



Beyond expectations: residual dynamic mode decomposition and variance for stochastic dynamical systems

Matthew J. Colbrook · Qin Li · Ryan V. Raut · Alex Townsend

Received: 6 September 2023 / Accepted: 22 November 2023
© The Author(s) 2023

Abstract Koopman operators linearize nonlinear dynamical systems, making their spectral information of crucial interest. Numerous algorithms have been developed to approximate these spectral properties, and dynamic mode decomposition (DMD) stands out as the poster child of projection-based methods. Although the Koopman operator itself is linear, the fact that it acts in an infinite-dimensional space of observables poses challenges. These include spurious modes, essential spectra, and the verification of Koopman mode decompositions. While recent work has addressed these challenges for deterministic systems, there remains a notable gap in verified DMD methods for stochastic systems, where the Koopman operator measures the expectation of observables. We show that

it is necessary to go beyond expectations to address these issues. By incorporating variance into the Koopman framework, we address these challenges. Through an additional DMD-type matrix, we approximate the sum of a squared residual and a variance term, each of which can be approximated individually using batched snapshot data. This allows verified computation of the spectral properties of stochastic Koopman operators, controlling the projection error. We also introduce the concept of variance-pseudospectra to gauge statistical coherency. Finally, we present a suite of convergence results for the spectral information of stochastic Koopman operators. Our study concludes with practical applications using both simulated and experimental data. In neural recordings from awake mice, we demonstrate how variance-pseudospectra can reveal physiologically significant information unavailable to standard expectation-based dynamical models.

M. J. Colbrook (✉)
DAMTP, University of Cambridge, Cambridge CB3 0WA, UK
e-mail: m.colbrook@damtp.cam.ac.uk

Q. Li
Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, USA
e-mail: qinli@math.wisc.edu

R. V. Raut
Allen Institute, Seattle, WA 98109, USA
e-mail: ryan.raut@alleninstitute.org

Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, USA

A. Townsend
Department of Mathematics, Cornell University, Ithaca, NY 14853, USA
e-mail: townsend@cornell.edu

Keywords Dynamical systems · Koopman operator · Data-driven discovery · Dynamic mode decomposition · Spectral theory · Error bounds · Stochastic systems

Mathematics Subject Classification 37M10 · 37H99 · 37N25 · 47A10 · 47B33 · 65P99

1 Introduction

Stochastic dynamical systems are widely used to model and study systems that evolve under the influence of both deterministic and random effects. They offer a framework for understanding, predicting, and controlling systems exhibiting randomness. This makes them invaluable across various scientific, engineering, and economic applications.

Given a state-space $\Omega \subset \mathbb{R}^d$ and a sample space Ω_s , we consider a discrete-time stochastic dynamical system

$$\mathbf{x}_n = F(\mathbf{x}_{n-1}, \tau_n), \quad n \geq 1, \quad \mathbf{x}_n \in \Omega, \quad (1)$$

where $\{\tau_n\}_{n \in \mathbb{N}} \in \Omega_s$ are independent and identically distributed (i.i.d.) random variables with distribution ρ supported on Ω_s , $\mathbf{x}_0 \in \Omega$ is an initial condition, and $F : \Omega \times \Omega_s \rightarrow \Omega$ is a function. In many applications, the function F is unknown or cannot be studied directly, which is the premise of this paper. We adopt the notation $F_\tau(\mathbf{x}) = F(\mathbf{x}, \tau)$ for convenience and express $\mathbf{x}_n = (F_{\tau_n} \circ \dots \circ F_{\tau_1})(\mathbf{x}_0)$, where ‘ \circ ’ denotes the composition of functions.

With the assumptions above, equation (1) describes a discrete-time Markov process. For such systems, the Kolmogorov backward equation governs the evolution of an observable [34,40], with the right-hand side defined as the stochastic Koopman operator [51]. The works [51,57] have spurred increased interest in the data-driven approximation of both deterministic and stochastic Koopman operators and in analyzing their spectral properties [11,43,54]. Prominent applications span a variety of fields including fluid dynamics [31,52,66,68], epidemiology [64], neuroscience [9,14,47], finance [46], robotics [6,8], power systems [75,76], and molecular dynamics [39,59,69,70].

Although the function F is usually nonlinear, the stochastic Koopman operator is always *linear*; however, it operates on an infinite-dimensional space of observables. Of particular interest is the spectral content of the Koopman operator near the unit circle, which corresponds to slow subspaces encapsulating the long-term dynamics. If finite-dimensional eigenspaces can capture this spectral content effectively, they can serve as a finite-dimensional approximation. Numerous algorithms have been developed to approximate the spectral properties of Koopman operators [1,2,10,12,26,30,42,48,52,55]. Among these, dynamic mode decomposi-

tion (DMD) is particularly popular [44]. Initially introduced in the fluids community [67,68], DMD’s connection to the Koopman operator was established in [66]. Since then, several extensions and variants of DMD have been developed [4,15,19,63,84,85], including methods tailored for stochastic systems [24,72,82,87].

At its core, DMD is a projection method. It is widely recognized that achieving convergence and meaningful applications of DMD can be challenging due to the infinite-dimensional nature of Koopman operators [12,23,37,84]. Challenges include the presence of spurious (unphysical) modes resulting from projection, essential spectra,¹ the absence of non-trivial finite-dimensional invariant subspaces, and the verification of Koopman mode decompositions (KMDs). Residual Dynamic Mode Decomposition (ResDMD) has been introduced to address these issues for deterministic systems [20,23]. ResDMD facilitates a data-driven approach to compute residuals associated with the full infinite-dimensional Koopman operator, thus enabling the computation of spectral properties with controlled errors and the verification of learned dictionaries and KMDs. Despite the evident importance of analyzing stochastic systems through the Koopman perspective, similar verified DMD methods in this setting are absent.

This paper presents several infinite-dimensional techniques for the data-driven analysis of stochastic systems. The central concept we explore is going beyond expectations to include higher moments within the Koopman framework. Figure 1 illustrates this point by depicting the evolution of two eigenfunctions associated with the stochastic Van der Pol oscillator (detailed in Sect. 5.2), alongside the expectation determined by the stochastic Koopman operator. Both eigenvalues and eigenfunctions are computed with a negligible projection error.² Notably, although both corresponding eigenvalues oscillate at the same frequency due to having identical arguments, the variances of the trajectories exhibit significant differences. This diver-

¹ For an illustrative example of a transition operator with non-trivial essential spectra, refer to [3]. If the operator in question is either self-adjoint or an L^2 isometry, the methodologies described in [18,21] and [23] respectively, can be applied to compute spectral measures.

² Here, ‘projection error’ refers to the error incurred when projecting the infinite-dimensional Koopman operator onto a finite-dimensional space of observables.

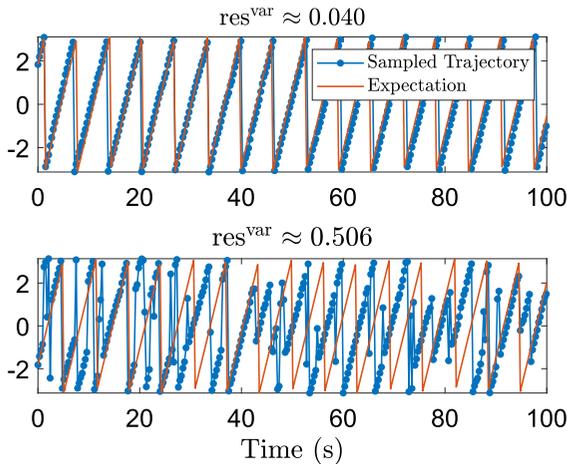


Fig. 1 The evolution of two eigenfunctions on the attractor of the stochastic Van der Pol oscillator from Sect. 5.2. The plots show the arguments. In blue, we see a sample of the true trajectories, while the expected values predicted from the stochastic Koopman operator are shown in red. Top: Eigenfunction associated with $m = 0$ and $k = 1$ in Table 1. The variance residual is small, and trajectories hug the expectation closely. Bottom: Eigenfunction associated with $m = 1$ and $k = 1$ in Table 1. The variance residual is large, and trajectories deviate from the expectation

gence is quantified by what we define as a *variance residual* (see Sect. 3.2).

1.1 Contributions

The contributions of our paper are as follows:

- *Variance Incorporation*: We integrate the concept of variance into the Koopman framework and establish its relationship with batched Koopman operators. Proposition 2 decomposes a mean squared Koopman error into an infinite-dimensional residual and a variance term. Additionally, we present methodologies (see Algorithms 1 and 2) for independently calculating these components, thereby enhancing the understanding of the spectral properties of the Koopman operator and the deviation from mean dynamics.
- *Variance-Pseudospectra*: We introduce a novel concept of pseudospectra, termed *variance-pseudospectra* (see Definition 2), which serves as a measure of statistical coherency.³ We also offer

³ In the setting of dynamical systems, coherent sets or structures are subsets of the phase space where elements (e.g., particles,

algorithms for computing these pseudospectra (see Algorithms 3 and 4) and prove their convergence.

- *Convergence Theory*: Sect. 4 of our paper is dedicated to proving a suite of convergence theorems. These pertain to the spectral properties of stochastic Koopman operators, the accuracy of KMD forecasts, and the derivation of concentration bounds for estimating Koopman matrices from a finite set of snapshot data.

Various examples are given in Sect. 5 and code is available at: <https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition>.

1.2 Previous work

Existing literature on stochastic Koopman operators primarily addresses the challenge of noisy observables in extended dynamic mode decomposition (EDMD) methodologies [82], and in techniques for debiasing DMD [27, 35, 77]. A related concern is the estimation error in Koopman operator approximations due to the finite nature of data sets. This issue is present in both deterministic and stochastic scenarios. As [84] describes, EDMD converges with large data sets to a Galerkin approximation of the Koopman operator. The work in [58] thoroughly analyzes kernel autocovariance operators, including nonasymptotic error bounds under classical ergodic and mixing assumptions. In [60], the authors offer the first comprehensive probabilistic bounds on the finite-data approximation error for truncated Koopman generators in stochastic differential equations (SDEs) and nonlinear control systems. They examine two scenarios: (1) i.i.d. sampling and (2) ergodic sampling, with the latter assuming exponential stability of the Koopman semigroup. Additionally, the variational approach to conformational dynamics (VAC), which bears similarities to DMD, is known for providing spectral estimates of time-reversible processes that result in a self-adjoint transition operator. The connection of VAC with Koopman operators is detailed in [83], and the approximation of spectral information with error bounds is discussed in [39].

agents, etc.) exhibit similar behavior over some time interval. This behavior remains relatively consistent despite potential perturbations or the chaotic nature of the system. In essence, within a coherent structure, the dynamics of elements are closely linked and evolve coherently.

1.3 Data-driven setup

We present data-driven methods that utilize a dataset of “snapshot” pairs alongside a dictionary of observables. While numerous approaches for selecting a dictionary exist in the literature [17, 32, 80–82, 84, 85], this topic is not the primary focus of our current study.⁴ Following the methodology outlined in [79], we consider our given data to consist of pairs of snapshots, which are

$$S = \left\{ (\mathbf{x}^{(m)}, \mathbf{y}^{(m)}) \right\}_{m=1}^M, \quad \mathbf{y}^{(m)} = F(\mathbf{x}^{(m)}, \tau_m). \quad (2)$$

Unlike in deterministic systems, for stochastic systems, it can be beneficial for S to include the same initial condition $\mathbf{x}^{(m)}$ multiple times, as each execution of the dynamics yields an independent realization of a trajectory. We say that S is M_1 -batched if it can be split into M_1 subsets such that

$$S = \cup_{j=1}^{M_1} S_j, \\ S_j = \{(\mathbf{x}^{(j)}, \mathbf{y}^{(j,k)}) : k = 1, \dots, M_2, \mathbf{y}^{(j,k)} = F(\mathbf{x}^{(j)}, \tau_{j,k})\}.$$

In other words, for each $\mathbf{x}^{(j)}$, we have multiple realizations of $F_\tau(\mathbf{x}^{(j)})$. Using batched data, we can approximate higher-order stochastic Koopman operators representing the moments of the trajectories. An unbatched dataset can be adapted to approximate a batched dataset by categorizing or “binning” the \mathbf{x} points in the snapshot data. In practical scenarios, one may encounter a combination of both batched and unbatched data. Depending on the type of snapshot data used, Galerkin approximations of stochastic Koopman operators can be achieved in the limit of large datasets (as discussed in Sect. 2.2).

2 Mathematical preliminaries

This section discusses several foundational concepts upon which our paper builds.

⁴ ResDMD has been shown to effectively verify learned dictionaries in deterministic dynamical systems [20].

2.1 The stochastic Koopman operator

Let $g : \Omega \rightarrow \mathbb{C}$ be a function, commonly called an *observable*. Given an initial condition $\mathbf{x}_0 \in \Omega$, measuring the initial state of the dynamical system through g yields the value $g(\mathbf{x}_0)$. One time-step later, the measurement $g(\mathbf{x}_1) = g(F_\tau(\mathbf{x}_0)) = (g \circ F_\tau)(\mathbf{x}_0)$ is obtained, where τ is a realization from a probability distribution supported on Ω_s , i.e., $\tau \sim \rho$. The “pull-back” operator, given g , outputs the “look ahead” measurement function $g \circ F_\tau$. This function is a random variable, and the stochastic Koopman operator is its expectation [56]:

$$\mathcal{K}_{(1)}[g] = \mathbb{E}_\tau [g \circ F_\tau] = \int_{\Omega_s} g \circ F_\tau \, d\rho(\tau). \quad (3)$$

Here, \mathbb{E}_τ represents the expectation with respect to the distribution ρ . The subscript (1) indicates this is the first moment. Throughout the paper, we assume that the domain of the operator $\mathcal{K}_{(1)}$ is $L^2(\Omega, \omega)$, where ω is a positive measure on Ω . This space is equipped with an inner product and norm, denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We do not assume that $\mathcal{K}_{(1)}$ is compact or self-adjoint.

We now introduce the batched Koopman operator, designed to capture the variance and other higher-order moments in the trajectories of dynamical systems. For $r \in \mathbb{N}$ and $g : \Omega^r \rightarrow \mathbb{C}$, we define

$$\mathcal{K}_{(r)}[g] = \mathbb{E}_\tau [g(F_\tau, \dots, F_\tau)], \quad (4)$$

where the same realization $\tau \sim \rho$ is used for the r arguments of g . Notably, both the classical and the batched versions of the Koopman operators adhere to the semigroup property, as we will demonstrate.

