DMD algorithm that can approximate the stochastic Koopman operator and demonstrate how this algorithm can be applied to Krylov subspace based methods using a single observable measured over a single trajectory. We test the performance of the algorithms over several examples.

1 Introduction

For many complex systems and processes, governing equations cannot be derived through first principles, or the models generated by them may be too complicated to be of practical use. Additionally, for such a system, the true state of the system may be difficult or even impossible to measure, making a state-space model impractical for applications such as control or prediction. Instead, only a limited set of measurements, or observables, will be made available. One of the tools available to model such a system is the Koopman operator. The Koopman operator represents a system in a high-dimensional linear space, which allows us to use spectral methods to analyze the system.

Originally introduced in [7], the Koopman Operator has gained traction for its utility as a data driven method through various form of Koopman Mode Decomposition (KMD), which decomposes the system based on eigenfunctions of the Koopman operator [9], [10]. Introduced in [8], Generalized Laplace Analysis (GLA) is an early data driven method of KMD based on the generalized Laplace transform. Another data driven method is Dynamic Mode Decomposition (DMD), which was introduced in [14] and shown to be connected to KMD in [13]. There are many different variations of DMD and it can be used for a wide array of applications. Despite their widespread use, many DMD algorithms possess a major drawback; they can fail if the data contains noise or some other randomness.

For systems with random dynamics, the eigenvalues produced by standard DMD algorithms converge to the spectrum of the stochastic Koopman operator provided the observables themselves do not contain any randomness and lie within a finite dimensional invariant subspace. Total Least Squares (TLS) DMD, proposed in [5], was developed to remove the bias for systems with measurement noise, but only converges when the underlying dynamics are deterministic. In [15], subspace DMD was introduced to converge for observables with additive noise even when the underlying dynamics are random. While many of these methods can combat the bias from measurement noise in DMD, they im-

pose relatively strict assumptions on either the dynamics (the underlying dynamics must be deterministic for TLS DMD) or the structure of the noise (i.i.d. additive noise in subspace DMD)

Of particular interest are Krylov subspace based DMD methods, where the iterates of a single observable under the Koopman evolution is used to generate an invariant subspace of the Koopman operator [11]. For deterministic systems, Hankel DMD uses time delays of a single observable to generate the Krylov subspace, and was shown to converge in [1]. This allows us to generate a model of a deterministic system using the data from a single trajectory of a single observable. However, for random systems, the time delayed observables contain randomness from the dynamics, and Hankel DMD does not converge. Further, the noise introduced is neither i.i.d. nor independent of the state. In [4], a new Stochastic Hankel DMD algorithm was shown to converge, but it requires the Stochastic Koopman evolution of the observable, which in general requires multiple realizations of the system.

In this paper, we introduce a new DMD algorithm which allows us to work with a more general set of observables with noise. We replace the i.i.d. assumption on the noise with a slightly weaker independence condition and we allow the distribution of the noise to depend on the state of the system. With these weaker conditions, we can use time delayed observables to form a variant of Hankel DMD. This allows us to compute a realization of the stochastic Koopman operator using data from a single observable over a single realization of the system. The paper is organized as follows: First we review the basics of random dynamical systems and the stochastic Koopman operator. Then, we establish the convergence of standard DMD algorithms for random systems in the absence of noise. Finally, we demonstrate the failure of standard DMD algorithms in the presence of noise and introduce a new algorithm which can accurately approximate the stochastic Koopman operator using noisy observables.

2 Random Dynamical Systems

2.1 Preliminaries

We define a random dynamical system as follows:

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let $\mathbb{T}$ be a semigroup. Let $\{\theta_t\}_{t\in\mathbb{T}}$ be a group or semigroup of measurable transformations on Ω which preserve the measure P. This forms a metric dynamical system on Ω. Now, let $(M, \mathfrak{B})$ be a measurable space, and let $T : \Omega \times \mathbb{T} \times M \to M$ be a measurable map. We say T forms a random dynamical system on M if the maps $T^t_\omega := T(\omega, t, \cdot) : M \to M$ form a cocycle over $\theta(\cdot)$, i.e.

$$T^0_\omega = id_M, \quad \text{and} \qquad T^{t+s}_\omega = T^t_{\theta_s(\omega)} \circ T^s_\omega. \tag{1}$$

We call $(\Omega, \mathfrak{F}, P, \theta)$ a driving dynamical system and θ a driving flow. If $\mathbb{T}$ is a group, then T^t_ω is invertible, with inverse T^{-t}_ω [2]. We will denote $T^t_\omega x_0 = x_t$ and $\theta_t\omega_0 = \omega_t$ for the remainder of the paper. We note that given an initial distribution for x_0, $\{(x_t), t \geq 0\}$ is a stochastic process.

Typically, we will not have access the state of the system at any given time. Instead, we will be able to measure some set of functions on the state space.

Definition 1. *An observable is any $\mathfrak{B}$ measureable map $f : M \to \mathbb{C}$.*

For any observable, we will denote $f(t) = f(x_t)$. We are interested in the evolution of observables over time. For a deterministic system the Koopman family of operators is defined to evolve an observable, f, on the state space under the flow, S^t of the system: $U^t f = f \circ S^t$. However, since a random dynamical system has many possible realizations, the stochastic Koopman operators is defined using the expectation of its evolution.

Definition 2. *The stochastic Koopman operator, $\mathcal{K}$, is defined by*

$$\mathcal{K}^t f(x) = \mathbb{E}_\Omega(f \circ T^t_\omega(x)) = \int_\Omega f \circ T^t_\omega(x) dP.$$

2.2 Measure Preservation and Stationarity

The preceding discussion allows for a very general random dynamical system. For our analysis, however, we will need restrict our attention to systems satisfying a few properties. First, we would like the stochastic Koopman operator family to have the semigroup property:

$$\mathcal{K}^{t+s} f = \mathcal{K}^s \circ \mathcal{K}^t f, \qquad s, t \geq 0. \tag{2}$$

For the deterministic Koopman operators, this is clearly true provided the system is autonomous and solutions exist and are unique, since in this case the flow forms a semigroup. To obtain the semigroup property for the stochastic Koopman family of operators, we need an analogous condition on the system. The condition that $\{(x_t),\ x_0 \in M\}$ is a time homogeneous Markov family suffices to give the semigroup property [4]. With this condition, the semigroup property follows from the Chapman-Kolmogorov equation. For a time homogeneous Markov family, the transition probabilities depend only on the current state; they do not depend on the time or the past of the process.

We will continue to specialize to systems which are weakly measure preserving. We say random dynamical system is weakly measure preserving if there exists a measure μ such that

$$\mathbb{E}(\mu((T_\omega^t)^{-1}A)) = \int_\Omega \mu((T_\omega^t)^{-1}A) dP = \mu(A) \tag{3}$$

for each $A \in \mathfrak{B}$. The measure μ is called an invariant measure. Weakly measure preserving systems give the stochastic Koopman operator the following nice property: For any $f \in L^1(\mu)$, we have

$$\int_M \mathcal{K}^t f d\mu = \int_M \int_\Omega f \circ T_\omega^t \, dP d\mu = \int_M f d\mu. \tag{4}$$

To see this, for any simple function $g = \sum_{i=1}^{n} a_i \chi_{A_i}$, we have

$$\int_M \mathcal{K}^t g \, d\mu = \int_M \int_\Omega g \circ T_\omega^t \, dP d\mu = \sum_{i=1}^{n} a_i \int_\Omega \mu((T_\omega^t)^{-1} A_i) \, dP = \sum_{i=1}^{n} a_i \mu(A_i) = \int_M g \, d\mu,$$

which we can use to extend (4) for any integrable f.

If our system is both a homogenous process and possesses an invariant measure μ, then we can extend (4) a little farther. If we let μ be our initial distribution for x_0, the stochastic process $\{(x_t), t \geq 0\}$ is a stationary process ([6], p.86). In this context μ is also called the stationary distribution. Stationary processes are determined by the transition kernal $P : M \times \mathfrak{B} \times \mathbb{T} \to [0, 1]$. The transition probability $P(x, A, t)$ is the probability that the state x evolves into the set A at time t:

$$P(x, A, t) = \int_\Omega \chi_A(T_\omega^t x) dP.$$

Stationarity gives us the equality, for $t_1, ..., t_n \geq 0$,

$$\int_M \int_\Omega f(T_\omega^{t_1+s} x, ..., T_\omega^{t_n+s} x) \, dP d\mu = \int_M \int_\Omega f(T_\omega^{t_1} x, ..., T_\omega^{t_n} x) \, dP d\mu. \tag{5}$$

Stationary processes have a useful representation as a measure preserving deterministic system in a probability space of bi-infinite sequences, $(M^{\mathbb{T}}, \mathfrak{C}, \nu)$ ([6]). Restricting to the discrete time case (i.e. $\mathbb{T} = \mathbb{Z}$ or $\mathbb{Z}^+$), for any $u = (..., u_0, u_1, u_2, ...) \in M^{\mathbb{T}}$, the evolution is given by the left shift map: $(Su)_t = u_{t+1}$. The invariant measure ν is constructed as follows:

We say that C is a cylinder set over the coordinates $t_1 < ... < t_k$ with the basis $B \in \mathfrak{B}^k$ if $C = \{u : (u_{t_1}, ..., u_{t_k}) \in B\}$. Let $\bar{\mathfrak{C}}$ be the semiring of cylinder sets. Define the

function $\nu : \bar{\mathfrak{C}}$ by

$$\nu(C) = \int_M \int_\Omega \chi_B(T_\omega^x, T_\omega^{t_2-t_1}x, ..., T_\omega^{t_k-t_1}x)dP(\omega)d\mu(x) = \mathbb{E}(\chi_B(x_{t_1}, x_{t_2}, ..., x_{t_k})). \quad (6)$$

It is easy to verify that (6) is well defined and countably additive, so we can extend ν to a measure on $\mathfrak{C}$, the completion of $\bar{\mathfrak{C}}$ by the Caratheodory extension theorem. Additionally, since ν is clearly invariant by (5), its extension to $\mathfrak{C}$ will also be invariant. The space $(M^{\mathbb{T}}, \mathfrak{C}, \nu)$ represents all possible realizations of the RDS T, and S gives the evolution of each realization over time. Using this, we can write the Koopman evolution of an observable on M using a conditional expectation on $M^{\mathbb{T}}$:

$$\mathcal{K}^t f(x) = \int_\Omega f(T_\omega^t x)dP = \mathbb{E}_{M^{\mathbb{T}}}(f(u_{s+t})|u_s = x).$$

The representation of the system in the space of sequences also allows us to define the adjoint of the stochastic Koopman operator, if we restrict our space of observables to $L^2(\mu)$.

$$(\mathcal{K}^{t*} f)(x) = \mathbb{E}_{M^{\mathbb{T}}}(f(u_{s-t})|u_s = x). \quad (7)$$

To see that this is the adjoint, using (5) and (7), we have

$$\begin{aligned}\langle \mathcal{K}^t f, g\rangle &= \int_M \mathbb{E}_{M^{\mathbb{T}}}(f(u_{t+s})|u_s = x)g(x)d\mu = \int_{M^{\mathbb{T}}} f(u_{s+t})g(u_s)d\nu \\ &= \int_{M^{\mathbb{T}}} f(u_s)g(u_{s-t})d\nu = \int_M f(x)\mathbb{E}_{M^{\mathbb{T}}}(g(u_{s-t})|u_t = x)d\mu = \langle f, \mathcal{K}^{t*} g\rangle.\end{aligned}$$

The adjoint gives the expected previous value of observables. If $\mathbb{T}$ is a group (and therefore T_ω is invertible), this can be expressed more simply in terms of the negative time evolution of observables:

$$\mathcal{K}^t f(x) = \int_\Omega f(T_\omega^{-t} x)dP.$$

We use (7) to define $\mathcal{K}^*$ since T_ω^t is not necessarily invertible. However,

2.3 Invariant Sets and Ergodicity

We define an invariant set for a stationary random system to be a set A such that

$$\int_A \int_\Omega \chi_A \circ T_\omega^t x \, dP d\mu = \mu(A).$$

In other words, for almost every ω and almost every $x \in A$, $T_\omega^t x \in A$. Clearly the countable union of invariant sets is an invariant set. Similarly the complement of A, A^c is an invariant set, since

$$\int_{A^c} \int_\Omega \chi_A \circ T_\omega^t x dP d\mu = \int_M \int_\Omega \chi_A \circ T_\omega^t x dP d\mu - \int_A \int_\Omega \chi_A \circ T_\omega^t x dP d\mu = \mu(A) - \mu(A) = 0,$$

so the invariant sets form a σ-algebra. For a deterministic measure preserving system, Birkhoff's Ergodic theorem gives us information about the time average of a function. In the discrete time case, it tells us

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(x_j) = \mathbb{E}(f \mid \mathfrak{I}),$$

where $\mathfrak{I}$ is the σ-algebra of invariant sets [6]. We will need a similar fact to hold on random systems in order to work with time averages.

Lemma 1. *Let $\mathfrak{I}$ be the σ-algebra of invariant sets of M for the stationary random dynamical system T. Let $\hat{f} : M^k \to \mathbb{C}$ be $\mathfrak{B}^k$ measurable and let $f : \Omega \times M$ be defined by*

$$f(x, \omega) = \hat{f}(T_\omega^{t_1} x, ..., T_\omega^{t_k} x).$$

Then, if $\int_M \int_\Omega |f| \, dP d\mu \leq \infty$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(x_j, \omega_j) = \mathbb{E}_M(\mathbb{E}_\Omega(f(x,\omega)) \mid \mathfrak{I})$$

for almost every initial condition (x_0, ω_0) *(with respect to* $\mu \times P$*).*

Proof. Consider the representation of the Markov process in a probability space of sequences, $(M^{\mathbb{T}}, \mathfrak{C}, \nu)$ with the measure given in (6) and the evolution given by the shift map. Denote $\hat{f}(u) = \hat{f}(u_{t_1}, ..., u_{t_k})$, where u is the sequence corresponding to the initial condition (x_0, ω_0). Since S is a measure preserving map, by Birkhoff's Ergodic Theorem we have for almost every u,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(x_n, \omega_n) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \hat{f}(S^j u) = \mathbb{E}_{M^{\mathbb{T}}}(\hat{f} \,|\, \hat{\mathfrak{I}}),$$

where $\hat{\mathfrak{I}}$ is the σ-algebra of S invariant sets in $M^{\mathbb{T}}$.

Now we need to show that $\mathbb{E}_{M^{\mathbb{T}}}(\hat{f} \mid \hat{\mathfrak{I}}) = \mathbb{E}_M(\mathbb{E}_\Omega(f(x,\omega)) \mid \mathfrak{I})$. For any cylinder set C with basis B, let $g_C(x,\omega) = \chi_B(x, T_\omega^{t_2-t_1}, ..., T_\omega^{t_k-t_1})$ and

$$h_C(x) = \int_\Omega \chi_B((x, T_\omega^{t_2-t_1}x, ..., T_\omega^{t_k-t_1}x)) dP(\omega) = \int_\Omega g_C(x,\omega) \, dP.$$

Let A be an S-invariant set and let $h(x) = \inf_{C \supset A} h_C(x)$. If C_1 and C_2 are both cylinder sets containing A, if t_1 and t_k are the first and last times for C_1 and $s > t_k - t_1$, we have

$$h(x) \leq \int_\Omega g_{C_1}(x,\omega) g_{C_2}(T_\omega^s x, \theta_s \omega) dP = \int_\Omega g_{C_1}(x,\omega) \int_\Omega g_{C_2}(T_\omega^s x, \alpha) dP(\alpha) dP(\omega)$$

Taking the infimum over all $C_2 \supset A$, this gives us

$$h(x) \leq \int_\Omega g_{C_1}(x,\omega) h(T_\omega^s x) dP.$$

Additionally, choosing $C_1 = M^{\mathbb{T}}$, we have $h \leq \mathcal{K}^s h$. But then

$$\begin{aligned}\|h\|^2 = \int_M h^2 \, d\mu &\leq \int_M \int_\Omega h(h \circ T_\omega) dP d\mu \\ &\leq \left(\int_\Omega \int_M h^2 dP d\mu \int_\Omega \int_M (h \circ T_\omega)^2 dP d\mu \right)^{\frac{1}{2}} = \|h\|^2,\end{aligned}$$

which shows, by Cauchy-Schwarz, that h and $h \circ T_\omega$ are linearly dependent. Then $h(T_\omega(x)) = h(x)$ for almost every x, ω. Then for almost every x, ω we have

$$h(x) \leq \int_\Omega g_{C_1}(x, \omega) h(T_\omega^s x) dP = h(x) \int_\Omega g_{C_1}(x, \omega) dP.$$

Taking the infimum over all $C_2 \supset A$, we obtain $h(x) \leq h(x)^2$, so $h(x) = 0$ or $h(x) = 1$ for almost every x. Since $\tilde{A} = f^{-1}(1)$ is an invariant set, and $\tilde{A}^{\mathbb{T}} = A \pmod{\nu}$, we have

$$\int_A \hat{f}(x, T_\omega x, ..., T_\omega^{k-1} x) d\nu = \int_{\tilde{A}^{\mathbb{T}}} \hat{f}(x, T_\omega x, ..., T_\omega^{k-1} x) d\nu = \int_{\tilde{A}} \int_\Omega f(x, \omega) dP d\mu,$$

which gives the result. □

Lemma 1 shows that the value of the time average of an observable depends only on which ergodic set our initial condition lies within. If every invariant set of our system has μ-measure 0 or 1, we say that the system is ergodic. The proof above shows that this is equivalent to the process $\{x_t\}$ being an ergodic stationary process, which means that every S-invariant set in $M^{\mathbb{T}}$ has ν measure 0 or 1, and we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} f(x_n, \omega_n) = \int_M \int_\Omega f(x, \omega) dP d\mu \tag{8}$$

for any f as defined in Lemma 1. If we have a function $f : M \to C$ with no dependence on ω, this property allows us to use time averages to evaluate integrals over the state space. For the rest of this paper, we will assume that $\{x_t\}$ is an ergodic stationary process and that $\mathbb{T}$ is discrete ($\mathbb{T} = \mathbb{Z}$ or $\mathbb{Z}^+$).

Remark 1. *The definitions of invariant measure and ergodicity presented here are those used for stationary Markov processes, and differ from the standard ones for random dynamical systems. Typically, invariant and ergodic measures are defined in terms of the skew product system on $M \times \Omega$ given by $\Theta^t(x, \omega) = (T^t_\omega x, \theta_t \omega)$. The stationary measure on M is related to the disintegration of the invariant measure on $M \times \Omega$. We choose to focus on the stationary measures since we are interested in the evolution on the state space M and the evolution on Ω may not be possible to measure. See [2],[3],[12].*

3 Dynamic Mode Decomposition

Dynamic Mode Decomposition is an algorithm which allows the computation of an approximation of the Koopman operator from data. For what follows, we assume $\mathbb{T} = \mathbb{Z}$ or $\mathbb{Z}^+$. Assuming the eigenfunctions of $\mathcal{K}^t$, span our function space, we can decompose any (possibly vector valued) observable $\mathbf{f}$ as

$$\mathbf{f} = \sum_j v_j \phi_j.$$

The expected evolution of f is then given by

$$\mathbb{E}(\mathbf{f}(T^t_\omega x)) = \sum_j \lambda_j v_j \phi_j(x). \tag{9}$$

In this Koopman mode decomposition, the functions ϕ_j are the Koopman eigenfunctions with eigenvalue λ_j, and the vectors v_j are called the Koopman modes associated with f. However, the expansion above can contain an infinite number of terms. In order to work with (9) using finite arithmetic, we must restrict ourselves to a finite dimensional subspace of our original function space.