Proposition 1 For any $r, n \in \mathbb{N}$,

$$\mathcal{K}_{(r)}^n[g] = \mathbb{E}_{\tau_1, \dots, \tau_n} [g(F_{\tau_n} \circ \dots \circ F_{\tau_1}, \dots, F_{\tau_n} \circ \dots \circ F_{\tau_1})].$$

Proof For $r = 1$, see [24]. For $r > 1$, note that $\mathcal{K}_{(r)}$ is a first-order Koopman operator of a dynamical system on Ω^r . \square

This proposition indicates that n applications of the stochastic Koopman operator yield the expected value of an observable after n time steps. It is crucial to understand that $\mathcal{K}_{(1)}$ only calculates the expected value. To gain insights into the variability around this mean and to

understand the projection error inherent in DMD methods, we need to consider higher-order statistics, such as the variance. These aspects are further explored in Sect. 3.

2.2 Extended dynamic mode decomposition

EDMD is a widely-used method for constructing a finite-dimensional approximation of the Koopman operator $\mathcal{K}_{(1)}$, utilizing the snapshot data S in (2). This approach involves projecting the infinite-dimensional Koopman operator onto a finite-dimensional matrix and approximating its entries. For notational simplicity, we will omit the subscript (1) when referring to the Koopman operator in this section. Originally, EDMD assumes that the initial conditions are independently drawn from a distribution ω [84]. However, in our adaptation, we apply EDMD to any given S , treating the $\mathbf{x}^{(m)}$ as quadrature nodes for integration with respect to ω . This flexibility allows us to use different quadrature weights depending on the specific scenario.

One first chooses a dictionary $\{\psi_1, \dots, \psi_N\}$ in the space $L^2(\Omega, \omega)$. This dictionary consists of a list of observables that form a finite-dimensional subspace $V_N = \text{span}\{\psi_1, \dots, \psi_N\}$. EDMD computes a matrix $K \in \mathbb{C}^{N \times N}$ that approximates the action of \mathcal{K} within this subspace. Specifically, the goal is to achieve $K = \mathcal{P}_{V_N} \mathcal{K} \mathcal{P}_{V_N}^*$, where $\mathcal{P}_{V_N} : L^2(\Omega, \omega) \rightarrow V_N$ is the orthogonal projection onto V_N . In the Galerkin framework, this equates to:

$$\langle \mathcal{K}[\psi_j], \psi_i \rangle = \sum_{s=1}^N K_{s,j} \langle \psi_s, \psi_i \rangle, \quad 1 \leq i, j \leq N.$$

A matrix K satisfying this relationship is given by

$$K = G^\dagger A, \quad G_{i,j} = \langle \psi_j, \psi_i \rangle, \quad A_{i,j} = \langle \mathcal{K}[\psi_j], \psi_i \rangle.$$

Commonly, we stack the Ψ and define the feature map

$$\Psi(\mathbf{x}) = [\psi_1(\mathbf{x}) \cdots \psi_N(\mathbf{x})]^\top \in \mathbb{C}^{1 \times N}.$$

Then, for any $g \in V_N$, we use the shorthand $g = \Psi \mathbf{g}$ for $g(\mathbf{x}) = \sum_{j=1}^N g_j \psi_j(\mathbf{x})$. With the previously defined K , the approximation becomes

$$\mathcal{K}[g](\mathbf{x}) \approx \sum_{i=1}^N \left(\sum_{j=1}^N K_{i,j} g_j \right) \psi_i(\mathbf{x}) = \Psi(\mathbf{x}) K \mathbf{g}.$$

The accuracy of this approximation depends on how well V_N can approximate $\mathcal{K}g$.

The entries of the matrices G and A are inner products and must be approximated using the trajectory data S . For quadrature weights $\{w_m\}$, we define \tilde{G} as the numerical approximation of G :

$$\tilde{G}_{i,j} = \sum_{m=1}^M w_m \psi_j(\mathbf{x}^{(m)}) \overline{\psi_i(\mathbf{x}^{(m)})} \approx \langle \psi_j, \psi_i \rangle = G_{i,j}. \quad (5)$$

The weights $\{w_m\}$ reflect the significance assigned to each snapshot in the dataset, influenced by factors such as data distribution or reliability, which we will explore further. Similarly, for A , we define

$$\tilde{A}_{i,j} = \sum_{m=1}^M w_m \psi_j(\mathbf{y}^{(m)}) \overline{\psi_i(\mathbf{x}^{(m)})} \approx \langle \mathcal{K}[\psi_j], \psi_i \rangle = A_{i,j}. \quad (6)$$

Let $\Psi_X, \Psi_Y \in \mathbb{C}^{M \times N}$ collect the dictionary's evaluations of these samples:

$$\Psi_X = \begin{pmatrix} \Psi^\top(\mathbf{x}^{(1)}) \\ \vdots \\ \Psi^\top(\mathbf{x}^{(M)}) \end{pmatrix}, \quad \Psi_Y = \begin{pmatrix} \Psi^\top(\mathbf{y}^{(1)}) \\ \vdots \\ \Psi^\top(\mathbf{y}^{(M)}) \end{pmatrix}, \quad (7)$$

and let $W = \text{diag}(w_1, \dots, w_M)$. Then we can succinctly write

$$\tilde{G} = \Psi_X^* W \Psi_X, \quad \tilde{A} = \Psi_X^* W \Psi_Y. \quad (8)$$

Throughout this paper, the symbol \tilde{X} denotes an estimation of the quantity X .

Various sampling methods converge in the large data limit, meaning that

$$\lim_{M \rightarrow \infty} \tilde{G} = G, \quad \lim_{M \rightarrow \infty} \tilde{A} = A. \quad (9)$$

We detail three convergent sampling methods:

- (i) *Random sampling*: In the initial definition of EDMD, ω is a probability measure and $\{\mathbf{x}^{(m)}\}_{m=1}^M$ are independently drawn according to ω with each quadrature weight set to $w_m = 1/M$. The strong law of large numbers guarantees that (9) holds with probability one [38, Section 3.4] [41, Section 4]. Typically, convergence occurs at a Monte Carlo rate of $\mathcal{O}(M^{-1/2})$ [13].
- (ii) *Ergodic sampling*: If the stochastic dynamical system is ergodic, the Birkhoff–Khinchin theorem [33, Theorem II.8.1, Corollary 3] supports convergence using data from a single trajectory for almost every initial point. Specifically, we use: $\mathbf{x}^{(m+1)} = F(\mathbf{x}^{(m)}, \tau_m)$, $w_m = 1/M$. This sampling method’s analysis for stochastic Koopman operators is detailed in [82]. An advantage is that knowledge of ω is not required. However, the convergence rate depends on the specific problem [36]. Note that in an ergodic system, the stochastic Koopman operator is an isometry on $L^1(\Omega, \omega)$ but typically not on $L^2(\Omega, \omega)$.
- (iii) *High-order quadrature*: When the dictionary and F are sufficiently regular, and the dimension d is not too large, and if we can choose the $\{\mathbf{x}^{(m)}\}_{m=1}^M$, employing a high-order quadrature rule is advantageous. For deterministic systems, this approach can significantly increase convergence rates in (9) [23]. In stochastic systems, high-order quadrature applies primarily to batched snapshot data. We may select $\{\mathbf{x}^{(j)}\}_{j=1}^{M_1}$ based on an M_1 -point quadrature rule with associated weights $\{w_j\}_{j=1}^{M_1}$. Convergence is achieved as $M_2 \rightarrow \infty$, effectively applying Monte Carlo integration of the random variable τ over Ω_s for each fixed $\mathbf{x}^{(j)}$.

The convergence described in (9) implies that the eigenvalues obtained through EDMD converge to the spectrum of $\mathcal{P}_{V_N} \mathcal{K} \mathcal{P}_{V_N}^*$ as $M \rightarrow \infty$. Therefore, approximating the spectrum of \mathcal{K} , denoted $\text{Sp}(\mathcal{K})$, by the eigenvalues of \tilde{K} is closely related to the so-called finite section method [7]. However, just as the finite section method can be prone to spectral pollution, which refers to the appearance of spurious modes that accumulate even as the size of the dictionary increases, this is also a concern for EDMD [84]. Consequently, having a method to validate the accuracy of the proposed eigenvalue-eigenvector pairs becomes crucial, which is one of the key functions of ResDMD.

2.3 Residual dynamic mode decomposition (ResDMD)

Accurately estimating the spectrum of \mathcal{K} is critical for analyzing dynamical systems. For deterministic systems, ResDMD achieves this goal, providing robust spectral estimates [20,23]. Unlike classical DMD methods, ResDMD introduces an additional matrix specifically designed to approximate $\mathcal{K}^* \mathcal{K}$. This enhancement not only offers rigorous error guarantees for the spectral approximation but also enables a posteriori assessment of the reliability of the computed spectra and Koopman modes. This capability is particularly valuable in addressing issues such as spectral pollution, which are common challenges in DMD-type methods.

ResDMD is built around the approximation of residuals associated with \mathcal{K} , providing an error bound. For any given candidate eigenvalue-eigenvector pair (λ, g) , with $\lambda \in \mathbb{C}$ and $g = \Psi \mathbf{g} \in V_N$, one can consider the relative squared residual as follows:

$$\frac{\int_{\Omega} |\mathcal{K}[g](\mathbf{x}) - \lambda g(\mathbf{x})|^2 d\omega(\mathbf{x})}{\int_{\Omega} |g(\mathbf{x})|^2 d\omega(\mathbf{x})} = \frac{\langle \mathcal{K}[g], \mathcal{K}[g] \rangle - \lambda \langle g, \mathcal{K}[g] \rangle - \bar{\lambda} \langle \mathcal{K}[g], g \rangle + |\lambda|^2 \langle g, g \rangle}{\langle g, g \rangle} \tag{10}$$

This pair (λ, g) can be computed either from K or other methods. A small residual means that λ can be approximately considered as an eigenvalue of \mathcal{K} , with g as the corresponding eigenfunction. The relative residual in (10) serves as a measure of the *coherency* of observables, indicating that observables with smaller residuals play a significant role in the dynamics of the system. If the relative (non-squared) residual is bounded by ϵ , then $\mathcal{K}^n g = \lambda^n g + \mathcal{O}(n\epsilon)$. In other words, λ characterizes the coherent oscillation and the decay/growth in the observable g with time.

The residual is closely related to the notion of pseudospectra [78].

Definition 1 For any $\lambda \in \mathbb{C}$, define:

$$\sigma_{\text{inf}}(\lambda) = \inf \left\{ \|\mathcal{K}[g] - \lambda g\| : g \in L^2(\Omega, \omega), \|g\| = 1 \right\}.$$

For $\epsilon > 0$, the approximate point⁵ ϵ -pseudospectrum is

$$\text{Sp}_\epsilon(\mathcal{K}) = \text{Cl}(\{\lambda \in \mathbb{C} : \sigma_{\text{inf}}(\lambda) < \epsilon\}),$$

where Cl denotes closure of a set. Furthermore, we say that g is a ϵ -pseudoeigenfunction if there exists $\lambda \in \mathbb{C}$ such that the relative squared residual in (10) is bounded by ϵ^2 .

To compute (10), notice that three of the four inner products appearing in the numerator are:

$$\langle \mathcal{K}[g], g \rangle = \mathbf{g}^* A \mathbf{g}, \quad \langle g, \mathcal{K}[g] \rangle = \mathbf{g}^* A^* \mathbf{g}, \quad \langle g, g \rangle = \mathbf{g}^* G \mathbf{g}, \tag{11}$$

with A, G numerically approximated by EDMD (8). Hence, the success of the computation relies on finding a numerical approximation to $\langle \mathcal{K}[g], \mathcal{K}[g] \rangle$. To that end, we deploy the same quadrature rule discussed in (5)-(6) and set

$$L = [L_{i,j}], \quad L_{i,j} = \langle \mathcal{K}[\psi_j], \mathcal{K}[\psi_i] \rangle, \quad \tilde{L} = \Psi_Y^* W \Psi_Y, \tag{12}$$

then $\langle \mathcal{K}[g], \mathcal{K}[g] \rangle \approx \mathbf{g}^* \Psi_Y^* W \Psi_Y \mathbf{g} = \mathbf{g}^* \tilde{L} \mathbf{g}$. We obtain a numerical approximation of (10) as

$$[\text{res}(\lambda, g)]^2 = \frac{\mathbf{g}^* [\tilde{L} - \lambda \tilde{A}^* - \bar{\lambda} \tilde{A} + |\lambda|^2 \tilde{G}] \mathbf{g}}{\mathbf{g}^* \tilde{G} \mathbf{g}}. \tag{13}$$

The matrix L introduced by ResDMD formally corresponds to an approximation of $\mathcal{K}^* \mathcal{K}$. The computation utilizes the same dataset as that employed for \tilde{G} and \tilde{A} and is computationally efficient to construct. The work presented in [23] demonstrates that the approximation outlined in (13) can be effectively used in various algorithms for rigorously computing the spectra and pseudospectra of \mathcal{K} for deterministic systems. However, these results from [23] are not directly applicable to stochastic systems.

⁵ In the presence of residual spectrum, the full pseudospectrum requires the injection modulus of complex shifts of the adjoint of \mathcal{K} . We have refrained from this discussion for the sake of simplicity.

3 Variance from the Koopman perspective

When analyzing a system with inherent stochasticity, basing conclusions only on the mean trajectory can lead to misleading interpretations, as illustrated in Fig. 1. To achieve a more accurate statistical understanding of such systems, it is crucial to quantify how much and in what ways the trajectory deviates from this mean. This need for a more comprehensive analysis underpins our exploration into quantifying the variance.

3.1 Variance via Koopman operators

For any observable $g \in L^2(\Omega, \omega)$ and $\mathbf{x} \in \Omega$, $g(F_\tau(\mathbf{x}))$ is a random variable. One can define its moments:

$$\mathbb{E}_\tau[(g(F_\tau(\mathbf{x})))^r] = \int_{\Omega_s} [g(F_\tau(\mathbf{x}))]^r d\rho(\tau), \quad r \in \mathbb{N}.$$

Recalling the definitions in (4), this becomes:

$$\mathbb{E}_\tau[(g(F_\tau(\mathbf{x})))^r] = \mathcal{K}_{(r)}[g \otimes \dots \otimes g](\mathbf{x}, \dots, \mathbf{x}).$$

This means that the r -th order Koopman operator directly computes the moments of the trajectory. In particular, the combination of the first and the second moment provides the following variance term:

$$\begin{aligned} \text{Var}_\tau[g(F_\tau(\mathbf{x}))] &= \mathbb{E}_\tau[|g(F_\tau(\mathbf{x}))|^2] - |\mathbb{E}_\tau[g(F_\tau(\mathbf{x}))]|^2 \\ &= \mathcal{K}_{(2)}[g \otimes \bar{g}](\mathbf{x}, \mathbf{x}) - |\mathcal{K}_{(1)}[g](\mathbf{x})|^2. \end{aligned}$$

We integrate the local definition of variance over the entire domain to define:

$$\text{Var}_\tau[g(F_\tau)] = \int_{\Omega} \text{Var}_\tau[g(F_\tau(\mathbf{x}))] d\omega(\mathbf{x}). \tag{14}$$

The following proposition provides a Koopman analog of decomposing an integrated mean squared error (IMSE).

Proposition 2 *Let $g, h \in L^2(\Omega, \omega)$, then*

$$\begin{aligned} \mathbb{E}_\tau[\|g \circ F_\tau + h\|^2] &= \|\mathcal{K}_{(1)}[g] + h\|^2 + \int_{\Omega} \text{Var}_\tau[(g \circ F_\tau)(\mathbf{x})] d\omega(\mathbf{x}). \end{aligned} \tag{15}$$

Proof We expand $|g(F_\tau(\mathbf{x})) + h(\mathbf{x})|^2$ for a fixed $\mathbf{x} \in \Omega$ and take expectations to find that

$$\begin{aligned} \mathbb{E}_\tau \left[|g(F_\tau(\mathbf{x})) + h(\mathbf{x})|^2 \right] &= \mathbb{E}_\tau \left[|g(F_\tau(\mathbf{x}))|^2 \right] + \mathcal{K}_{(1)}[g](\mathbf{x})\overline{h(\mathbf{x})} \\ &\quad + h(\mathbf{x})\overline{\mathcal{K}_{(1)}[g](\mathbf{x})} + |h(\mathbf{x})|^2 \\ &= |\mathcal{K}_{(1)}[g](\mathbf{x}) + h(\mathbf{x})|^2 + \mathbb{E}_\tau \left[|g(F_\tau(\mathbf{x}))|^2 \right] \\ &\quad - |\mathbb{E}_\tau [g(F_\tau(\mathbf{x}))]|^2. \end{aligned}$$

The result now follows by integrating over \mathbf{x} with respect to the measure ω . \square

Similarly, for any two functions $g, h \in L^2(\Omega, \omega)$, we define the covariance:

$$\mathcal{C}(g, h) = \int_\Omega \mathbb{E}_\tau [(g \circ F_\tau - \mathcal{K}_{(1)}[g])\overline{(h \circ F_\tau - \mathcal{K}_{(1)}[h])}] d\omega(\mathbf{x}) \tag{16}$$

and obtain the following similar result using covariance:

$$\begin{aligned} \int_\Omega \mathbb{E}_\tau [g(F_\tau(\mathbf{x}))\overline{h(F_\tau(\mathbf{x}))}] d\omega(\mathbf{x}) &= \langle \mathcal{K}[g], \mathcal{K}[h] \rangle + \mathcal{C}(g, h). \end{aligned}$$

Proposition 2 is analogous to the decomposition of an IMSE and is practically useful. Suppose we use an observation h to approximate $-g \circ F_\tau$, in an attempt to minimize $\|g \circ F_\tau + h\|^2$. An unbiased estimator is $-\mathcal{K}_{(1)}[g]$; however, this approximation will not be perfect due to the variance term in (15). Therefore, there is a variance-residual tradeoff for stochastic Koopman operators. Depending on the type of trajectory data collected, one can approximate the quantities $\mathbb{E}_\tau [\|g \circ F_\tau + h\|^2]$ and $\|\mathcal{K}_{(1)}[g] + h\|^2$ in (15) and hence, estimate the third variance term.

Example 1 [Circle map] Let $\Omega = [0, 1]_{\text{per}}$ be the periodic interval and consider

$$F(\mathbf{x}, \tau) = \mathbf{x} + c + f(\mathbf{x}) + \tau \pmod{1},$$

where $\Omega_s = [0, 1]_{\text{per}}$, ρ is absolutely continuous, and c is a constant. Let $\psi_j(\mathbf{x}) = e^{2\pi i j \mathbf{x}}$ for $j \in \mathbb{Z}$. Then

$$\mathcal{K}_{(1)}[\psi_j](\mathbf{x}) = \psi_j(\mathbf{x}) e^{2\pi i j f(\mathbf{x})} e^{2\pi i j c} \int_{\Omega_s} e^{2\pi i j \tau} d\rho(\tau). \tag{17}$$

Define the constants

$$\alpha_j = e^{2\pi i j c} \int_{\Omega_s} e^{2\pi i j \tau} d\rho(\tau).$$

Let D be the operator that multiplies each ψ_j by α_j . Then $\mathcal{K}_{(1)} = TD$, where T is the Koopman operator corresponding to $\mathbf{x} \mapsto \mathbf{x} + f(\mathbf{x})$. Since ρ is absolutely continuous, the Riemann–Lebesgue lemma implies that $\lim_{|j| \rightarrow \infty} \alpha_j = 0$ and hence D is a compact operator. It follows that if T is bounded, then $\mathcal{K}_{(1)}$ is a compact operator. A straightforward computation using (14) shows that

$$\int_\Omega \text{Var}_\tau [\psi_j(F_\tau(\mathbf{x}))] d\omega(\mathbf{x}) = 1 - |\alpha_j|^2. \tag{18}$$

For example, if $f = 0$, $\mathcal{K}_{(1)}$ has pure point spectrum with eigenfunctions ψ_j . However, as $|j| \rightarrow \infty$, the variance converges to one and ψ_j become less statistically coherent. This example is explored further in Sect. 5.1. \square

Another immediate application of the variance term is in providing an estimated bound for the Koopman operator prediction of trajectories.

Proposition 3 *We have*

$$\begin{aligned} \mathbb{P} \left(|g \circ F_{\tau_n} \circ \dots \circ F_{\tau_1}(\mathbf{x}) - \mathcal{K}^n[g](\mathbf{x})| \geq a \right) &\leq \frac{1}{a^2} \text{Var}_{\tau_1, \dots, \tau_n} [g \circ F_{\tau_n} \circ \dots \circ F_{\tau_1}(\mathbf{x})] \\ &= \frac{1}{a^2} \left(\mathcal{K}_{(2)}^n [g \otimes \bar{g}](\mathbf{x}, \mathbf{x}) - |\mathcal{K}_{(1)}^n [g](\mathbf{x})|^2 \right) \end{aligned} \tag{19}$$

for any $a > 0$.

Proof the result follows from combining Proposition 1 and (14) with Chernoff’s bound. \square

The bound can be combined with concentration bounds for $\Psi \tilde{K}^n - \mathcal{K}^n$ (see Sect. 4.2).

3.2 ResDMD in stochastic systems

In the deterministic setting, ResDMD provides an efficient way to evaluate the accuracy of candidate eigenpairs through the computation of an additional matrix L in (12). However, what happens in the stochastic setting?

Suppose that (λ, g) is a candidate eigenpair of $\mathcal{K}_{(1)}$ with $g \in V_N$. Resembling (10), we consider

$$\frac{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]}{\|g\|^2}. \tag{20}$$

We can write the numerator in terms of $A, G,$ and $L,$ i.e.,

$$\begin{aligned} \mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2] &= \mathbf{g}^*(L - \lambda A^* - \bar{\lambda}A + |\lambda|^2 G)\mathbf{g} \\ &= \lim_{M \rightarrow \infty} \mathbf{g}^*(\tilde{L} - \lambda \tilde{A}^* - \bar{\lambda}\tilde{A} + |\lambda|^2 \tilde{G})\mathbf{g}. \end{aligned}$$

Hence, we define

$$[\text{res}^{\text{var}}(\lambda, g)]^2 = \frac{\mathbf{g}^* [\tilde{L} - \lambda \tilde{A}^* - \bar{\lambda}\tilde{A} + |\lambda|^2 \tilde{G}]\mathbf{g}}{\mathbf{g}^* \tilde{G}\mathbf{g}}, \tag{21}$$

which furnishes an approximation of (20). Setting $h = -\lambda g$ in Proposition 2, we see that

$$\begin{aligned} \mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2] &= \mathbb{E}_\tau \left[\int_\Omega |g(F_\tau(\mathbf{x})) - \lambda g(\mathbf{x})|^2 d\omega(\mathbf{x}) \right] \\ &= \underbrace{\|\mathcal{K}_{(1)}[g] - \lambda g\|^2}_{\text{squared residual}} + \underbrace{\int_\Omega \text{Var}_\tau [g(F_\tau(\mathbf{x}))] d\omega(\mathbf{x})}_{\text{integrated variance of } g \circ F_\tau}. \end{aligned} \tag{22}$$

Thus, $\text{res}^{\text{var}}(\lambda, g)$ approximates the sum of the squared residual $\|\mathcal{K}[g] - \lambda g\|^2$ and the integrated variance of $g \circ F_\tau$. For stochastic systems, the integrated variance of $g \circ F_\tau$ is usually nonzero so that

$$\lim_{M \rightarrow \infty} \text{res}^{\text{var}}(\lambda, g) > \|\mathcal{K}_{(1)}[g] - \lambda g\| \|g\|. \tag{23}$$

Based on this notion and drawing an analogy with Definition 1, we make the following definition.

Definition 2 For any $\lambda \in \mathbb{C}$, define:

$$\sigma_{\text{inf}}^{\text{var}}(\lambda) = \inf \left\{ \sqrt{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]} : g \in L^2(\Omega, \omega), \|g\| = 1 \right\}.$$

For $\epsilon > 0$, we define the variance- ϵ -pseudospectrum as

$$\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) = \text{Cl} \left(\{ \lambda \in \mathbb{C} : \sigma_{\text{inf}}^{\text{var}}(\lambda) < \epsilon \} \right),$$

where Cl denotes the closure of a set. Furthermore, we say that g is a variance- ϵ -pseudoeigenfunction if there exists $\lambda \in \mathbb{C}$ such that $\sqrt{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]} \leq \epsilon$.

Superficially, this definition is a straightforward extension of Definition 1. However, there are some essential differences. Both the conceptual understanding and the computation methods need to be modified.

First, the relation (22) shows that $\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$ takes into account uncertainty through the variance term. Hence, the variance-pseudospectrum provides a notion of *statistical coherency*. Furthermore, comparing Definitions 1 and 2, we have

$$\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) \subset \text{Sp}_\epsilon(\mathcal{K}_{(1)}).$$

If the dynamical system is deterministic, then $\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$ is equal to the approximate point ϵ -pseudospectrum. However, in the presence of variance, they are no longer equal.

Second, the relation (22) gives a computational surprise. Following the same derivation between (10)–(13), with $L, A,$ and G accordingly adjusted through replacing \mathcal{K} by $\mathcal{K}_{(1)}$ in (11)–(12), we can still compute the variance-residual term. However, the original residual itself, $\text{res}(\lambda, g)$, needs a modification. Recalling (10), in the same spirit of EDMD, if $g \in V_N$, we write

$$\begin{aligned} \|\mathcal{K}_{(1)}[g] - \lambda g\|^2 &= \langle \mathcal{K}_{(1)}[g], \mathcal{K}_{(1)}[g] \rangle - \lambda \langle g, \mathcal{K}_{(1)}[g] \rangle \\ &\quad - \bar{\lambda} \langle \mathcal{K}_{(1)}[g], g \rangle + |\lambda|^2 \langle g, g \rangle \\ &= \mathbf{g}^*(H - \lambda A^* - \bar{\lambda}A + |\lambda|^2 G)\mathbf{g}, \end{aligned}$$

where H is a newly introduced matrix with

$$H_{i,j} = \langle \mathcal{K}_{(1)}[\psi_j], \mathcal{K}_{(1)}[\psi_i] \rangle. \tag{24}$$

We employ the quadrature rule for the \mathbf{x} -domain to approximate this new term. If S is batched with $M_2 = 2$, then we can form the matrix

$$\tilde{H}_{i,j} = \sum_{l=1}^{M_1} w_l \psi_j(\mathbf{y}^{(l,1)}) \overline{\psi_i(\mathbf{y}^{(l,2)})}.$$

Since $\tau_{l,1}$ and $\tau_{l,2}$ are independent, we have

$$\lim_{M_1 \rightarrow \infty} \tilde{H}_{i,j} = H_{i,j} = \langle \mathcal{K}[\psi_j], \mathcal{K}[\psi_i] \rangle. \tag{25}$$

We stress that $\mathcal{K}_{(1)}$ is applied separately to ψ_i and ψ_j and thus $\tau_{l,1}$ and $\tau_{l,2}$ need to be independent realizations.

The convergence in (25) allows us to compute the spectral properties of $\mathcal{K}_{(1)}$ directly (see Sect. 3.3). In particular, instead of (13), we now have

$$[\text{res}(\lambda, g)]^2 = \frac{\mathbf{g}^* \left[\tilde{H} - \lambda \tilde{A}^* - \bar{\lambda} \tilde{A} + |\lambda|^2 \tilde{G} \right] \mathbf{g}}{\mathbf{g}^* \tilde{G} \mathbf{g}} \tag{26}$$

and the approximate decomposition

$$\begin{aligned} \int_{\Omega} \text{Var}_{\tau} [g(F_{\tau}(\mathbf{x}))] d\omega(\mathbf{x}) &= \mathbf{g}^* (L - H) \mathbf{g} \\ &\approx \mathbf{g}^* (\tilde{L} - \tilde{H}) \mathbf{g} = \|\mathbf{g}\|^2 \left([\text{res}^{\text{var}}(\lambda, g)]^2 - [\text{res}(\lambda, g)]^2 \right), \end{aligned} \tag{27}$$

which becomes exact in the large data limit.