Let $\mathscr{F}$ be a finite dimensional subspace of $L^2(\mu)$, and let $\bar{\mathscr{F}}$ be its orthogonal complement. Let P_1 and P_2 be the projections on to $\mathscr{F}$ and $\bar{\mathscr{F}}$. For any function $g \in L_2(\mu)$,

we can compute the Koopman evolution as

$$\mathcal{K}g = P_1\mathcal{K}g + P_2\mathcal{K}g = P_1\mathcal{K}P_1g + P_2\mathcal{K}P_1g + P_1\mathcal{K}P_2g + P_2\mathcal{K}P_2g.$$

The operator $P_1\mathcal{K}P_1$ maps $\mathscr{F}$ into itself. Since for any $g \in \mathscr{F}$, $P_2g = 0$, we can view $P_1\mathcal{K}P_1$ as an approximation of $\mathcal{K}$ provided $\|P_2\mathcal{K}P_1\|$ is small. If $\mathscr{F}$ is an invariant subspace under $\mathcal{K}$, we have $\|P_2\mathcal{K}P_1\| = 0$, and $\mathcal{K}g = P_1\mathcal{K}P_1g$ for all $g \in \mathscr{F}$. If we let $f_1, f_2, ..., f_k$ be a basis for $\mathscr{F}$, we can represent the restriction of $P_1\mathcal{K}P_1$ to $\mathscr{F}$ as a matrix $\mathbf{K}$ that acts on the basis by

$$\mathbf{K}\begin{bmatrix} f_1 & f_2 & \dots & f_k \end{bmatrix}^T = \begin{bmatrix} \mathcal{K}f_1 & \mathcal{K}f_2 & \dots & \mathcal{K}f_k \end{bmatrix}^T. \tag{10}$$

Remark 2. *The matrix K can also be thought of as the matrix acting (on the right) on the vector of coefficients of functions represented in the basis $f_1, \dots, f_K$: for any function $g \in \mathscr{F}$ we can write*

$$g = \sum_{j=1}^{k} a_j f_j = \mathbf{a}\begin{bmatrix} f_1 & \dots & f_k \end{bmatrix}^T,$$

and $\mathbf{a} = \begin{bmatrix} a_1 & \dots & a_k \end{bmatrix}$ is the row vector of coefficients of g. Then $(\mathbf{aK})$ is the row vector of coefficients for $\mathcal{K}g$, since

$$\mathcal{K}g = \mathcal{K}(\mathbf{a}\begin{bmatrix} f_1 & \dots & f_k \end{bmatrix}^T) = \mathbf{a}\begin{bmatrix} \mathcal{K}f_1 & \dots & \mathcal{K}f_k \end{bmatrix} = \mathbf{aK}\begin{bmatrix} f_1 & \dots & f_k \end{bmatrix}^T$$

Dynamic mode decomposition algorithms compute an approximation of the matrix $\mathbf{K}$ from data. If we can measure the observables $f_1, f_2, ..., f_k$ along a trajectory $x_0, x_1, ..., x_n$, we can form the vector valued observable $\mathbf{f} : M \to \mathbb{R}^k$ by

$$\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \dots & f_k \end{bmatrix}^T.$$

Given $\mathcal{K}$, Each $\mathbf{f}(t)$ is called a data snapshot. Given a data matrix of these snapshots

$$D = \begin{bmatrix} \mathbf{f}(0) & \mathbf{f}(1) & \dots & \mathbf{f}(n) \end{bmatrix},$$

we can construct an operator $A : \mathbb{R}^k \to \mathbb{R}^k$, called the DMD operator, which (approximately) maps each data snapshot to the next one, i.e.

$$A\mathbf{f}(i) \approx \mathbf{f}(i+1).$$

Standard DMD constructs a matrix C to minimize the error

$$\sum_{i=0}^{n-1} \|C\mathbf{f}(i) - \mathbf{f}(i+1)\|_2^2.$$

Algorithm 1: Standard DMD

Let $x_0, x_1, ..., x_n$ be a trajectory of our random dynamical system and $\mathbf{f} : M \to \mathbb{C}^k$ be a vector valued observable on our system.

1: Construct the data matrices

$$X = \begin{bmatrix} \mathbf{f}(0) & \mathbf{f}(1) & \dots & \mathbf{f}(n-1) \end{bmatrix}, \qquad Y = \begin{bmatrix} \mathbf{f}(1) & \mathbf{f}(2) & \dots & \mathbf{f}(n) \end{bmatrix}.$$

2: Form the matrix

$$C = YX^\dagger,$$

where $X^\dagger$ is the Moore-Penrose psuedoinverse.

3. Compute the eigenvalues and left and right eigenvectors, (λ_i, w_i, v_i) $i = 1, 2, ..., k$, of C. Then the dynamic eigenvalues are λ_i, the dynamic modes are v_i, and the numerical

eigenfunctions are given by

$$\hat{\phi}_i = w_i^T X.$$

Let $f_1, f_2, ..., f_k$ be the components of $\mathbf{f}$. If we let $\hat{f}_i$ be the i^{th} column of X,

$$\hat{f}_i = \begin{bmatrix} f_i(0) & f_i(1) & \dots & f_i(n-1) \end{bmatrix} = \begin{bmatrix} f_i(x_0) & f_i(x_1) & \dots & f_i(x_{n-1}) \end{bmatrix},$$

we see that $\hat{f}_i$ represents f_i by evaluating it along a trajectory. With standard DMD, we construct the DMD operator C represented in the basis $\hat{f}_1, \hat{f}_2, ..., \hat{f}_k$. Similarly, the numerical eigenfunctions, $\hat{\phi}_i$ will be approximations of eigenfunctions of the stochastic Koopman operator evaluated along our trajectory. Unfortunately, depending on the choice of basis, this DMD construction may be numerically unstable. This leads to the second algorithm [14].

Algorithm 2: SVD based DMD

Let $x_0, x_1, ..., x_n$ be a trajectory of our random dynamical system and, $f_1, f_2, ..., f_l$, $l \geq k$, be a set of l observables on our system.

1: Construct the data matrices

$$X = \begin{bmatrix} \mathbf{f}(0) & \mathbf{f}(1) & \dots & \mathbf{f}(n-1) \end{bmatrix}, \qquad Y = \begin{bmatrix} \mathbf{f}(1) & \mathbf{f}(2) & \dots & \mathbf{f}(n) \end{bmatrix}.$$

2: Compute the truncated SVD of X using the first k singular values.

$$X = W_k S_k V_k^*.$$

3: Form the matrix

$$A = S_k^{-1} W_k^* Y V_k.$$

4. Compute the eigenvalues and left and right eigenvectors, (λ_i, w_i, u_i) $i = 1, 2, ..., k$, of A. Then the dynamic eigenvalues are λ_i, the dynamic modes are

$$v_i = WSu_i,$$

and the numerical eigenfunctions are given by

$$\hat{\phi}_i = w_i^T V_k^*.$$

The benefit of SVD based DMD is that it is more numerically stable. If X has a large condition number, the pseudoinversion of X can introduce large errors to the DMD operator and make Algorithm 1 unstable. To combat this, Algorithm 2 computes the SVD of X and truncates to include only the dominant singular values. Since S_k has a smaller condition number than X, the inversion of S_k in Algorithm 2 is more numerically stable than the psuedoinversion of X. Algorithm 2 uses singular values and vectors to choose a basis of observables to construct the DMD operator; the matrix A generated is the same as the one produced by Algorithm 1 using the observable $\hat{\mathbf{f}} = S_k^{-1} W^* \mathbf{f}$.

4 Convergence of DMD for Random Systems

The utility of Algorithms 1 and 2 comes from the convergence of the dynamic eigenvalues and modes to eigenvalues and eigenfunctions of $\mathcal{K}$.

Proposition 1. *Let T be an ergodic random dynamical system. Let $\mathscr{F}$ be a k dimensional subspace of $L^2(\mu)$ which is invariant under the action of $\mathcal{K}$, and let $f_1, f_2, ..., f_k$ span $\mathscr{F}$. Let $\lambda_{j,n}$ be the dynamic eigenvalues and $v_{j,n}$ be the dynamic modes produced by Algorithm 1 using the trajectory $x_0, x_1, ..., x_n$. Then as $n \to \infty$, the dynamic eigenvalues converge to the eigenvalues of $\mathcal{K}$ restricted to $\mathscr{F}$ for almost every initial condition (x_0, ω_0) with*

respect to $(\mu \times \omega)$. *If the eigenvalues of* $\mathcal{K}$ *are distinct, the numerical eigenfunctions converge to a sampling of the eigenfunctions along the trajectory.*

Proof. Let $f_1, f_2, ..., f_k$, and $\mathbf{K}$ be as described in (10). Let X_n, Y_n, and C_n be the matrices produced by Algorithm 1 for the trajectory $x_0, x_1, ..., x_n$, and let $\omega_0, \omega_1, ..., \omega_n$ be the evolution of the noise. Let $\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \dots & f_k \end{bmatrix}^T$ as above. If we define the matrices

$$G_0 = \int_M \begin{bmatrix} f_1 & f_2 & \dots & f_k \end{bmatrix}^T \begin{bmatrix} f_1^* & f_2^* & \dots & f_k^* \end{bmatrix} d\mu = \int_M \mathbf{f}\,\mathbf{f}^*\, d\mu$$

and

$$G_1 = \int_M \begin{bmatrix} \mathcal{K}f_1 & \mathcal{K}f_2 & \dots & \mathcal{K}f_k \end{bmatrix}^T \begin{bmatrix} f_1^* & f_2^* & \dots & f_k^* \end{bmatrix} d\mu = \int_M \mathbf{K}\,\mathbf{f}\,\mathbf{f}^*\, d\mu = \mathbf{K}G_0.$$

We can see that G_0 has full rank, since if $\mathbf{v}$ is in its nullspace, we would have

$$\|\mathbf{f}^*\mathbf{v}\|^2 = \mathbf{v}^* G_0 \mathbf{v} = 0,$$

which implies $\mathbf{v} = 0$ since $f_1, f_2, ..., f_k$ are linearly independent. This gives us $\mathbf{K} = G_0^{-1} G_1$.