3.3 Algorithms

In the derivations above, we noticed that one-batched data permits computation only of $\text{res}^{\text{var}}(\lambda, g)$, while two-batched data also permits the computation of $\text{res}(\lambda, g)$. Algorithms 1 and 2 approximate the relative residuals of EDMD eigenpairs in the scenario of unbatched and batched data, respectively. In Algorithm 2, we have taken an average when computing \tilde{A} and \tilde{L} to reduce quadrature error, and an average when computing \tilde{H} to ensure that it is self-adjoint (and positive semi-definite). Algorithm 3 approximates the pseudospectrum and corresponding pseudoeigenfunctions, given batched snapshot data. Algorithm 4 approximates the variance-pseudospectrum and corresponding variance-pseudoeigenfunctions, and does not need batched data. Note that the computational complexity of all of these algorithms scales the same as those for ResDMD, which is discussed in [20, 23]. In particular, Algorithms 1 and 2 scale the same as EDMD.

Algorithm 1 : Eigenpairs and residuals.

Input: Snapshot data $\{\mathbf{x}^{(m)}\}_{m=1}^M, \{\mathbf{y}^{(m)}\}_{m=1}^M$ ($\mathbf{y}^{(m)} = F(\mathbf{x}^{(m)}, \tau_m)$), quadrature weights $\{w_m\}_{m=1}^M$, and dictionary of observables $\{\psi_j\}_{j=1}^N$.

- 1: Compute $\tilde{G} = \Psi_X^* W \Psi_X, \tilde{A} = \Psi_X^* W \Psi_Y, \tilde{L} = \Psi_Y^* W \Psi_Y$, where Ψ_X and Ψ_Y are given in (7).
- 2: Solve $\tilde{A} \mathbf{g} = \lambda \tilde{G} \mathbf{g}$ for eigenpairs $\{(\lambda_j, \mathbf{g}_j) = \Psi \mathbf{g}_j\}$.
- 3: Compute $\text{res}^{\text{var}}(\lambda_j, \mathbf{g}_j)$ for all j (see (21)).

Output: Eigenpairs $\{(\lambda_j, \mathbf{g}_j)\}$ and variance residuals $\{\text{res}^{\text{var}}(\lambda_j, \mathbf{g}_j)\}$.

Algorithm 2 : Eigenpairs and residuals (batched data).

Input: Snapshot data $\{\mathbf{x}^{(m)}\}_{m=1}^M, \{\mathbf{y}^{(m,1)}\}_{m=1}^M, \{\mathbf{y}^{(m,2)}\}_{m=1}^M$ (batched), quadrature weights $\{w_m\}_{m=1}^M$, dictionary of observables $\{\psi_j\}_{j=1}^N$.

- 1: Compute $\tilde{G} = \Psi_X^* W \Psi_X,$
 $\tilde{A} = \left[\Psi_X^* W \Psi_Y^{(1)} + \Psi_X^* W \Psi_Y^{(2)} \right] / 2,$
 $\tilde{L} = \left[\Psi_Y^{(1)*} W \Psi_Y^{(1)} + \Psi_Y^{(2)*} W \Psi_Y^{(2)} \right] / 2,$
 $\tilde{H} = \left[\Psi_Y^{(1)*} W \Psi_Y^{(2)} + \Psi_Y^{(2)*} W \Psi_Y^{(1)} \right] / 2,$

- where Ψ_X and $\Psi_Y^{(i)}$ are given in (7) and the superscript for Ψ_Y corresponds to each batch of snapshot data.
- 2: Solve $\tilde{A} \mathbf{g} = \lambda \tilde{G} \mathbf{g}$ for eigenpairs $\{(\lambda_j, \mathbf{g}_j) = \Psi \mathbf{g}_j\}$.
 - 3: Compute $\text{res}^{\text{var}}(\lambda_j, \mathbf{g}_j)$ and $\text{res}(\lambda_j, \mathbf{g}_j)$ for all j (see (21) and (26)).

Output: Eigenpairs $\{(\lambda_j, \mathbf{g}_j)\}$ and residuals $\{\text{res}^{\text{var}}(\lambda_j, \mathbf{g}_j), \text{res}(\lambda_j, \mathbf{g}_j)\}$.

4 Theoretical guarantees

We now prove the correctness of the algorithms mentioned above. Specifically, through a series of theorems, we demonstrate that the computations of $\tilde{A}, \tilde{G}, \tilde{L},$ and \tilde{H} are accurate and that the spectral estimates can be trusted. To achieve this, we divide the section into three subsections, each focusing on demonstrating the accuracy of the spectrum, the predictive power, and the matrices, respectively. The universal assumptions made in this section are as follows:

- $\mathcal{K}_{(1)}$ is bounded.
- $\{\psi_j\}_{j=1}^N$ are linearly independent for any finite N .
- $V_N \subset V_{N+1}$ and the union, $\cup_N V_N$, is dense in $L^2(\Omega, \omega)$.

The algorithms and proofs can be readily adapted for an unbounded $\mathcal{K}_{(1)}$. The latter two assumptions can also be relaxed with minor modifications.

Algorithm 3 : Pseudospectra (batched data).

Input: Snapshot data $\{\mathbf{x}^{(m)}\}_{m=1}^M, \{\mathbf{y}^{(m,1)}, \mathbf{y}^{(m,2)}\}_{m=1}^M$ (batched), quadrature weights $\{w_m\}_{m=1}^M$, dictionary of observables $\{\psi_j\}_{j=1}^N$, an accuracy goal $\epsilon > 0$, and a grid $z_1, \dots, z_k \in \mathbb{C}$ (e.g., see (28)).

1: Compute

$$\begin{aligned} \tilde{G} &= \Psi_X^* W \Psi_X, \\ \tilde{A} &= [\Psi_X^* W \Psi_Y^{(1)} + \Psi_X^* W \Psi_Y^{(2)}] / 2, \\ \tilde{L} &= [\Psi_Y^{(1)*} W \Psi_Y^{(1)} + \Psi_Y^{(2)*} W \Psi_Y^{(2)}] / 2, \\ \tilde{H} &= [\Psi_Y^{(1)*} W \Psi_Y^{(2)} + \Psi_Y^{(2)*} W \Psi_Y^{(1)}] / 2, \end{aligned}$$

where Ψ_X and $\Psi_Y^{(i)}$ are given in (7) and the superscript for Ψ_Y corresponds to each batch of snapshot data.

2: For each z_j , compute $r_j = \min_{\mathbf{g} \in \mathbb{C}^N} \text{res}(z_j, \Psi \mathbf{g})$ (see (26)) and the corresponding singular vectors \mathbf{g}_j . This step is a generalized SVD problem.

Output: $\{z_j : r_j < \epsilon\}$, an estimate of $\text{Sp}_\epsilon(\mathcal{X}_{(1)})$, and pseudoeigenfunctions $\{\mathbf{g}_j : r_j < \epsilon\}$.

Algorithm 4 : Variance-pseudospectra.

Input: Snapshot data $\{\mathbf{x}^{(m)}\}_{m=1}^M, \{\mathbf{y}^{(m)}\}_{m=1}^M$ ($\mathbf{y}^{(m)} = F(\mathbf{x}^{(m)}, \tau_m)$), quadrature weights $\{w_m\}_{m=1}^M$, dictionary of observables $\{\psi_j\}_{j=1}^N$, an accuracy goal $\epsilon > 0$, and a grid $z_1, \dots, z_k \in \mathbb{C}$ (e.g., see (28)).

1: Compute

$$\begin{aligned} \tilde{G} &= \Psi_X^* W \Psi_X, \\ \tilde{A} &= \Psi_X^* W \Psi_Y, \\ \tilde{L} &= \Psi_Y^* W \Psi_Y, \end{aligned}$$

where Ψ_X and Ψ_Y are given in (7).

2: For each z_j , compute $r_j = \min_{\mathbf{g} \in \mathbb{C}^N} \text{res}^{\text{var}}(z_j, \Psi \mathbf{g})$ (see (21)) and the corresponding singular vectors \mathbf{g}_j . This step is a generalized SVD problem.

Output: $\{z_j : r_j < \epsilon\}$, an estimate of $\text{Sp}_\epsilon^{\text{var}}(\mathcal{X}_{(1)})$, and variance-pseudoeigenfunctions $\{\mathbf{g}_j : r_j < \epsilon\}$.

4.1 Accuracy in finding spectral quantities

In this subsection, we prove the convergence of our algorithms. We have already discussed the convergence of residuals in Algorithms 1 and 2, under the assumption of convergence of the finite matrices $\tilde{G}, \tilde{A}, \tilde{L}$, and \tilde{H} in the large data limit. Hence, we focus on Algorithm 4. We first define the functions

$$f_{M,N}(\lambda) = \min_{\mathbf{g} \in \mathbb{C}^N} \text{res}^{\text{var}}(\lambda, \Psi \mathbf{g}),$$

and note that $r_j = f_{M,N}(z_j)$ in Algorithm 4. Our first lemma describes the limit of these functions as $M \rightarrow \infty$ and $N \rightarrow \infty$.

Lemma 1 Suppose that

$$\lim_{M \rightarrow \infty} \tilde{G} = G, \quad \lim_{M \rightarrow \infty} \tilde{A} = A, \quad \lim_{M \rightarrow \infty} \tilde{L} = L,$$

then $f_N(\lambda) = \lim_{M \rightarrow \infty} f_{M,N}(\lambda)$ exists. Moreover, f_N is a nonincreasing function of N and converges to $\sigma_{\text{inf}}^{\text{var}}$ from above and uniformly on compact subsets of \mathbb{C} as a function of the spectral parameter λ .

Proof The limit $f_N(\lambda) = \lim_{M \rightarrow \infty} f_{M,N}(\lambda)$ follows trivially from the convergence of matrices. Moreover, we have

$$\begin{aligned} f_N(\lambda) &= \min_{\mathbf{g} \in \mathbb{C}^N} \sqrt{\frac{\mathbf{g}^*(L - \lambda A^* - \bar{\lambda} A + |\lambda|^2 G)\mathbf{g}}{\mathbf{g}^* G \mathbf{g}}} \\ &= \inf \left\{ \sqrt{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]} : g \in V_N, \|g\| = 1 \right\}. \end{aligned}$$

Since $V_N \subset V_{N+1}$, $f_N(\lambda)$ is nonincreasing in N . By definition, we also have

$$f_N(\lambda) \geq \sigma_{\text{inf}}^{\text{var}}(\lambda).$$

Let $\delta > 0$ and choose $g \in L^2(\Omega, \omega)$ such that $\|g\| = 1$ and

$$\sqrt{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]} \leq \sigma_{\text{inf}}^{\text{var}}(\lambda) + \delta.$$

Since $\cup_N V_N$ is dense in $L^2(\Omega, \omega)$, there exists some n and $g_n \in V_n$ such that $\|g_n\| = 1$ and

$$\sqrt{\mathbb{E}_\tau [\|g_n \circ F_\tau - \lambda g_n\|^2]} \leq \sqrt{\mathbb{E}_\tau [\|g \circ F_\tau - \lambda g\|^2]} + \delta.$$

It follows that $f_n(\lambda) \leq \sigma_{\text{inf}}^{\text{var}}(\lambda) + 2\delta$. Since this holds for any $\delta > 0$, $\lim_{N \rightarrow \infty} f_N(\lambda) = \sigma_{\text{inf}}^{\text{var}}(\lambda)$. Since $\sigma_{\text{inf}}^{\text{var}}(\lambda)$ is continuous in λ , f_N converges uniformly down to $\sigma_{\text{inf}}^{\text{var}}$ on compact subsets of \mathbb{C} by Dini’s theorem. \square

Let $\{\text{Grid}(N) = \{z_{1,N}, z_{2,N}, \dots, z_{k(N),N}\}\}$ be a sequence of grids, each finite, such that for any $\lambda \in \mathbb{C}$,

$$\lim_{N \rightarrow \infty} \text{dist}(\lambda, \text{Grid}(N)) = 0.$$

For example, we could take

$$\text{Grid}(N) = \frac{1}{N} [\mathbb{Z} + i\mathbb{Z}] \cap \{z \in \mathbb{C} : |z| \leq N\}. \quad (28)$$

In practice, one considers a grid of points over the region of interest in the complex plane. Lemma 1 tells

us that to study Algorithm 4 in the large data limit, we must analyze

$$\Gamma_N^\epsilon(\mathcal{K}_{(1)}) = \{\lambda \in \text{Grid}(N) : f_N(\lambda) < \epsilon\}.$$

To make the convergence of Algorithm 4 precise, we use the Attouch–Wets metric defined by [5]:

$$d_{\text{AW}}(C_1, C_2) = \sum_{n=1}^\infty 2^{-n} \min \left\{ 1, \sup_{|x| \leq n} |\text{dist}(x, C_1) - \text{dist}(x, C_2)| \right\},$$

where C_1, C_2 are closed nonempty subsets of \mathbb{C} . This metric corresponds to local uniform convergence on compact subsets of \mathbb{C} . For any closed nonempty sets C and C_n , $d_{\text{AW}}(C_n, C) \rightarrow 0$ if and only if for any $\delta > 0$ and $B_m(0)$ (closed ball of radius $m \in \mathbb{N}$ about 0), there exists N such that if $n > N$ then $C_n \cap B_m(0) \subset C + B_\delta(0)$ and $C \cap B_m(0) \subset C_n + B_\delta(0)$. The following theorem contains our convergence result.

Theorem 1 (Convergence to variance-pseudospectrum) *Let $\epsilon > 0$. Then, $\Gamma_N^\epsilon(\mathcal{K}_{(1)}) \subset \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$ and*

$$\lim_{N \rightarrow \infty} d_{\text{AW}}(\Gamma_N^\epsilon(\mathcal{K}_{(1)}), \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})) = 0.$$

Proof Lemma 1 shows that $\Gamma_N^\epsilon(\mathcal{K}_{(1)}) \subset \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$. To prove convergence, we use the characterization of the Attouch–Wets topology. Suppose that m is large such that $B_m(0) \cap \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) \neq \emptyset$. Since $\Gamma_N^\epsilon(\mathcal{K}_{(1)}) \subset \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$, we clearly have $\Gamma_N^\epsilon(\mathcal{K}_{(1)}) \cap B_m(0) \subset \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$. Hence, we must show that given $\delta > 0$, there exists n_0 such that if $N > n_0$ then $\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) \cap B_m(0) \subset \Gamma_N^\epsilon(\mathcal{K}_{(1)}) + B_\delta(0)$. Suppose for a contradiction that this statement is false. Then, there exists $\delta > 0$, $\lambda_{n_j} \in \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) \cap B_m(0)$, and $n_j \rightarrow \infty$ such that $\text{dist}(\lambda_{n_j}, \Gamma_{n_j}^\epsilon(\mathcal{K}_{(1)})) \geq \delta$.