Now, let $G_{0,n} = \frac{1}{n} X_n X_n^*$ and $G_{1,n} = \frac{1}{n} X_n Y_n^*$. We have $G_{0,n} \to G_0$ and $G_{1,n} \to G_1$ for almost every initial condition (x_0, ω_0). To see this, by Lemma 1 we have

$$\begin{aligned} \lim_{n\to\infty} G_{1,n} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{f}(x_{m+1})\mathbf{f}^*(x_m) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{f}(T_{\omega_m} x_m)\mathbf{f}^*(x_m) \\ &= \int_M \int_P \mathbf{f}(T_\omega x)\mathbf{f}^*(x)\, dP d\mu = \int_M \mathbf{K}\,\mathbf{f}(x)\,\mathbf{f}^*(x)\, d\mu = G_1, \end{aligned}$$

and similarly for G_0, we have

$$\lim_{n\to\infty} G_{0,n} = \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{f}(x_m)\mathbf{f}^*(x_m) = \int_M \int_\Omega \mathbf{f}(x)\mathbf{f}^*(x)\, dP d\mu = G_0.$$

Since G_0 has full rank and $G_{0,n} \to G_0$, $G_{0,n}$ is full rank for n large enough, so $G_{0,n}^{-1}$ exists and

$$\lim_{n\to\infty} G_{0,n}^{-1} G_{1,n} = G_0^{-1} G_1 = \mathbf{K}.$$

Since $G_{0,n} = \frac{1}{n} X_n X_n^*$, we have X_n has full row rank for n large enough, so

$$C_n = Y_n (X_n)^\dagger = Y_n X_n^* (X_n X_n^*)^{-1} = \left(\frac{1}{n} Y_n X_n^*\right)\left(\frac{1}{n} X_n X_n^*\right)^{-1} = G_{0,n}^{-1} G_{1,n},$$

which shows that $C_n \to \mathbf{K}$. This shows that the dynamic eigenvalues $\lambda_{j,n}$ converge to the eigenvalues of $\mathbf{K}$, λ_j as $n \to \infty$.

To show the numerical eigenfunctions converge to samplings of our eigenfunctions, let $w_{j,n}$ and w_j be the left eigenvectors of C_n and $\mathbf{K}$, respectively. Consider the functions $\phi_{j,n} = w_{j,n}^T \mathbf{f}$ and $\phi_j = w_j^T \mathbf{f}$. We know ϕ_j is a Koopman eigenfunction, since

$$\mathcal{K}\phi_j = \mathcal{K}(w_j^T \mathbf{f}) = w_j^T \mathbf{K}\, \mathbf{f} = \lambda_j w_j^T \mathbf{f} = \lambda_j \phi_j.$$

If $\mathbf{K}$ has distinct eigenvalues, the vectors $w_{j,n}$ each converge to w_j, so $\phi_{j,n} \to \phi_j$. The numerical eigenfunctions, $\hat{\phi}_{j,n}$ are the values of the function $\phi_{j,n}$ sampled along the trajectory $x_0, ..., x_{n-1}$. □

The convergence of Proposition 1 is based on the inner product of functions in $L^2(\mu)$. As such, we cannot glean any information about dynamics outside the support of μ. There could be an eigenvalue/eigenfunction pair, (λ, ϕ), such that ϕ is zero on the support of μ. Such a pair cannot be captured by Algorithm 1, since $\phi = 0$ almost everywhere with respect to μ. In particular, if μ is a singular measure concentrated on some attractor, the eigenvalues governing the dissipation to the attractor cannot be found using ergodic sampling. For random systems, the invariant measure is often absolutely continuous, so

the dissipation can be captured, but may take a large amount of data depending on the distribution μ.

5 DMD with Noisy Observables

5.1 Preliminaries

The proof above shows that Dynamic Mode Decomposition converges for random dynamical systems. However, it is important to note that although the systems can have randomness, the observables cannot. The stochastic Koopman operator acts on functions, $f : M \to \mathbb{C}$, which depend only on the state of the system. If we allow our observables to have some noise, (i.e. dependence on ω), the proof fails. In particular, observables with i.i.d. measurement noise and time delayed observables (used in Hankel DMD) both have some dependence on ω, and therefore cannot be used with the above DMD methods. Examining the failure of standard DMD with noisy observables is instructive. First we must define our requirements for "noisy observables."

Definition 3. *A noisy observable is a measurable map $\tilde{f} : M \times \Omega \to \mathbb{C}$, such that the random function $\tilde{f}_\omega = \tilde{f}(\,\cdot\,, \omega) : M \to \mathbb{C}$ is $\mathfrak{B}$ measurable for almost every ω.*

For notation, we will always denote a noisy observable, $\tilde{f}$, with a tilde and let the space of noisy observables be $\mathscr{H}$. We will also denote $\tilde{f}(t) = \tilde{f}_{\omega_t}(x_t)$ and will define f to be its mean:

$$f(x) = \int_\Omega \tilde{f}_\omega(x) dP.$$

In what follows, we will assume that f exists and is in $L^2(M)$.

In order to evaluate the stochastic Koopman evolution of f, we will need to place further restrictions on $\tilde{f}$. If $\tilde{f}_{\omega_t}$ is independent of T_{ω_s} for all $s < t$, we will say $\tilde{f} \in \mathscr{H}_+$. Roughly speaking, this means the random function $\tilde{f}_{\omega_t}$ cannot be predicted by the past of the dynamics on M. This gives us

$$\int_\Omega \tilde{f}_{\theta^j\omega}(T_\omega^j x)dP = \int_\Omega \int_\Omega \tilde{f}_\psi(T_\omega^j x)dP(\psi)dP(\omega) = \int_\Omega f(T_\omega^j x)dP = \mathcal{K}^j f(x). \tag{11}$$

Finally, in order to approximate integrals from data, we will need some ergodicity assumptions on our noisy observables. Namely, we will need time averages to converge in a similar sense to Lemma 1 and (8). In particular, we will need

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \tilde{\mathbf{f}}(t+j)\tilde{\mathbf{g}}(t) = \int_M \int_\Omega \tilde{\mathbf{f}}_{\theta_j}(T_\omega^j x)\tilde{\mathbf{g}}_\omega(x)dPd\mu, \tag{12}$$

for two vector valued noisy observables $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ and almost every initial condition (x_0, ω_0).

Remark 3. *While we make the ergodicity assumption for generality, we will show that (12) holds for observables with i.i.d. measurement noise and time delayed observables, the primary observables of interest in this paper. More generally, we can consider the skew product system Θ on $M \times \Omega$ treat $\tilde{f}$ as an observable on $M \times \Omega$, $\tilde{f}(x, \omega) = \tilde{f}_\omega(x)$.If Θ has an invariant measure and $\tilde{f}$ is independent of the σ-algebra of Θ invariant sets, we can evaluate its average with respect to the invariant measure on $M \times \Omega$.*

5.2 Failure of Dynamic Mode Decomposition with Noisy Observables

Now, assuming some of the above properties, we can see exactly how DMD fails. The convergence of DMD depends largely on estimation of the inner products using time averages. As before let $f_1, ..., f_k$ be observables which span a k-dimensional subspace $\mathscr{F}$, and let $\mathbf{K}$ be the restriction of $\mathcal{K}$ to $\mathscr{F}$ as in (10). Let $\mathbf{f} = \begin{bmatrix} f_1 & \dots & f_k \end{bmatrix}^T$. We have from Lemma 1 that

$$G_j = \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{f}(x_{m+j})\mathbf{f}^*(x_m) = \int_M \int_\Omega \mathbf{f}(T_\omega^j x)\mathbf{f}^*(x)dPd\mu = \int_M \mathbf{K}^j \mathbf{f}\, \mathbf{f}^*\, d\mu.$$

We can use the fact that $G_j = \mathbf{K}G_{j-1}$ to estimate $\mathbf{K}$. However, suppose have an ergodic (in the sense that (12) holds) noisy observable $\tilde{\mathbf{f}} \in \mathscr{H}^+$ such that $\mathbb{E}_\Omega(\tilde{\mathbf{f}}_\omega) = \mathbf{f}$. When we take the comparable time average, we do not obtain the same results:

$$\tilde{G}_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \tilde{\mathbf{f}}(m+j)\tilde{\mathbf{f}}(m) = \int_M \int_\Omega \tilde{\mathbf{f}}_{\theta_j\omega}(T_\omega^j x)\tilde{\mathbf{f}}_\omega^*(x) dP d\mu.$$

We cannot simplify this inner product any farther because $\tilde{\mathbf{f}}_\omega$ and $\tilde{\mathbf{f}}_{\theta_j\omega} \circ T_\omega^j$ are not necessarily independent. In fact, if we examine the error in these averages, we obtain

$$\tilde{G}_j - G_j = \int_M Cov(\tilde{\mathbf{f}}_{\theta_j\omega} \circ T_\omega^j, \tilde{\mathbf{f}}_\omega) d\mu,$$

where $Cov(\tilde{\mathbf{f}}_{\theta_j\omega} \circ T_\omega^j, \tilde{\mathbf{f}}_\omega)(x)$ denotes the covariance of $\tilde{\mathbf{f}}_{\theta_j\omega}(T_\omega^t x)$ and $\tilde{\mathbf{f}}_\omega(x)$.
Since Algorithms 1 and 2 depend on the numerical approximations of G_j, we can conclude that the error stems from the covariances of the observables. However, if we had an observable $\tilde{\mathbf{g}}_\omega$ which met some independence conditions with $\tilde{\mathbf{f}}_\omega$ and T_ω^t, we could still compute $\mathbf{K}$. This brings us to our third algorithm.

5.3 Noise Resistant DMD Algorithms

Algorithm 3: Noise Resistant DMD

Let $\tilde{\mathbf{f}} \in \mathscr{H}_+^k$, and $\tilde{\mathbf{g}} \in \mathscr{H}^l,\ l \geq k$.

1: Construct the data matrices

$$X = \begin{bmatrix} \tilde{\mathbf{f}}(0) & \tilde{\mathbf{f}}(1) & \dots & \tilde{\mathbf{f}}(n-1) \end{bmatrix}^T,$$

$$Y = \begin{bmatrix} \tilde{\mathbf{f}}(1) & \tilde{\mathbf{f}}(2) & \dots & \tilde{\mathbf{f}}(n) \end{bmatrix}^T,$$

and

$$Z = \begin{bmatrix} \tilde{\mathbf{g}}(0) & \tilde{\mathbf{g}}(1) & \dots & \tilde{\mathbf{g}}(n-1) \end{bmatrix}^T.$$

2: Form the matrices $\tilde{G}_0 = \frac{1}{n} X Z^*$ and $\tilde{G}_1 = \frac{1}{n} Y Z^*$.

3: Compute the matrix

$$C = \tilde{G}_1 \tilde{G}_0^\dagger.$$

4: Compute the eigenvalues and left and right eigenvectors, (λ_i, w_i, v_i) of C. The dynamic eigenvalues are λ_i, the dynamic modes are v_i, and the numerical eigenfunctions are given by

$$\hat{\phi}_i = w_i^T X.$$

The idea behind Algorithm 3 is to use a second noisy observable, $\tilde{\mathbf{g}}$, which meets some independence requirements with $\tilde{\mathbf{f}}$, to generate a second basis for $\mathscr{F}$. If $\tilde{\mathbf{g}}$ meets the proper independence requirements, the convergence can be shown in a similar manner to Proposition 1.

Proposition 2. *Let $\tilde{\mathbf{f}} \in \mathscr{H}_+^k$ and $\tilde{\mathbf{g}} \in \mathscr{H}^l$ be such that $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ satisfy (12). Suppose $\tilde{\mathbf{g}}_{\omega_t}$ is independent of $\tilde{\mathbf{f}}_{\omega_t}$, $\tilde{\mathbf{f}}_{\omega_{t+1}}$, and T_{ω_t}. Define $\mathbf{f}(x) = \mathbb{E}_\Omega(\tilde{\mathbf{f}}_\omega(x))$ and $\mathbf{g}(x) = \mathbb{E}_\Omega(\tilde{\mathbf{g}}_\omega(x))$. Suppose the components of $\mathbf{f}$, $f_1, ..., f_k$, span a k-dimensional invariant subspace, $\mathscr{F}$, of $\mathcal{K}$ and $\mathscr{F} \subset span\{g_1, ..., g_l\}$, where $g_1, ..., g_l$ are the components of $\mathbf{g}$. Then the matrix C generated by Algorithm 3 converges to the restriction of $\mathcal{K}$ to $\mathscr{F}$ as $n \to \infty$.*

Proof. Let $\mathbf{K}$ be the restriction of $\mathcal{K}$ to $\mathscr{F}$. Let $\tilde{G}_{0,n}$ and $\tilde{G}_{1,n}$ be the matrices generated in Algorithm 3 with n data points. Using the independence conditions on $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{f}}$ and (11), define

$$G_0 = \int_M \int_\Omega \tilde{\mathbf{f}}_\omega(x) \tilde{\mathbf{g}}_\omega^*(x) dP d\mu = \int_M \mathbf{f}\, \mathbf{g}^* d\mu \tag{13}$$

and

$$G_1 = \int_M \int_\Omega \tilde{\mathbf{f}}_{\theta\omega}(T_\omega x)\tilde{\mathbf{g}}^*_\omega(x) dP d\mu = \int_M \int_\Omega \tilde{\mathbf{f}}(T_\omega x) dP \int_\Omega \tilde{\mathbf{g}}^*_\omega(x) dP d\mu = \mathbf{K} \int_M \mathbf{f}\, \mathbf{g}^* d\mu. \tag{14}$$

We can show that G_0 has full row rank, since if $\mathbf{v}$ is in its left nullspace, we would have

$$\langle \mathbf{v}^T \mathbf{f}, g_i \rangle = 0$$

for each i, which shows $\mathbf{v} = 0$ since $\mathscr{F} \subset \text{span}\{g_i\}$. This gives us $\mathbf{K} = G_1 G_0^\dagger$. We will show that $\tilde{G}_{0,n} \to G_0$ and $\tilde{G}_{1,n} \to G_1$ as $n \to \infty$. Taking the limit of $G_{0,n}$ with (12) and using (13), we have

$$\lim_{n\to\infty} \tilde{G}_{0,n} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \tilde{\mathbf{f}}(m)\tilde{\mathbf{g}}^*(m) = \int_M \int_\Omega \tilde{\mathbf{f}}_\omega(x)\tilde{\mathbf{g}}^*_\omega(x)\, dP d\mu = G_0$$

and similarly $\tilde{G}_{1,n} \to G_1$ using (14). Since G_0 has full rank and $\tilde{G}_{0,n} \to G_0$, we have $\tilde{G}^\dagger_{0,n} \to G_0^\dagger$, so $\tilde{G}_{1,n}\tilde{G}^\dagger_{0,n} \to \mathbf{K}$. □

It follows from Proposition 2 that the eigenvalues and eigenvectors of C go to those of $\mathbf{K}$. Therefore, the dynamic eigenvalues limit to Koopman eigenvalues. The numerical eigenfunctions, however, are more complicated. If w_i is a left eigenvector of $\mathbf{K}$, we have $w_i^T \mathbf{f}$ is a Koopman eigenfunction. The numerical eigenfunctions, however, limit to $w_i^T X$, which a sampling of $w_i^T \tilde{\mathbf{f}}$. In this regard, the numerical eigenfunction is a sampling of an eigenfunction with some zero mean noise added to it.

The key idea in the proof of Proposition 2 is the assumption that we have a second observable $\tilde{\mathbf{g}}$ that is uncorrelated with $\tilde{\mathbf{f}}$. This allows us to estimate the inner product of $\mathbf{g}$ and $\mathbf{f}$ using time averages without introducing a covariance term. We call $\tilde{\mathbf{g}}$ our "dual observable" since we are using it to evaluate these inner products. While the necessity of a second observable may seem restrictive, Proposition 2 allows us to work with very general observables. If we specialize to more specific classes of observables, we will find

that we often do not need a second observable. Often, we can use time delays of single observable $\tilde{\mathbf{f}}$ so that $\tilde{\mathbf{f}}_{\omega_0}$ and $\tilde{\mathbf{f}}_{\omega_s}$ are independent.

5.4 Observables with i.i.d. Measurement Noise

Often, when measuring an observable on a system, the measurement will be imprecise. The error in the measurement are often modeled as an i.i.d. random variable. We call observables with this type of noise an observable with measurement noise:

Definition 4. *A noisy observable, $\tilde{f}$, is an observable with i.i.d. measurement noise if $\tilde{f}_\omega$ is an i.i.d. random function and is independent of the dynamics T_{ω_t}.*

We note that for any given ω, the measurement error,

$$\tilde{e}_\omega(x) = \tilde{f}_\omega(x) - f(x),$$

can vary over the state space M; it does not need to be a constant additive noise. Since $\tilde{f}_{\omega_t}$ is an i.i.d. random variable and independent of T_{ω_t} for all t, the ordered pair $(x_t, \tilde{f}_{\omega_t}) \in M \times L^2(M)$ is an ergodic stationary process, with ergodic measure $\nu = \mu \times \tilde{f}_*(P)$, where $\tilde{f}_*(P)$ is the pushforward of P. This allows us to evaluate the time averages as in (12). The proof of this follows from the lemma below and the fact that i.i.d. processes are mixing.

Lemma 2. *Let x_t and y_t be independent stationary processes with invariant measures μ_1 and μ_2, respectively. Then the ordered pair (x_t, y_t) is stationary with stationary measure $\mu_1 \times \mu_2$. If x_t is ergodic and y_t is mixing, (x_t, y_t) is ergodic.*

Proof. Let P_1 and P_2 be the transition kernals for x_t and y_t, respectively, and let P be the transition kernal for (x_t, y_t). The pair (x_t, y_t) is stationary with measure $\mu_1 \times \mu_2$,

since

$$\int_{X\times Y} P((x,y),A\times B,t)d(\mu_1\times\mu_2) = \int_X P_1(x,A,t)d\mu_1 \int_Y P_2(y,B,t)d\mu_2 = \mu_1(A)\mu_2(B).$$

Now, let ν_1 and ν_2 be the cylinder measures defined by (6) for x_t and y_t, respectively. Let S_1 and S_2 be the shift map on $X^{\mathbb{T}}$ and $Y^{\mathbb{T}}$, respectively. If S_1 is ergodic and S_2 is mixing, then the map $S(\mathbf{x},\mathbf{y}) = (S_1(\mathbf{x}), S_2(\mathbf{y}))$ is ergodic with respect to $\nu_1\times\nu_2$. □

If the components of $\tilde{\mathbf{f}}$ are observables with measurement noise, it turns out we don't need second observable to use in Algorithm 3. Instead, we can use a time shift of $\tilde{\mathbf{f}}$ to generate $\tilde{\mathbf{g}}$. The i.i.d. property of $\tilde{\mathbf{f}}$ will give us the independence properties we need.

Corollary 1. *Suppose $\tilde{\mathbf{f}}$ is a vector valued observable with i.i.d. measurement noise, and the components of $\mathbf{f}$, $f_1,...,f_k$ span a k-dimensional invariant subspace, $\mathscr{F}$ such that the restriction of $\mathcal{K}$ to $\mathscr{F}$ has full rank. Then Algorithm 3 converges setting $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-1)$.*

Proof. Let $\mathbf{K}$ be the resriction of $\mathcal{K}$ to $\mathscr{F}$. By Lemma 2, (x_t, f_{ω_t}) is an ergodic stationary sequence. Then, using ergodicity and the independence properties of $\tilde{\mathbf{f}}$, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\tilde{\mathbf{f}}(m)\tilde{\mathbf{g}}^*(m) = \lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\tilde{\mathbf{f}}(m)\tilde{\mathbf{f}}^*(m-1) &= \int_M\int_\Omega \tilde{\mathbf{f}}_{\theta\omega}(T_\omega^{j+1}x)\tilde{\mathbf{f}}_\omega^*(x)\,dPd\mu \\ &= \int_M\int_\Omega \tilde{\mathbf{f}}_{\theta\omega}(T_\omega x)dP\int_\Omega \tilde{\mathbf{f}}_\omega(x)dPd\mu = \mathbf{K}\int_M \mathbf{f}\,\mathbf{f}^*d\mu,\end{aligned}$$

which has full rank since $\mathbf{K}$ has full rank. Similarly,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\tilde{\mathbf{f}}(m+1)\tilde{\mathbf{g}}^*(m) = \lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\tilde{\mathbf{f}}(m+1)\tilde{\mathbf{f}}^*(m-1) = \mathbf{K}^2\int_M \mathbf{f}\,\mathbf{f}^*d\mu.$$

The rest of the proof follows Proposition 2. □

Remark 4. *It is useful to note that if $\mathbb{T}$ is a group (e.g. $\mathbb{T}=\mathbb{Z}$), we would be able to define $\tilde{\mathbf{g}}_\omega = \tilde{\mathbf{f}}_{\theta_{-1}\omega}\circ(T_\omega^{-1})$, and $\tilde{\mathbf{g}}$ would meet the conditions of Proposition 2 exactly.*

However, if $\mathbb{T} = \mathbb{Z}^+$, we cannot define $\tilde{\mathbf{g}}_\omega \in L^2(M)$ explicitly, since T_ω may not be invertible. However, since we are still able to evaluate time averages, the proof is nearly identical.