Without loss of generality, we can assume that $\lambda_{n_j} \rightarrow \lambda \in \text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)}) \cap B_m(0)$. There exists some z with $\sigma_{\text{inf}}^{\text{var}}(z) < \epsilon$ and $|\lambda - z| \leq \delta/2$. Let $z_{n_j} \in \text{Grid}(n_j)$ such that $|z - z_{n_j}| \leq \text{dist}(z, \text{Grid}(n_j)) + n_j^{-1}$. Since $\sigma_{\text{inf}}^{\text{var}}$ is continuous and f_N converges locally uniformly to $\sigma_{\text{inf}}^{\text{var}}$, we must have $f_{n_j}(z_{n_j}) < \epsilon$ for large n_j so that $z_{n_j} \in \Gamma_{n_j}^\epsilon(\mathcal{K}_{(1)})$. But $|z_{n_j} - \lambda| \leq |z - \lambda| + |z_{n_j} - z| \leq \delta/2 + |z - z_{n_j}|$, which is smaller than δ for large n_j , and we reach the desired contradiction. \square

4.2 Error bounds for iterations

We now aim to bound the difference between \tilde{K}^n and \mathcal{K}^n , a step crucial for measuring the accuracy of our

approximation of the mean trajectories in $L^2(\Omega, \omega)$. This effort, in conjunction with the Chernoff-like bound presented in (19), enables us to compute the statistical properties of the trajectories and their forecasts. Our approach to establishing these bounds is twofold. First, we consider the difference between \tilde{K}^n and \mathcal{K}^n , taking into account both the estimation errors and the errors intrinsic to the subspace. Subsequently, we establish concentration bounds for the estimation errors of \tilde{G} , \tilde{A} , and \tilde{L} .

Theorem 2 (Error bound for forecasts) *Define the quantities*

$$\begin{aligned} I_G &= G^{\frac{1}{2}} \tilde{G}^{-\frac{1}{2}}, \\ \Delta_G &= \|I_G\| \|I - I_G^{-1}\| + \|(I - I_G)\|, \\ \Delta_A &= \|\mathcal{K}\| (1 + \|I_G\|) \|I_G - I\| + \|I_G\|^2 \|G^{-\frac{1}{2}}(A - \tilde{A})G^{-\frac{1}{2}}\|. \end{aligned}$$

Let $g = \sum_{j=1}^N \mathbf{g}_j \psi_j \in V_N$ and suppose that

$$\|\mathcal{K}_{(1)}^n g - \mathcal{P}_{V_N}^* (\mathcal{P}_{V_N} \mathcal{K}_{(1)} \mathcal{P}_{V_N}^*)^n g\| \leq \delta_n(g) \|g\|.$$

Then

$$\|\Psi \tilde{K}^n g - \mathcal{K}_{(1)}^n g\| \leq C_n \|g\|,$$

where

$$C_n = \left[\frac{\|\mathcal{K}\|^n - \Delta_A^n}{\|\mathcal{K}\| - \Delta_A} \Delta_A (\Delta_G + 1) + \|\mathcal{K}\|^n \Delta_G + \delta_n(g) \right].$$

Proof We introduce the two matrices

$$T = G^{-1/2} A G^{-1/2}, \quad \tilde{T} = \tilde{G}^{-1/2} \tilde{A} \tilde{G}^{-1/2}.$$

Note that

$$\begin{aligned} \|T\| &= \sup_{x \in \mathbb{C}^N} \frac{\|T G^{1/2} x\|}{\|G^{1/2} x\|} = \sup_{x \in \mathbb{C}^N} \frac{\|G^{1/2} K x\|}{\|G^{1/2} x\|} \\ &= \|\mathcal{P}_{V_N} \mathcal{K} \mathcal{P}_{V_N}^*\| \leq \|\mathcal{K}\|. \end{aligned}$$

We can re-write \tilde{T} as

$$\begin{aligned} \tilde{T} &= I_G^* G^{-1/2} \tilde{A} G^{-1/2} I_G \\ &= I_G^* T I_G + I_G^* G^{-1/2} (\tilde{A} - A) G^{-1/2} I_G \\ &= T + (I_G - I)^* T I_G + T (I_G - I) \\ &\quad + I_G^* G^{-1/2} (\tilde{A} - A) G^{-1/2} I_G. \end{aligned}$$

It follows that

$$\|T - \tilde{T}\| \leq \|\mathcal{K}\| (1 + \|I_G\|) \|I_G - I\|$$

$$\begin{aligned}
 & + \|I_G\|^2 \|G^{-1/2}(A - \tilde{A})G^{-1/2}\| \\
 & = \Delta_A.
 \end{aligned}$$

We have that

$$T^n - \tilde{T}^n = T(T^{n-1} - \tilde{T}^{n-1}) + (T - \tilde{T})\tilde{T}^{n-1}.$$

A simple proof by induction now shows that

$$\begin{aligned}
 \|T^n - \tilde{T}^n\| & \leq \|T - \tilde{T}\| \sum_{j=0}^{n-1} \|T\|^j \|\tilde{T}\|^{n-1-j} \\
 & \leq \Delta_A \sum_{j=0}^{n-1} \|\mathcal{K}\|^j (\|\mathcal{K}\| + \Delta_A)^{n-1-j} \\
 & = \Delta_A \frac{\|\mathcal{K}\|^n - \Delta_A^n}{\|\mathcal{K}\| - \Delta_A}.
 \end{aligned}$$

We wish to bound the quantity

$$\begin{aligned}
 \|\Psi K^n \mathbf{g} - \Psi \tilde{K}^n \mathbf{g}\| & = \|T^n G^{1/2} \mathbf{g} - I_G \tilde{T}^n \tilde{G}^{1/2} \mathbf{g}\| \\
 & \leq \|T^n - \tilde{T}^n\| \|\mathbf{g}\| + \|\tilde{T}^n G^{1/2} \mathbf{g} - I_G \tilde{T}^n \tilde{G}^{1/2} \mathbf{g}\|.
 \end{aligned}$$

We can express the final term on the right-hand side as

$$\begin{aligned}
 \tilde{T}^n G^{1/2} \mathbf{g} - I_G \tilde{T}^n \tilde{G}^{1/2} \mathbf{g} & = I_G \tilde{T}^n (I - I_G^{-1}) G^{1/2} \mathbf{g} \\
 & \quad + (I - I_G) \tilde{T}^n G^{1/2} \mathbf{g}.
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \|\tilde{T}^n G^{1/2} \mathbf{g} - I_G \tilde{T}^n \tilde{G}^{1/2} \mathbf{g}\| & \leq \|\tilde{T}^n\| \|G^{1/2} \mathbf{g}\| \Delta_G \\
 & \leq \left(\|\mathcal{K}\|^n + \|T^n - \tilde{T}^n\| \right) \Delta_G \|\mathbf{g}\|
 \end{aligned}$$

and hence that

$$\begin{aligned}
 \|\Psi K^n \mathbf{g} - \Psi \tilde{K}^n \mathbf{g}\| & \leq \left[\|T^n - \tilde{T}^n\| (\Delta_G + 1) + \|\mathcal{K}\|^n \Delta_G \right] \|\mathbf{g}\| \\
 & \leq \left[\frac{\|\mathcal{K}\|^n - \Delta_A^n}{\|\mathcal{K}\| - \Delta_A} \Delta_A (\Delta_G + 1) + \|\mathcal{K}\|^n \Delta_G \right] \|\mathbf{g}\|.
 \end{aligned}$$

The theorem now follows from the triangle inequality \square

This theorem explicitly tells us how much to trust the prediction using the computed Koopman matrix, compared with the true Koopman operator. The quantities Δ_G and Δ_A represent errors due to estimation or quadrature. They are both expected to be small. The

quantity $\delta_n(g)$ is an intrinsic invariant subspace error that depends on the dictionary and observable g . To approximate $\delta_n(g)$, note that

$$\mathcal{K}^n[g] - \Psi K^n \mathbf{g} = \sum_{j=1}^n \mathcal{K}^{n-j} [\mathcal{K}[\Psi K^{j-1} \mathbf{g}] - \Psi K^j \mathbf{g}]$$

and hence

$$\|\mathcal{K}^n[g] - \Psi K^n \mathbf{g}\| \leq \sum_{j=1}^n \|\mathcal{K}\|^{n-j} \|\mathcal{K}[\Psi K^{j-1} \mathbf{g}] - \Psi K^j \mathbf{g}\|. \tag{29}$$

To bound the term on the right-hand side, we can use the matrix H in (24) and the fact that

$$\|\mathcal{K} \Psi \mathbf{v} - \Psi K \mathbf{v}\| = \sqrt{\mathbf{v}^* H \mathbf{v} - 2\text{Re}(\mathbf{v}^* K^* A \mathbf{v}) + \mathbf{v}^* K^* G K \mathbf{v}} \tag{30}$$

for any $\mathbf{v} \in \mathbb{C}^N$.

4.3 Estimation error for computation of A , G , and L

To effectively estimate $\mathcal{K}_{(1)}g$ and $\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$ in practical applications, it is imperative to have reliable approximations of A , G , and L . We provide a justification for our ability to construct such approximations from trajectory data with high probability, employing concentration bounds. The subsequent result delineates the requisite number of samples and basis functions needed to achieve a desired level of accuracy with high probability. To ensure this level of accuracy, several reasonable assumptions about the stochastic dynamical system are necessary.

Assumption 1 We suppose that $\mathbf{x}^{(m)}$ in the snapshot data are sampled at random according to ω , independent of τ , and for simplicity, assume that ω is a probability measure.⁶ We assume that $\tau : \Omega_S \rightarrow \mathcal{H}$ for some Hilbert space \mathcal{H} and let $\kappa = (\mathbf{x}, \tau)$. In this section, \mathbb{E} and \mathbb{P} are with respect to the joint distribution of κ . We assume that

⁶ Similar types of bounds to Theorem 3 can be derived for ergodic sampling and high-order quadrature sampling.

- The random variable κ is sub-Gaussian, meaning that there exists some $a > 0$ such that

$$\mathbb{E} \left[e^{\|\kappa - \mathbb{E}(\kappa)\|^2/a^2} \right] < \infty.$$

This allows us to define the following finite quantity:

$$\Upsilon = \inf \left\{ s > 0 : e^{\frac{\mathbb{E}[\|\kappa - \mathbb{E}(\kappa)\|^2]}{s^2}} \mathbb{E} \left[e^{\frac{1}{s^2} \|\kappa - \mathbb{E}(\kappa)\|^2} \right] \leq 2 \right\}.$$

- The dictionary functions are uniformly bounded and satisfy the following Lipschitz condition:

$$|\psi_k(\mathbf{x}) - \psi_k(\mathbf{x}')| \leq c_k \|\mathbf{x} - \mathbf{x}'\|.$$

- The function F is Lipschitz with

$$\|F(\kappa) - F(\kappa')\| \leq c \|\kappa - \kappa'\|.$$

With these assumptions, we can show that our approximations of A , G , and L are good with high probability.

Theorem 3 (Concentration bound on estimation errors) *Under Assumption 1 we have, for any $t > 0$,*

$$\mathbb{P} \left(\|\tilde{A} - A\|_{\text{Fr}} < t \right) \geq 1 - \exp \left(2 \log(2N) - \frac{Mt^2}{24\Upsilon^2(c^2+1)\alpha^2\beta^2} \right)$$

$$\mathbb{P} \left(\|\tilde{G} - G\|_{\text{Fr}} < t \right) \geq 1 - \exp \left(2 \log(2N) - \frac{Mt^2}{48\Upsilon^2\alpha^2\beta^2} \right)$$

$$\mathbb{P} \left(\|\tilde{L} - L\|_{\text{Fr}} < t \right) \geq 1 - \exp \left(2 \log(2N) - \frac{Mt^2}{48\Upsilon^2c^2\alpha^2\beta^2} \right),$$

where $\|\cdot\|_{\text{Fr}}$ denotes the Frobenius norm, and α and β are given by

$$\alpha = \sqrt{\sum_{k=1}^N c_k^2}, \quad \beta = \sqrt{\sum_{k=1}^N \|\psi_k\|_{L^\infty}^2}.$$

Proof We first argue for $\|\tilde{A} - A\|_{\text{Fr}}$. Fix $j, k \in \{1, \dots, N\}$ and define the random variable

$$X = \psi_k(F(\mathbf{x}, \tau)) \overline{\psi_j(\mathbf{x})}.$$

Then

$$|X(\kappa) - X(\kappa')| \leq (c_k c \|\psi_j\|_{L^\infty} + c_j \|\psi_k\|_{L^\infty}) \|\kappa - \kappa'\|.$$

Let $c_{j,k} = c_k c \|\psi_j\|_{L^\infty} + c_j \|\psi_k\|_{L^\infty}$. The above Lipschitz bound for X implies that

$$|\mathbb{E}[X] - X(\kappa')| \leq c_{j,k} \int_{\Omega \times \Omega_s} \|\kappa - \kappa'\| d\mathbb{P}(\kappa)$$

$$\leq c_{j,k} \sqrt{\|\kappa - \mathbb{E}(\kappa)\|^2 + \mathbb{E}(\|\kappa - \mathbb{E}(\kappa)\|^2)},$$

where we have used Hölder’s inequality to derive the last line. It follows that

$$\mathbb{E} \left[\exp \left(\frac{|\mathbb{E}[X] - X|^2}{\Upsilon^2 c_{j,k}^2} \right) \right] \leq 2.$$

Let $Y = \text{Re}(\mathbb{E}[X] - X)$ and $\lambda \geq 0$. Since $\mathbb{E}[Y] = 0$, we have

$$\begin{aligned} \mathbb{E}[\exp(\lambda Y)] &= 1 + \sum_{l=2}^{\infty} \frac{\lambda^l \mathbb{E}[Y^l]}{l!} \\ &\leq 1 + \frac{\lambda^2}{2} \mathbb{E}[Y^2 \exp(\lambda|Y|)]. \end{aligned}$$

For any $b > 0$, we have $\lambda|Y| \leq \lambda^2/(2b) + b|Y|^2/2$. We also have $bY^2 \leq \exp(bY^2/2)$. It follows that

$$\mathbb{E}[\exp(\lambda Y)] \leq 1 + \frac{\lambda^2}{2b} e^{\lambda^2/(2b)} \mathbb{E}[\exp(bY^2)].$$

We select $b = 1/(\Upsilon^2 c_{j,k}^2)$ and use the fact that $\mathbb{E}[\exp(bY^2)] \leq \mathbb{E}[\exp(b|\mathbb{E}[X] - X|^2)] \leq 2$ to obtain

$$\mathbb{E}[\exp(\lambda Y)] \leq 1 + \frac{\lambda^2}{b} e^{\frac{\lambda^2}{2b}} \leq \left(1 + \frac{\lambda^2}{b} \right) e^{\frac{\lambda^2}{2b}} \leq e^{\frac{3\lambda^2}{2b}}.$$

Now let $\{Y^{(m)}\}_{m=1}^M$ independent copies of Y , then

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{M} \sum_{m=1}^M Y^{(m)} \geq t \right) \\ &= \mathbb{P} \left(\exp(\lambda \sum_{m=1}^M Y^{(m)}) \geq \exp(\lambda Mt) \right) \\ &\leq e^{-\lambda Mt} \mathbb{E} \left[\exp \left(\lambda \sum_{m=1}^M Y^{(m)} \right) \right] \\ &= e^{-\lambda Mt} \prod_{m=1}^M \mathbb{E}[\exp(\lambda Y)] \\ &\leq \exp \left(3M\lambda^2/(2b) - \lambda Mt \right), \end{aligned}$$

where we use Markov’s inequality in the first inequality. Minimizing over λ , we obtain

$$\mathbb{P} \left(\frac{1}{M} \sum_{m=1}^M Y^{(m)} \geq t \right) \leq \exp \left(-Mbt^2/6 \right).$$

We can argue in the same manner for $-Y$ and deduce that

$$\mathbb{P}\left(\frac{1}{M}\left|\sum_{m=1}^M Y^{(m)}\right|\geq t\right)\leq 2\exp(-Mbt^2/6).$$

Similarly, we can argue for the imaginary part of $\mathbb{E}[X] - X$.

We now allow j, k to vary and let $X_{j,k} = \psi_k(F(\mathbf{x}, \tau))\overline{\psi_j(\mathbf{x})}$. For $t > 0$, consider the events

$$\begin{aligned} S_{j,k,1} &: \frac{1}{M}\left|\sum_{m=1}^M \operatorname{Re}(\mathbb{E}[X_{j,k}] - X_{j,k}(\kappa_m))\right| \\ &< \frac{t\Upsilon c_{j,k}}{\sqrt{2\Upsilon^2\sum_{l,p=1}^N c_{l,p}^2}}, \\ S_{j,k,2} &: \frac{1}{M}\left|\sum_{m=1}^M \operatorname{Im}(\mathbb{E}[X_{j,k}] - X_{j,k}(\kappa_m))\right| \\ &< \frac{t\Upsilon c_{j,k}}{\sqrt{2\Upsilon^2\sum_{l,p=1}^N c_{l,p}^2}}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{P}(\cap_{j,k,i} S_{j,k,i}) &\geq 1 - \sum_{j,k=1}^N (\mathbb{P}(S_{j,k,1}^c) + \mathbb{P}(S_{j,k,2}^c)) \\ &\geq 1 - 4N^2 \exp\left(-\frac{Mt^2}{12\Upsilon^2\sum_{l,p=1}^N c_{l,p}^2}\right). \end{aligned}$$

Moreover, the AM-GM inequality implies that

$$c_{l,p}^2 \leq 2c^2 c_k^2 \|\psi_j\|_{L^\infty}^2 + 2c_j^2 \|\psi_k\|_{L^\infty}^2$$

and hence

$$\sum_{l,p=1}^N c_{l,p}^2 \leq 2(c^2 + 1)\alpha^2\beta^2.$$

It follows that

$$\mathbb{P}(\cap_{j,k,i} S_{j,k,i}) \geq 1 - \exp\left(2\log(2N) - \frac{Mt^2}{24\Upsilon^2(c^2 + 1)\alpha^2\beta^2}\right).$$

If $\cap_{j,k,i} S_{j,k,i}$, then $\|\tilde{A} - A\|_{\text{Fr}} < t$. We can argue in the same manner, without the function F , to deduce that

$$\mathbb{P}(\|\tilde{G} - G\|_{\text{Fr}} < t) \geq 1 - \exp\left(2\log(2N) - \frac{Mt^2}{48\Upsilon^2\alpha^2\beta^2}\right).$$

Finally, for the matrix L and its estimate \tilde{L} , we derive similar concentration bounds for $\psi_k(F(\mathbf{x}, \tau))\overline{\psi_j(F(\mathbf{x}, \tau))}$ to see that

$$\mathbb{P}(\|\tilde{L} - L\|_{\text{Fr}} < t) \geq 1 - \exp\left(2\log(2N) - \frac{Mt^2}{48\Upsilon^2c^2\alpha^2\beta^2}\right).$$

The statement of the theorem now follows. \square

This theorem explicitly spells out the number of basis functions and samples required to approximate the three matrices appearing in Theorem 2. Roughly speaking, if we set

$$\exp\left(2\log(2N) - Mt^2\right) \sim N^2 \exp(-Mt^2) \leq \delta,$$

then

$$M \sim |\ln \delta - 2 \ln N|/t^2.$$

For any fixed tolerance t , the confidence exponentially tightens up when M , the number of samples, increases. The idea is similar to other concentration inequality type bounds: if one samples from the same distribution many times, the sample mean becomes closer and closer to the true mean, and this bound gives the confidence interval for the tail bound. On the other hand, when N increases, more entries in the matrices need to be approximated, so it brings a logarithmically negative effect. More samples are needed to balance out the increase of N .

5 Examples

We now present three examples. The first two are based on numerically sampled trajectory data, while the final example utilizes collected experimental data.

5.1 Arnold's circle map

For our first example, we revisit the circle map discussed in Example 1, setting $c = 1/5$, ρ as the uniform distribution on $[0, 1]$, and defining

$$f(\mathbf{x}) = \frac{1}{4\pi} \sin(2\pi\mathbf{x}).$$

Our dictionary consists of Fourier modes $\{\exp(ij\mathbf{x}) : j = -n, \dots, n\}$ with $n = 20$ (yielding $N = 41$), and we use batched trajectory data with $M_1 = 100$ equally

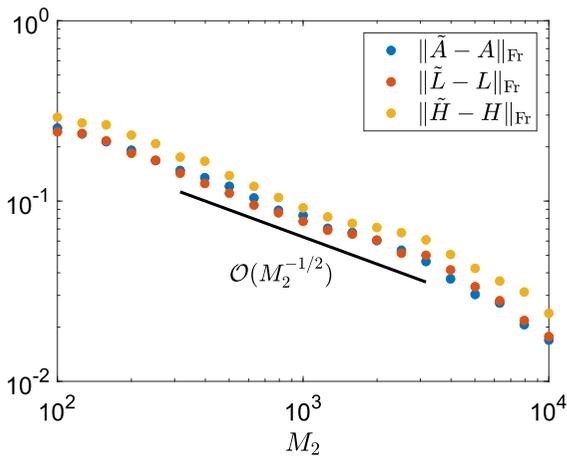


Fig. 2 Estimation error for the matrices \tilde{A} , \tilde{L} and \tilde{H} for the circle map. The solid line shows the expected Monte-Carlo convergence rate

spaced $\{\mathbf{x}^{(j)}\}$, and $M_2 = 2 \times 10^4$. Figure 2 illustrates the convergence of the matrices \tilde{A} , \tilde{L} , and \tilde{H} . We do not display the convergence of \tilde{G} as its error was on the order of machine precision, a result of the exponential convergence achieved by the trapezoidal quadrature rule across different batches. Figure 3 shows the residuals computed using Algorithm 2. The quantity $\text{res}^{\text{var}}(\lambda, g)$ deviates from (18) (the formula for $f = 0$), particularly when $|\lambda|$ is small. As n increases, the residuals $\text{res}(\lambda, g)$ converge to zero, indicating more accurate computation of the spectral content of $\mathcal{K}_{(1)}$. However, the residuals $\text{res}^{\text{var}}(\lambda, g)$ converge to finite positive values, except for the trivial eigenvalue 1, which satisfies $\lim_{M \rightarrow \infty} \text{res}^{\text{var}}(\lambda, g) = 0$.

To underscore the significance of variance in our analysis, Fig. 4 displays the absolute value of the matrix $\tilde{L} - \tilde{H}$, which approximates the covariance matrix defined in (16). Notably, the covariance disappears for the constant function $\exp(ij\mathbf{x})$ with $j = 0$, and the matrix is diagonally dominated. Figure 5 presents the results obtained from applying Algorithms 3 and 4. These results align in areas where the variance is minimal (large $|\lambda|$). However, in regions where $|\lambda|$ is small, the variance component in (27) becomes significant. This observation leads us to infer that only about seven eigenpairs are of meaningful significance in a statistically coherent framework.

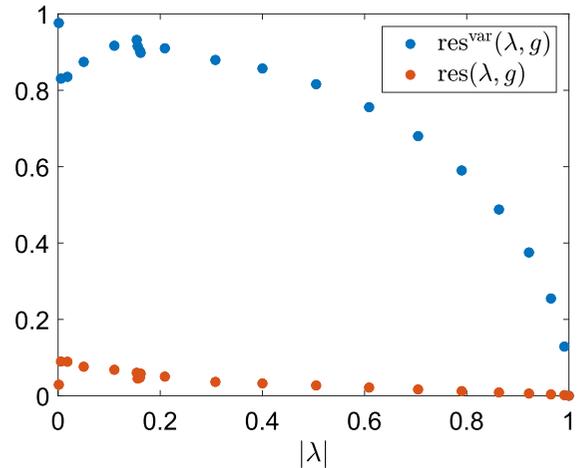


Fig. 3 Residuals for the circle map computed using Algorithm 2

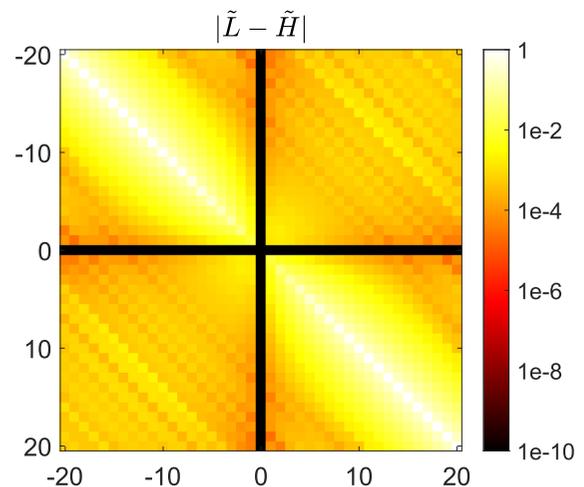


Fig. 4 Absolute values of the matrix $\tilde{L} - \tilde{H}$ for the circle map. This difference corresponds to the covariance matrix in (16)

5.2 Stochastic Van der Pol oscillator

We now consider the stochastic differential equation

$$\begin{aligned} dX_1 &= X_2 dt \\ dX_2 &= \left[\mu(1 - X_1^2)X_2 - X_1 \right] dt + \sqrt{2\delta} dB_t, \end{aligned}$$

where B_t denotes standard one-dimensional Brownian motion, $\delta > 0$, and $\mu > 0$.⁷ This equation represents

⁷ The inclusion of Brownian motion only in the dX_2 term is motivated by the physical interpretation of the random driving force. However, adding a similar term to the dX_1 equation would only affect the Kolmogorov operator by altering the parameter δ .

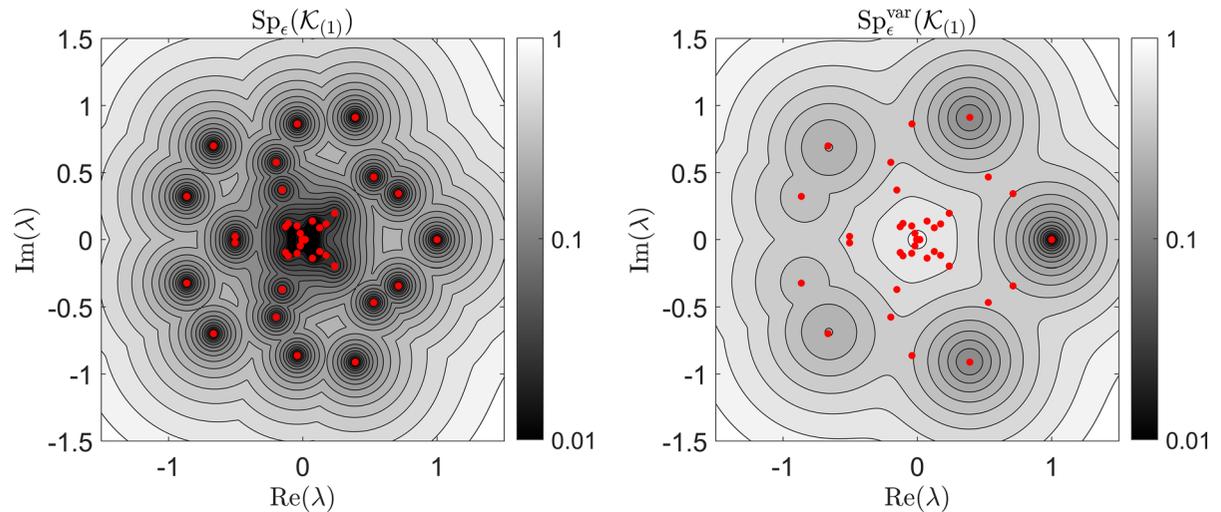


Fig. 5 Pseudospectra versus variance pseudospectra. Left: Output of Algorithm 3 for the circle map. Right: Output of Algorithm 4 for the circle map. We have shown the minimized resid-

als over a contour plot of ϵ in both cases. The red dots correspond to the EDMD eigenvalues

a noisy version of the Van der Pol oscillator. In the absence of noise, the Van der Pol oscillator exhibits a limit cycle to which all initial conditions converge, except for the unstable fixed point at the origin. The introduction of noise transforms the system, resulting in a global attractor that forms a band around the deterministic system's limit cycle.