6 Time Delayed Observables and Krylov Subspace Methods

Another important type of noisy observable are time delayed observables. Allowing time delayed observables in DMD is useful for two reasons. First, time delays allow us to enrich our space of observables. Oftentimes, there are functions on our state space which cannot be measured by a certain set of observables, but can be observed if we allow time delays. For example, the velocity of a moving mass cannot be observed by any function on the position, but can be approximated using the position at two different times. Second, using time delays allows us to identify an invariant (or nearly invariant) subspace spanned by the Krylov sequence $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$.

Of particular interest is an analogue of Hankel DMD for random systems, which uses a Krylov sequence of observables to generate our finite subspace. With Hankel DMD, we use a single observable, $\tilde{f}$, and its time delays to approximate the sequence $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$. If $\tilde{f}$ is an observable with measurement noise (or has no noise), we can define

$$\tilde{\mathbf{f}}(t) = \begin{bmatrix} \tilde{f}(t) & \tilde{f}(t+1) & \dots & \tilde{f}(t+k-1) \end{bmatrix}^T.$$

By (11), its mean is

$$\int_\Omega \tilde{\mathbf{f}}\, dP = \begin{bmatrix} f & \mathcal{K}f & \dots & \mathcal{K}^{k-1}f \end{bmatrix}^T.$$

We can then use time delays of $\tilde{f}$ to approximate the Krylov sequence $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$. Additionally, if we set $\tilde{g}(t) = \tilde{f}(t-k)$ in Algorithm 3 we will have the necessary indepen-

dence conditions, and the time averages will converge as in (12) due to the pair $(x_t, \tilde{f}_{\omega_t})$ being an ergodic stationary variable.

Corollary 2. *(Hankel DMD for Random Systems) Let $\tilde{f}$ be an observable with measurement noise, and let its mean, f, be such that the Krylov sequence $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$ spans a k-dimensional invariant subspace $\mathscr{F}$ and the restriction of $\mathcal{K}$ to $\mathscr{F}$ has full rank. Let*

$$\tilde{\mathbf{f}}(t) = \begin{bmatrix} \tilde{f}(t) & \tilde{f}(t+1) & \dots & \tilde{f}(t+k-1) \end{bmatrix}^T,$$

and

$$\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-k) = \begin{bmatrix} \tilde{f}(t-k) & \tilde{f}(t-k+1) & \dots & \tilde{f}(t-1) \end{bmatrix}^T.$$

Then the matrix A generated by Algorithm 3 converges to the restriction of $\mathcal{K}$ to $\mathscr{F}$. If $\tilde{f}$ has no noise (i.e. $\tilde{f} = f$) we can use

$$\tilde{\mathbf{g}}'(t) = \tilde{\mathbf{f}}(t-k+1) = \begin{bmatrix} \tilde{f}(t-k+1) & \tilde{f}(t-k+2) & \dots & \tilde{f}(t) \end{bmatrix}^T.$$

Proof. Using (11), we can see that the components of $\mathbf{f}$ are $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$, which spans $\mathscr{F}$. Additionally, using the independence properties of $\tilde{f}$, we have $\tilde{\mathbf{f}}_{\omega_t}$ and $\tilde{\mathbf{f}}_{\omega_{t+s}}$ are independent for $s \geq k$. Since $(x_t, \tilde{f}_{\omega_t})$ is ergodic by Lemma 2, we can take the time averages

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{m=k}^{n+k-1} \mathbf{f}(m)\mathbf{g}^*(m) &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{f}(m+k)\mathbf{f}^*(m) = \int_M \int_\Omega \tilde{\mathbf{f}}_{\theta^k\omega}(T_\omega^k x)\tilde{\mathbf{f}}_\omega^*(x) dP d\mu \\ &= \int_M \int_\Omega \tilde{\mathbf{f}}_{\theta_k\omega}(T_\omega^k x)\tilde{\mathbf{f}}^*(x) dP d\mu = \mathbf{K}^k \int_M \mathbf{f}\,\mathbf{f}^* d\mu, \end{aligned}$$

which has full rank since $\mathbf{K}$ has full rank. Similarly, we can take the time average

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=k}^{n+k-1} \mathbf{f}(m+1)\mathbf{g}^*(m) = \mathbf{K}^{k+1} \int_M \mathbf{f}\,\mathbf{f}^* d\mu,$$

and the rest of the proof follows Proposition 2. If $\tilde{f}_\omega = f$, $\tilde{\mathbf{f}}_{\omega_t}$ and $\tilde{\mathbf{f}}_{\omega_{t+k-1}}$ are independent, and we can take the time averages using $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-k+1)$.

□

Corollary 2 allows us to compute an approximation of $\mathcal{K}$ using the data from a single observable evaluated along a single trajectory. However, the method does not require that the we only use time delays of a single observable. In general, even if $\tilde{f}$ is vector valued, we can take time delays of $\tilde{f}$ as in Corollary 2 so long as we span the proper subspace. The dual observable, $\tilde{\mathbf{g}}$, is also generated in the same way.

7 Conditioning of Algorithm 3

Asymptotically, the convergence rate of Algorithm 3 is governed by the rate at which $G_{0,n}$ and $G_{1,n}$ converges to G_0 and G_1, as defined in the proof of Proposition 2. This is governed by the convergence rate of ergodic sampling. However, Algorithm 3 also requires the pseudo-inversion of $G_{0,n} \approx G_0$. If the matrix G_0 is ill-conditioned, small errors in the time averages approximations of G_0 and G_1 can cause large errors in our DMD operator. The condition number of G_0, $\kappa(G_0)$, can become large if either set of observables, $f_1, ..., f_k$ or $g_1, ..., g_l$, are close to being linearly dependent.

Both of these issues arise particularly often when using Hankel DMD. With Hankel DMD, we use the basis $f, \mathcal{K}f, ..., \mathcal{K}^{k-1}f$ as our basis for $\mathscr{F}$. This is often a poor choice of basis, as f and $\mathcal{K}f$ may be close to being linearly dependent. This is particularly the case when data from a continuous time system is sampled with a short period, such as from a discretization of an ODE or SDE. Similarly, if j is large or $\mathcal{K}$ has eigenvalues close to zero, $\mathcal{K}^j f$ and $\mathcal{K}^{j+1}f$ may be close to being linearly dependent, which will also cause conditioning issues.

7.1 SVD Based Algorithms

To combat these conditioning issues, we have some leeway in the observables we choose for $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$. Looking at G_0, we have

$$G_0 = \int_M \mathbf{g}\,\mathbf{f}^*\, d\mu = \int_M \begin{bmatrix} f_1 & f_2 & \dots & f_k \end{bmatrix}^T \begin{bmatrix} g_1^* & g_2^* & \dots & g_l^* \end{bmatrix} d\mu. \tag{15}$$

Ideally, $\{g_1, ..., g_l\}$ and $\{f_1, ..., f_k\}$ would be orthonormal bases for , so $\kappa(G_0)$ would be 1. However, we rarely can choose such bases a priori. Instead, we can try to augment $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ with extra observables and use the singular value decomposition to choose a well conditioned basis. This brings us to the SVD implementation of Algorithm 3.

Algorithm 4: SVD implemented Noise Resistant DMD

Let $\tilde{\mathbf{f}} \in \mathscr{H}_+^{l_1}$, and $\tilde{\mathbf{g}} \in \mathscr{H}_-^{l_2},\ l_1, l_2 \geq k$ be noisy observables on our system. Fix the length of the time shift $s \geq 0$. 1: Construct the data matrices

$$X = \begin{bmatrix} \tilde{\mathbf{f}}(0) & \tilde{\mathbf{f}}(1) & \dots & \tilde{\mathbf{f}}(n-1) \end{bmatrix},$$

$$Y = \begin{bmatrix} \tilde{\mathbf{f}}(1) & \tilde{\mathbf{f}}(2) & \dots & \tilde{\mathbf{f}}(n) \end{bmatrix},$$

and

$$Z = \begin{bmatrix} \tilde{\mathbf{g}}(0) & \tilde{\mathbf{g}}(1) & \dots & \tilde{\mathbf{g}}(n-1) \end{bmatrix}.$$

2: Form the matrices $\tilde{G}_0 = \frac{1}{n} X Z^*$ and $\tilde{G}_1 = \frac{1}{n} Y Z^*$.

3: Compute the truncated SVD of $\tilde{G}_0$ using the first k singular values:

$$\tilde{G}_0 \approx W_k S_k V_k^*.$$

5: Form the matrix

$$A = S_k^{-1} W_k^* \tilde{G}_1 V_k.$$

6: Compute the eigenvalues and left and right eigenvectors, (λ_i, w_i, u_i) of A. The dynamic eigenvalues are λ_i, the dynamic modes are

$$v_i = W_k S_k u_i,$$

and the numerical eigenfunctions are

$$\hat{\phi}_i = w_i S_k^{-1} W_k^* X.$$

Similar to Algorithm 2, Algorithm 4 uses the SVD to choose a basis of observables to use in Algorithm 1. It is equivalent to performing Algorithm 3 using data from the observable $(S_k^{-1} W_k^*)\tilde{\mathbf{f}}$, while leaving $\tilde{\mathbf{g}}$ unchanged. It is important to note that Algorithm 4 uses the components of $(S_k^{-1} W_k^*)\mathbf{f}$ to as a basis for $\mathscr{F}$, so when we add observables to $\tilde{\mathbf{f}}$, we need their means to lie in our invariant subspace. One way to guarantee this is to only use time delays of our original observables.

7.2 Augmented Dual Observables

Typically, augmenting $\tilde{\mathbf{f}}$ with extra observables and using Algorithm 4 to truncate the singular values is an effective way to improve the conditioning of the problem. However, we have an alternate tool at our disposal. While each component of $\mathbf{f}$ must lie within $\mathscr{F}$, the components of $\mathbf{g}$ can be arbitrary, and we do not need to take an SVD to truncate the extra observables in $\mathbf{g}$. Since we do not need to worry about leaving our invariant subspace, we can add arbitrary functions of $\tilde{\mathbf{g}}$ (e.g. powers of $\tilde{\mathbf{g}}$) to our dual observable and still expect convergence. However, while this can improve conditioning, it also can

slow down the convergence of the time averages, so should only be done when the error stems from poor conditioning.

8 Numerical Examples

8.1 Random Rotation on a Circle

Consider a rotation on the circle. The dynamical system is defined by

$$x_{t+1} = x_t + \nu, \tag{16}$$

where $\nu \in S^1$. If we perturb (16) by adding noise to the rotation rate we obtain the random system

$$x_t + 1 = x_t + \nu + \pi(\omega_t) \tag{17}$$

where $\pi(\omega_t) \in S^1$ is an i.i.d. random variable. For the stochastic Koopman operator associated with (17), the functions $\varphi_n(x) = e^{inx}$ are eigenfunctions with eigenvalues $\lambda_i = \mathbb{E}(e^{in(\nu+\pi(\omega)}))$, since

$$\mathcal{K}\varphi_i(x) = \mathbb{E}(\varphi_i(T_\omega x)) = \int_\Omega e^{in(x+\nu+\pi(\omega))} dP = e^{inx} \int_\Omega e^{in(\nu+\pi(\omega))} dP = \varphi_i(x)\lambda_i.$$

We can compare these eigenvalues with the results obtained from our different DMD algorithms. We will set our system parameter to $\nu = \frac{1}{2}$ and draw $\pi(\omega_t)$ from the uniform distribution over $[-\frac{1}{2}, \frac{1}{2}]$. In this case the eigenvalues are $\lambda_i = \frac{i-ie^{in}}{n}$. For the first test, we will compare Algorithms 1 and 3 using a set of observables with measurement noise. We will let our observable be

$$\tilde{\mathbf{f}}(t) = [\sin(x_t), ..., \sin(5x_t), \cos(x_t), ..., \cos(5x_t)]^T + \mathbf{m}(t), \tag{18}$$

where $\mathbf{m}(t) \in [-0.5, 0.5]^{10}$ is measurement noise drawn from the uniform distribution. Algorithm 1 is applied directly to the data from measurements of $\tilde{\mathbf{f}}$, and for Algorithm 3 we let $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-1)$.

For the second test, we let $\tilde{f} = \sin(x) + \sin(2x) + \sin(3x) + \sin(4x)$, and use time delays to generate $\tilde{\mathbf{f}}$:

$$\tilde{\mathbf{f}}(t) = \begin{bmatrix} \tilde{f}(t) & \tilde{f}(t+1) & \dots & \tilde{f}(t+d). \end{bmatrix} \tag{19}$$

To perform Algorithms 1 and 3 we would use five time delays ($d = 5$) so the components of $\mathbf{f}$ span an invariant subspace. We perform Algorithm 1 directly from the data, while for Algorithm 3, we set $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-5)$. To combat the poor conditioning associated with Hankel DMD, Algorithm 4 is performed using using $\tilde{f}$ and 24 time delays (setting $d = 24$ in (19) and letting $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-24)$. The SVD is truncated to include only the first six singular values to compute the DMD operator. Finally, we run Algorithm 3 a second time with an augmented dual observable $\tilde{\mathbf{g}}$ while letting $\tilde{\mathbf{f}}$ contain $\tilde{f}$ and five time delays as before. For the augmented dual observable, we let $\tilde{\mathbf{g}}$ contain $\tilde{f}$, $\tilde{f}^2$, and $\tilde{f}^3$, as well as 44 time delays of each of these functions.

As can be seen in Figure 8.1, Algorithm 1 fails to accurately approximate the eigenvalues of $\mathcal{K}$ in both tests. For the first test, Algorithm 3 gives very accurate approximations to the eigenvalues of $\mathcal{K}$. Using the time delayed observables, (19), however, gives less accurate eigenvalues. This is because the conditioning of the matrix G_j is very poor, which amplifies the errors in our time averages. When we add extra time delays to $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ for Algorithm 4, and truncate to the leading singular values, we obtain much more accurate results. Alternatively, if we add extra functions and time delays to $\tilde{\mathbf{g}}$ and leave $\tilde{\mathbf{f}}$ unchanged, we also accurately capture the eigenvalues.

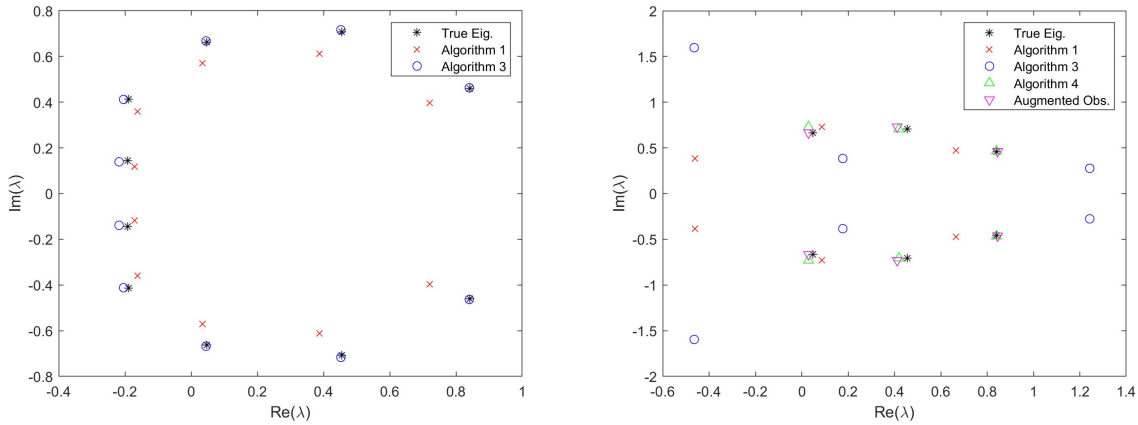

Figure 1: (Left) Outputs of Algorithm 1 and Algorithm 3 using (18) as observables on (17) with 20,000 data points. Algorithm 1 shows a clear error in the approximate eigenvalues while Algorithm 3 captures them accurately.
(Right) DMD outputs from Algorithms 1, 3, and 4 using (19) as observables on (17) with 20,000 data points. Algorithm 1 shows a bias in the eigenvalues while Algorithm 3 fails to accurately capture the eigenvalues due to conditioning issues. Algorithm 3 using the Augmented observables and Algorithm 4 both combat the conditioning issues and accurately capture the eigenvalues.

8.2 Linear System with Additive Noise

Consider the linear system in $\mathbb{R}^4$:

$$\mathbf{x}(t+1) = \begin{bmatrix} 0.75 & 0.5 & 0.1 & 2 \\ 0 & 0.2 & 0.8 & 1 \\ 0 & -0.8 & 0.2 & 0.5 \\ 0 & 0 & 0 & -0.85 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} = A\mathbf{x}(t). \tag{20}$$

If we perturb (20) by including a random forcing term b, we obtain

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + b_t, \tag{21}$$

where $b_t \in \mathbb{R}^4$ is assumed to be an i.i.d. random variable. Let $(w_i, \lambda_i), i = 1, ..., 4$ be the left eigenpairs of A. If b_t is assumed to have zero mean, $w_i^T \mathbf{x}$ is an eigenfunction of $\mathcal{K}$ with eigenvalue λ_i. For this example we will assume each component of b_t is drawn from

randomly from the uniform distribution on $[-0.5, 0.5]$. As before, we will test Algorithms 1 and 3 using observables with measurement noise and time delayed observables. For the first test, we will use a state observables with Gaussian measurement noise:

$$\tilde{\mathbf{f}}(t) = \mathbf{x}(t) + \mathbf{m}(t) \tag{22}$$

where each component of $\mathbf{m}(t) \in \mathbb{R}^4$ is drawn from the standard normal distribution. As before, will let $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-1)$.

For the second test, to generate the time delayed observables, we will let $\tilde{f}(x) = x_1(x) + x_2(x) + x_3(x) + x_4(x)$, an use three time delays:

$$\tilde{\mathbf{f}}(t) = \begin{bmatrix} \tilde{f}(t) & \tilde{f}(t+1) & \tilde{f}(t+2) & \tilde{f}(t+3) \end{bmatrix}. \tag{23}$$

We will apply Algorithm 1 directly to this matrix, while for Algorithm 3 we let $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-3)$.