The generator of the stochastic solutions, known as the backward Kolmogorov operator, is described in [25, Section 9.3]. It is a second-order elliptic type differential operator \mathcal{L} , defined by

$$[\mathcal{L}g](X_1, X_2) = \left(\mu(1 - X_1^2)X_2 - X_1 \right) \cdot \nabla g(X_1, X_2) + \delta \nabla^2 g(X_1, X_2).$$

For a discrete times step Δ_t , the Koopman operator is given by $\exp(\Delta_t \mathcal{L})$. In the absence of noise ($\delta = 0$), the Koopman operator has eigenvalues forming a lattice [53, Theorem 13]:

$$\left\{ \hat{\lambda}_{m,k} = \exp([-m\mu + ik\omega_0]\Delta_t) : k \in \mathbb{Z}, m \in \mathbb{N} \cup \{0\} \right\},$$

where $\omega_0 \approx 1 - \mu^2/16$ is the base frequency of the limit cycle [74]. When δ is moderate, the base frequency of the averaged limit cycle remains similar to that in the deterministic case [45].

We simulate the dynamics using the Euler–Maruyama method [65] with a time step of 3×10^{-3} . Data are collected along a single trajectory of length $M_1 = 10^6$ with $M_2 = 2$, starting the sampling after the trajectory reaches the global attractor. We employ 318 Laplacian radial basis functions with centers on the attractor as our dictionary. The parameters are set to $\mu = 0.5$, $\delta = 0.02$, and $\Delta_t = 0.3$.

Figure 6 displays the results obtained using Algorithms 3 and 4. Similar to observations from the circle map example, $\text{Sp}_\epsilon(\mathcal{K}_{(1)})$ and $\text{Sp}_\epsilon^{\text{var}}(\mathcal{K}_{(1)})$ exhibit greater similarity near the unit circle. The lattice-like structure in the eigenvalues is also evident, with the EDMD-computed eigenvalues appearing as perturbations of the set $\{\hat{\lambda}_{m,k}\}$. Table 1 lists some of these eigenvalues alongside the residuals calculated using Algorithm 2. We observe that as $|k|$ increases, $\text{res}(\lambda, g)$ also increases, and similarly, $\text{res}^{\text{var}}(\lambda, g)$ increases with m . For any given eigenvalue, $\text{res}(\lambda, g)$ decreases to zero with larger dictionaries. In contrast, $\text{res}^{\text{var}}(\lambda, g)$ approaches a finite nonzero value, except for the trivial eigenvalue, which has a constant eigenfunction exhibiting zero variance. Figure 7 illustrates the corresponding eigenfunctions on the attractor, showcasing their beautiful modal structure.

In this example, the norm of the Koopman operator $\|\mathcal{K}\|$ is approximately 1, and the subspace error $\delta_n(g)$ predominantly contributes to the bound established in

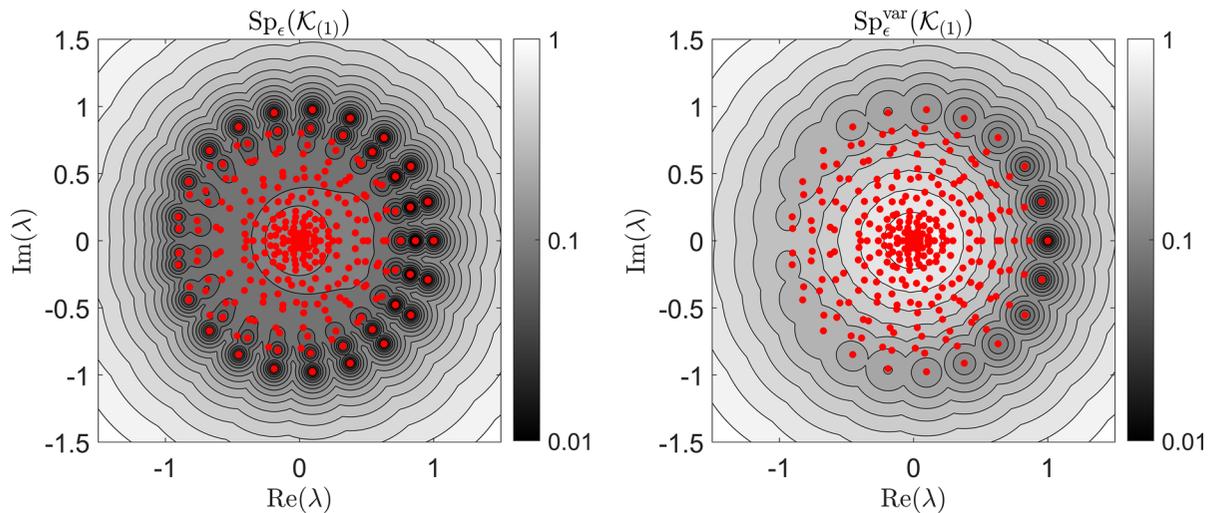


Fig. 6 Pseudospectra versus variance pseudospectra. Left: Output of Algorithm 3 for the stochastic Van der Pol oscillator. Right: Output of Algorithm 4 for the stochastic Van der Pol oscillator.

We have shown the minimized residuals over a contour plot of ϵ in both cases. The red dots correspond to the EDMD eigenvalues

Theorem 2. We analyze the two observables X_1 and X_2 , each starting from a point randomly selected on the attractor. Figure 8 presents the calculated values of $\delta_n(X_1)$ and $\delta_n(X_2)$ as per (29) and (30), along with the variance of the trajectory. Additionally, Fig. 9 compares the values computed using $K^n X_i$ with the actual values of $\mathcal{K}^n X_i$, obtained by integrating the generator \mathcal{L} . Together, these figures demonstrate the convergence of the mean trajectories toward the dominant subspace of \mathcal{H} .

5.3 Neuronal population dynamics

As a final example, we apply our approach to experimental neuroscience data. Recent technological advancements in this field now allow for the simultaneous monitoring of large neuronal populations in the brains of awake, behaving animals. This development has spurred significant interest in employing data-driven methods to derive physically meaningful insights from high-dimensional neural measurements [62].

To analyze complex neural data, researchers have employed a variety of analytical tools to uncover features like low-dimensional manifolds, latent population dynamics, within-trial variance, and trial-to-trial variability. However, existing methods often examine these

features in isolation [16, 29, 61, 73]. From a dynamical systems perspective, a unified model that captures these distinct aspects of neural data would be highly advantageous. In this context, the Koopman operator framework offers a compelling approach to analyzing high-dimensional neural observables [47]. DMD has emerged as a prominent method for the spatiotemporal decomposition of diverse datasets [9, 14]. Nevertheless, a limitation of DMD is its lack of explicit uncertainty quantification regarding the modes and forecasts it uncovers. This aspect is particularly vital in neural time series analysis, where it is challenging to identify physically meaningful spectral components [28].

Our framework offers a unified, data-driven solution to uncover validated latent dynamical modes and their associated variance in neural data. To demonstrate its efficacy, we applied it to high-dimensional neuronal recordings from the visual cortex of awake mice, as publicly shared by the Allen Brain Observatory [71], involving 400–800 neurons per mouse. Our focus was on the “Drifting Gratings” task epoch, wherein mice were presented with gratings drifting in one of eight directions (0° , 45° , etc.), modulated sinusoidally at one of five temporal frequencies. We specifically analyzed responses to gratings modulated at 15 Hz across all eight directions, as these stimuli consistently elicited an identifiable eigenvalue in the neural data corresponding to the expected frequency. This analysis encompassed

Table 1 Computed eigenvalues of the stochastic Van der Pol oscillator, and the residuals computed using Algorithm 2. We have ordered them according to perturbations of $\hat{\lambda}_{m,k}$. Due to conjugate symmetry, we have only shown eigenvalues with non-negative imaginary parts

$\lambda \approx \hat{\lambda}_{m,k}$	m	k	res ^{var}	res
1.000 + 0.000i	0	0	0.001	0.001
0.956 + 0.290i	0	1	0.040	0.001
0.829 + 0.554i	0	2	0.080	0.002
0.630 + 0.767i	0	3	0.120	0.005
0.378 + 0.912i	0	4	0.159	0.008
0.096 + 0.975i	0	5	0.198	0.012
-0.190 + 0.953i	0	6	0.237	0.016
-0.454 + 0.848i	0	7	0.275	0.022
-0.672 + 0.671i	0	8	0.313	0.029
0.864 + 0.000i	1	0	0.504	0.017
0.825 + 0.250i	1	1	0.506	0.009
0.715 + 0.477i	1	2	0.511	0.013
0.543 + 0.661i	1	3	0.518	0.024
0.325 + 0.784i	1	4	0.528	0.033
0.083 + 0.838i	1	5	0.541	0.041
-0.163 + 0.816i	1	6	0.555	0.051
-0.388 + 0.724i	1	7	0.571	0.062
-0.572 + 0.571i	1	8	0.589	0.074
0.751 + 0.000i	2	0	0.661	0.057
0.714 + 0.218i	2	1	0.665	0.066
0.614 + 0.415i	2	2	0.671	0.075
0.461 + 0.571i	2	3	0.679	0.084
0.271 + 0.673i	2	4	0.689	0.094
0.061 + 0.712i	2	5	0.700	0.104
-0.149 + 0.685i	2	6	0.713	0.117
-0.336 + 0.597i	2	7	0.729	0.131
-0.550 + 0.463i	2	8	0.696	0.144

120 trials per mouse (stimulus duration of 2 s) for a total of 20 mice, as detailed in [71]. We computed distinct stochastic Koopman operators for 15 different arousal levels, categorized by the average pupil diameter measured during the 500ms before each stimulus [49]. For this analysis, DMD was employed to identify 100 dictionary functions.

Our data-driven approach was effective in identifying an isolated, population-level coherent mode at the stimulus frequency. As illustrated in Fig. 10, this is evidenced by a distinct eigenvalue, highlighted in green, which consistently appears as a clear local minimum

in the variance pseudospectra contour plots across various arousal states. Without the variance pseudospectra, discerning which DMD eigenvalues are reliable and indicative of coherence can be challenging. We observed that individual neurons displayed a variety of waveforms, all linked to this single linear dynamic mode. Demonstrating the diversity of these responses, Fig. 11 showcases five randomly chosen sample trajectories from the KMD. These trajectories highlight the distinct spike counts and/or timings of different neurons, all parsimoniously represented by a single latent mode.

Importantly, neuronal responses demonstrate significant trial-to-trial variability, a phenomenon of considerable physiological interest due to its close relationship with ongoing fluctuations in an animal’s internal state. Dynamical systems approaches are adept at modeling this type of variability, which often stems from changes in the neural population’s pre-stimulus state [61]. Furthermore, the extent of this variability is heavily influenced by internal states like arousal and attention, as detailed in [50]. Our stochastic modeling approach enables us to additionally estimate this second source of trial-to-trial variability in neuronal responses.

To validate the physiological significance of our variance estimates, we analyzed the variance linked to the Koopman operators computed across each of 15 levels of pupil diameter, effectively using pupil diameter as a parameter for the Koopman operator in relation to arousal. Our hypothesis was that this analysis would reflect the well-known “U-shape” pattern described by the Yerkes–Dodson law [86], with variance minimized at intermediate arousal levels [49]. Figure 10 indicates that the eigenvalue or expectation derived from 10 remains consistent across various arousal states. However, from Fig. 12, a notable modulation in variance residuals is observed in accordance with arousal levels, aligning with our predictions: the variance associated with the leading mode is specifically reduced at intermediate arousal levels. This pattern underscores the physiological relevance of the variance estimates yielded by our modeling approach. Consequently, our findings suggest that arousal systematically influences dynamical variance, providing both practical and physiological rationales for employing dynamical models that explicitly estimate variance. Overall, our data-driven framework offers a unified and formal representation of neural dynamics, parsimoniously captur-

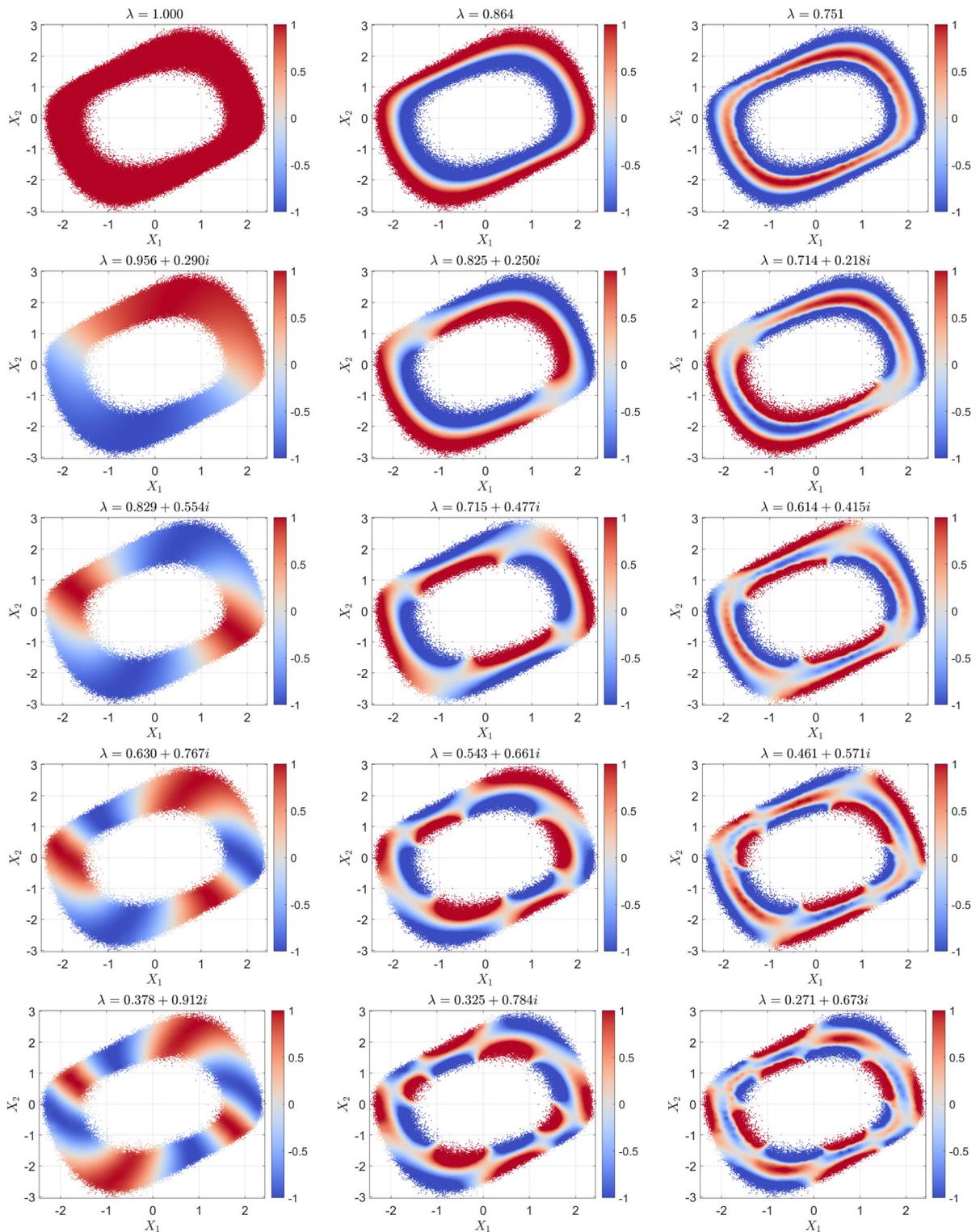


Fig. 7 Computed eigenfunctions (real part shown) of the stochastic Van der Pol oscillator. Due to conjugate symmetry, we have only shown eigenfunctions corresponding to eigenvalues with non-negative imaginary parts