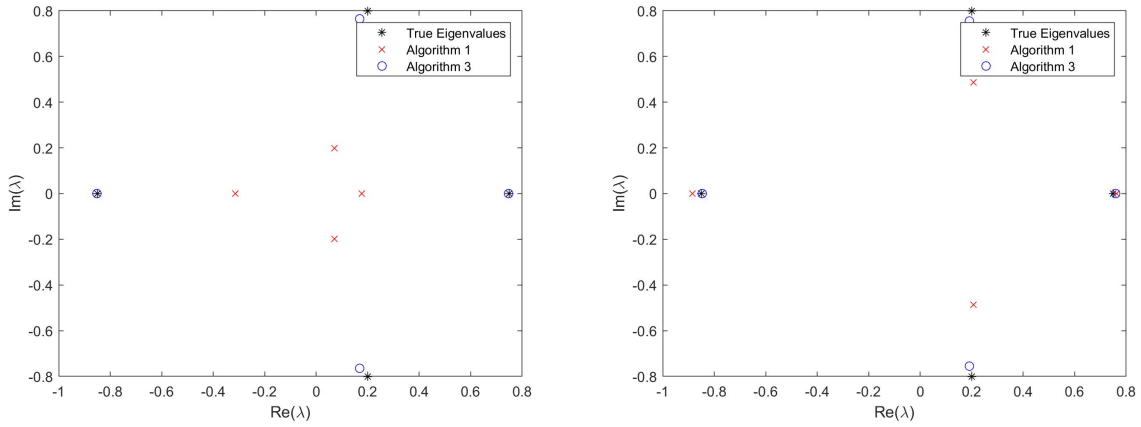

Figure 2: (Left) Outputs of Algorithm 1 and Algorithm 3 using state observables with measurement noise (22) on 10,000 data points from (21).
(Right) Outputs of Algorithm 1 and Algorithm 3 using (23) as observables on (21) with 10,000 data points. For both cases, Algorithm 3 accurately captures the eigenvalues while Algorithm 1 fails to do so.

Figure 8.2 shows that the eigenvalues generated by Algorithm 1 again fail to accurately

approximate those of $\mathcal{K}$. However, for both sets of observables, Algorithm 3 estimates the eigenvalues of $\mathcal{K}$ well. Since we did not run into conditioning issues, we did not test the results using Algorithm 4 or an augmented dual observable.

8.3 Stuart Landau Equations

Consider the stochastic Stuart Landau equations defined by

$$dr = (\delta r - r^3 + \frac{\epsilon^2}{r})dt + \epsilon dW_r \tag{24}$$

$$d\theta = (\gamma - \beta r^2)dt + \frac{\epsilon}{r} dW_\theta, \tag{25}$$

where W_r and W_θ satisfy

$$dW_r = \cos\theta \, dW_x + \sin\theta \, dW_y$$

$$dW_\theta = -\sin\theta \, dW_x + \cos\theta \, dW_y$$

for independent Wiener processes dW_x and dW_y. It was shown in [16] that for small ϵ and $\delta > 0$, the (continuous time) stochastic Koopman eigenvalues are given by

$$\lambda_{l,n} = \begin{cases} -\frac{n^2\epsilon^2(1+\beta^2)}{2\delta} + in\omega_0 + \mathcal{O}(\epsilon^4) & l = 0 \\ -2l\delta + in\omega_0 + \mathcal{O}(\epsilon^2) & l > 0, \end{cases}$$

where $\omega_0 = \gamma - \beta\delta$.

Let $\gamma = \beta = 1$, $\delta = 1/2$, and $\epsilon = 0.05$ in (24) and (25). Define the observables

$$f_k(r, \theta) = e^{ik(\theta - (\log(2r))}.$$

First, we will let

$$\tilde{\mathbf{f}}(t) = [f_1(t), f_{-1}(t), ..., f_6(t), f_{-6}(t)]^T + \mathbf{m}_1(t) + i\mathbf{m}_2(t), \tag{26}$$

where each component of $\mathbf{m}_1(t)$ and $\mathbf{m}_2(t)$ is drawn independently from a normal distribution with mean 0 and variance 1/4. In Algorithm 3, we let $\tilde{\mathbf{g}}(t) = \tilde{\mathbf{f}}(t-1)$. The (continuous time) eigenvalues generated by Algorithms 1 and 3 are shown from a simulation with 10,000 data points with a time step of 0.05 in Figure 8.3.

To test Hankel DMD, we use the observable

$$f = \sum_{k=1}^{6} (f_k + f_{-k}),$$

and let $\tilde{\mathbf{f}}$ contain f and d time delays of f:

$$\tilde{\mathbf{f}}(t) = \begin{bmatrix} f(t) & f(t+1) & \dots & f(t+d) \end{bmatrix}. \tag{27}$$

Due to the poor conditioning of Algorithms 1 and 3, the eigenvalues they generate are highly innaccurate, so we instead implement Algorithms 2 and 4. In each case, we let $d = 399$ and truncate the SVD to the leading 12 singular values. As usual, we let $\tilde{\mathbf{g}} = \tilde{\mathbf{f}}(t-d)$ in Algorithm 4. The results shown in Figure 8.3 are from a simulation with 100,000 data points and a time step of 0.05.

As can be seen in Figure 8.3, Algorithm 1 exhibits a clear bias towards the left of the complex plane using observables with measurement noise, although it appears to accurately estimate the imaginary part of the eigenvalue. Algorithm 3, on the other hand, appears to give a mostly accurate spectrum. When using time delayed observables for Hankel DMD, Algorithms 1 and 3 were very poorly conditioned, and gave eigenvalues far outside the windows shown in Figure 8.3. When using Algorithms 2 and truncating to the 12 dominant singluar values, we again see that the imaginary parts of the eigenvalues seem

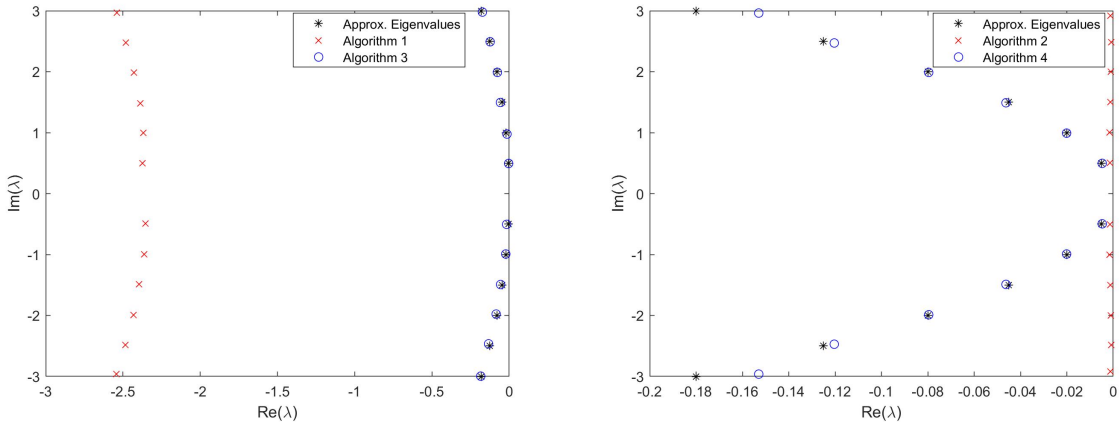

Figure 3: (Left) Outputs of Algorithm 1 and Algorithm 3 using observables with measurement noise (26). The data is taken over 100,000 data points from (24) and (25) with a time step of 0.05. The eigenvalues produced by Algorithm 1 are biased towards the left hand plane while Algorithm 1 captures them accurately. (Right) Outputs of Algorithm 2 and Algorithm 4 using (27) as observables on (24) and (25). The Algorithms used 100,000 data points with a time step of 0.05. Algorithm 4 captures most of the eigenvalues accurately while Algorithm 2 biases all eigenvalues towards the imaginary axis.

to be captured, but the real parts are all biased to the right. Algorithm 4, however, again captures the correct spectrum, but with some error for the most dissipative eigenvalues.

9 Conclusions

In this paper we analyzed the convergence of DMD algorithms for random dynamical systems, culminating in the introduction of a new DMD algorithm that converges to the spectrum of the stochastic Koopman operator in the presence of both random dynamics and noisy observables. We then specialized the algorithm to handle observables with i.i.d. measurement noise and time delayed observables and showed that measurements of a single set of observables was sufficient to generate an approximation of the stochastic Koopman operator. In particular, we demonstrated that a single trajectory of a single observable could be used to generate a Krylov subspace of the operator, which allows us to use DMD without needing to choose a basis of observables.

This algorithm provides a method for modeling complex systems where a deterministic model is unfeasible. This could be because a full state model would be to complex, observables of the full state are unavailable, or measurements come with uncertainty. A possible extension of this algorithm could adapt it to handle data from systems with control inputs, which could be used to develop control algorithms for random dynamical systems.

References

[1] Hassan Arbabi and Igor Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, 2017.

[2] Ludwig Arnold. *Random Dynamical Systems*. Springer-Verlag Berlin Heidelberg, 1998.

[3] Ludwig Arnold and Hans Crauel. Random dynamical systems. In *Lyapunov exponents*, pages 1–22. Springer, 1991.

[4] Nelida Črnjarić-Žic, Senka Maćešić, and Igor Mezić. Koopman operator spectrum for random dynamical systems. *Journal of Nonlinear Science*, pages 1–50, 2019.

[5] Scott TM Dawson, Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.

[6] I.I. Gikhman, A.V. Skorokhod, and S. Kotz. *The Theory of Stochastic Processes: I.* Classics in Mathematics. Springer Berlin Heidelberg, 2004.

[7] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the national academy of sciences of the united states of america*, 17(5):315, 1931.

[8] I Mezic and Andrzej Banaszuk. Comparison of systems with complex behavior: Spectral methods. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, volume 2, pages 1224–1231. IEEE, 2000.

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

[10] Igor Mezić. Spectrum of the koopman operator, spectral expansions in functional spaces, and state-space geometry. *Journal of Nonlinear Science*, pages 1–55, 2019.

[11] Igor Mezic. On numerical approximations of the koopman operator. *arXiv preprint arXiv:2009.05883*, 2020.

[12] Taijiro Ohno. Asymptotic behaviors of dynamical systems with random parameters. *Publications of the research Institute for Mathematical Sciences*, 19(1):83–98, 1983.

[13] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, Dans Henningson, et al. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641(1):115–127, 2009.

[14] Peter Schmid and Joern Sesterhenn. Dynamic mode decomposition of numerical and experimental data. *APS*, 61:MR–007, 2008.

[15] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Subspace dynamic mode decomposition for stochastic koopman analysis. *Physical Review E*, 96(3):033310, 2017.

[16] A Tantet, MD Chekroun, HA Dijkstra, and JD Neelin. Mixing spectrum in reduced phase spaces of stochastic differential equations. *Part II: Stochastic Hopf Bifurcation. ArXiv e-prints*, 2017.