Fig. 8 Left: Subspace errors $\delta_n(X_1)$ and $\delta_n(X_2)$ for the stochastic Van der Pol oscillator, computed using (29) and (30). Right: Variance of trajectory. We have rescaled the horizontal axis in both plots to correspond to time

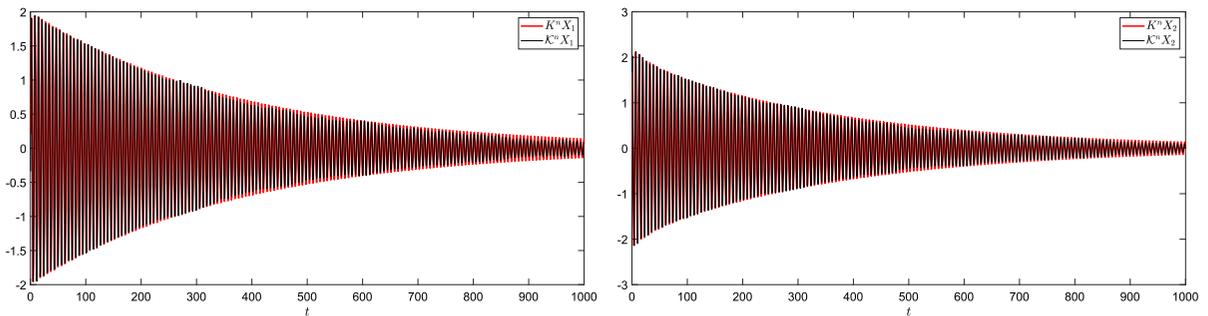
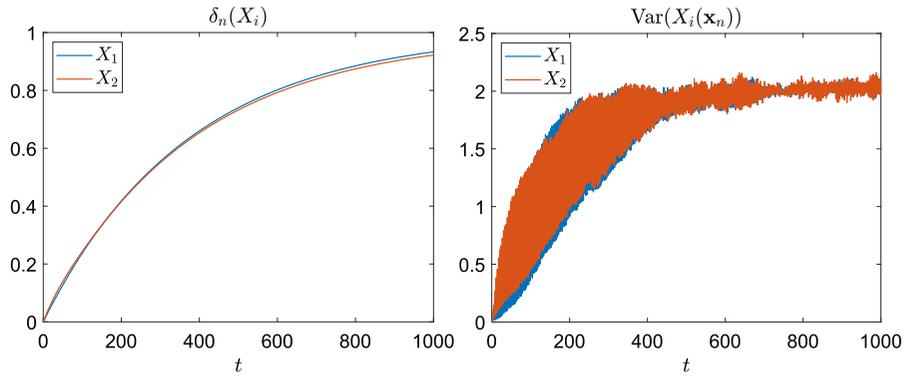


Fig. 9 Comparison of computed $K^n X_i$, where $K \in \mathbb{C}^{N \times N}$ is the EDMD matrix, and the true values of $\mathcal{X}^n X_i$

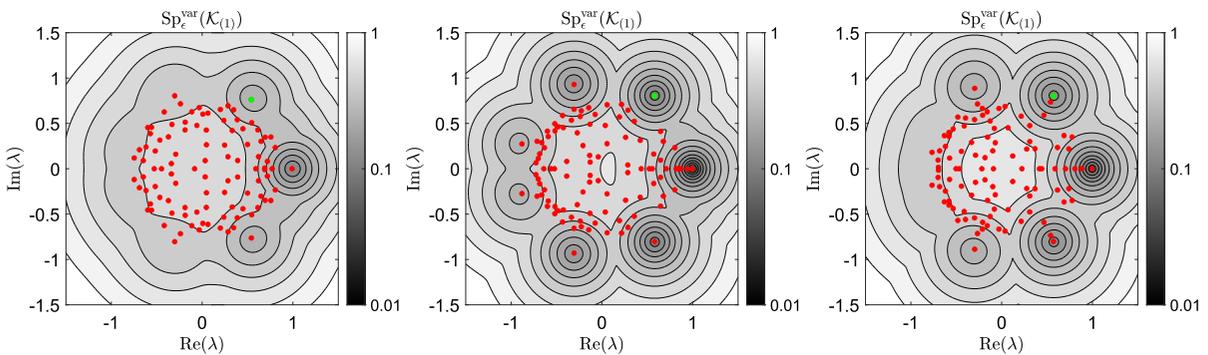


Fig. 10 Variance pseudospectra for a single mouse in the neuronal population dynamics example. Each case corresponds to a pupil diameter of 8% (left), 28% (middle), and 43% (right). The identified mode is shown in green, and the red dots show the

other DMD eigenvalues. The variance pseudospectra changes considerably as the arousal state changes, but the green eigenvalue shows little variability

ing multiple physiologically significant features in the data.

6 Conclusion

We have demonstrated the role of variance in the Koopman analysis of stochastic dynamical systems. To effectively

study projection errors in data-driven approaches for these systems, it is crucial to move beyond expectations and study more than just the stochastic Koopman operator. Incorporating variance into the Koopman framework enhances our understanding of spectral properties and the related projection errors. By analyzing various types of residuals, we have developed data-

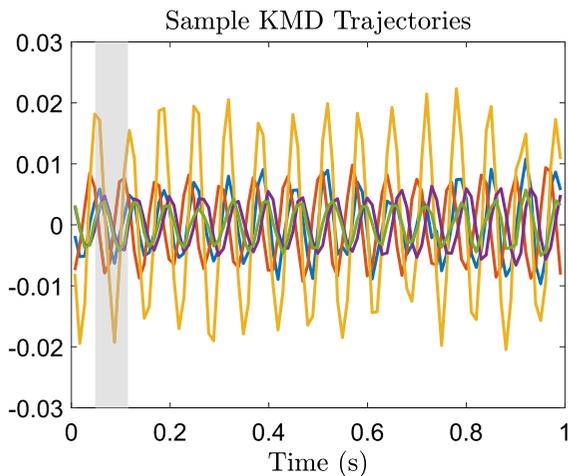


Fig. 11 Randomly selected sample trajectories from the Koopman mode corresponding to the eigenvalue shown in green in Fig. 10. The reference gray region in the left region shows the wavelength predicted by the eigenvalue

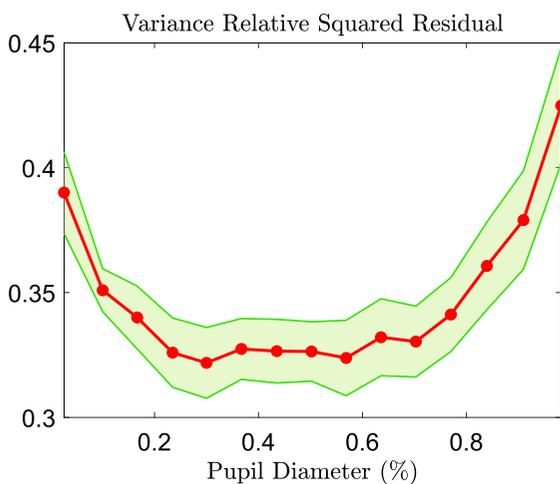


Fig. 12 The variance relative squared residual as a function of the arousal state. The red lines show the average across the mice, and the green error bounds correspond to the standard error of the mean. The “U-shape” is characteristic of the so-called Yerkes–Dodson law, which we produce in a data-driven fashion from the dynamics

driven algorithms capable of computing the spectral properties of infinite-dimensional stochastic Koopman operators. Furthermore, we introduced the concept of variance pseudospectra, a tool designed to assess statistical coherency. From a computational perspective, our work includes several convergence theorems pertinent to the spectral properties of these operators. In the realm of experimental neural recordings, our frame-

work has proven effective in extracting and compactly representing multiple data features with known physiological significance.

There are several avenues of future work related to this paper. One such direction involves an analysis of the algorithms and theorems presented in Sect. 4 in scenarios involving noisy snapshot data. Another avenue explores the trade-offs between computing the squared residual and variance terms, as outlined in (15), potentially reflecting variance-bias trade-offs in statistical analysis. Additionally, we aim to assess the robustness and generalizability of the proposed framework across further stochastic dynamical systems.

Acknowledgements We thank the Allen Institute for the publicly available data and the referees for valuable comments that helped improve the clarity of the paper. MJC would like to thank the Cecil King Foundation and the London Mathematical Society for a Cecil King Travel Scholarship that funded visits to the University of Wisconsin-Madison, the University of Washington, and Cornell University. QL would like to thank Vice Chancellor for Research and Graduate Education, DMS-2308440 and ONR-N000142112140. RVR would like to thank the Shanahan Family Foundation for support. AT work was partially supported by the NSF DMS-1952757, DMS-2045646, a Simons Mathematical Fellowship, and ONR-N000142312729.

Funding This study was funded by DMS-2308440, DMS-1952757, DMS-2045646, ONR-N000142112140, ONR-N000142312729.

Data availability All data generated or analyzed during this study is available upon request. A preprint of this work can be found at [22].

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Arbabi, H., Mezić, I.: Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM J. Appl. Dyn. Syst.* **16**(4), 2096–2126 (2017)
2. Arbabi, H., Mezić, I.: Study of dynamics in post-transient flows using Koopman mode decomposition. *Phys. Rev. Fluids* **2**(12), 124402 (2017)
3. Atchadé, Y.F., Perron, F.: On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics* **41**(1), 77–84 (2007)
4. Baddoo, P.J., Herrmann, B., McKeon, B.J., Nathan Kutz, J., Brunton, S.L.: Physics-informed dynamic mode decomposition. *Proc. R. Soc. A* **479**(2271), 20220576 (2023)
5. Beer, G.: *Topologies on Closed and Closed Convex Sets*, vol. 268. Springer, Berlin (1993)
6. Berger, E., Sastuba, M., Vogt, D., Jung, B., Ben Amor, H.: Estimation of perturbations in robotic behavior using dynamic mode decomposition. *Adv. Robot.* **29**(5), 331–343 (2015)
7. Böttcher, A., Silbermann, B.: The finite section method for Toeplitz operators on the quarter-plane with piecewise continuous symbols. *Math. Nachr.* **110**(1), 279–291 (1983)
8. Bruder, D., Gillespie, B., Remy, C.D., Vasudevan, R.: Modeling and control of soft robots using the Koopman operator and model predictive control. arXiv preprint [arXiv:1902.02827](https://arxiv.org/abs/1902.02827) (2019)
9. Brunton, B.W., Johnson, L.A., Ojemann, J.G., Kutz, J.N.: Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci Methods* **258**, 1–15 (2016)
10. Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N.: Chaos as an intermittently forced linear system. *Nat. Commun.* **8**(1), 1–9 (2017)
11. Brunton, S.L., Budišić, M., Kaiser, E., Kutz, J.N.: Modern Koopman theory for dynamical systems. *SIAM Rev.* **64**(2), 229–340 (2022)
12. Budišić, M., Mohr, R., Mezić, I.: Applied Koopmanism. *Chaos* **22**(4), 047510 (2012)
13. Caffisch, R.E.: Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.* **7**, 1–49 (1998)
14. Casorso, J., Kong, X., Chi, W., Van De Ville, D., Yeo, B.T., Liégeois, R.: Dynamic mode decomposition of resting-state and task fMRI. *Neuroimage* **194**, 42–54 (2019)
15. Chen, K.K., Tu, J.H., Rowley, C.W.: Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. *J. Nonlinear Sci.* **22**(6), 887–915 (2012)
16. Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al.: Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**(3), 369–378 (2010)
17. Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**(1), 5–30 (2006)
18. Colbrook, M., Horning, A., Townsend, A.: Computing spectral measures of self-adjoint operators. *SIAM Rev.* **63**(3), 489–524 (2021)
19. Colbrook, M.J.: The mpEDMD algorithm for data-driven computations of measure-preserving dynamical systems. *SIAM J. Numer. Anal.* **61**(3), 1585–1608 (2023)
20. Colbrook, M.J., Ayton, L.J., Szöke, M.: Residual dynamic mode decomposition: robust and verified Koopmanism. *J. Fluid Mech.* **955**, A21 (2023)
21. Colbrook, M.J., Horning, A., Townsend, A.: SpecSolve. github (online). <https://github.com/SpecSolve> (2020)
22. Colbrook, M.J., Li, Q., Raut, R.V., Townsend, A.: Beyond expectations: residual dynamic mode decomposition and variance for stochastic dynamical systems. arXiv preprint [arXiv:2308.10697](https://arxiv.org/abs/2308.10697) (2023)
23. Colbrook, M.J., Townsend, A.: Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Commun. Pure Appl. Math. (to appear)*
24. Črnjarić-Žic, N., Maćešić, S., Mezić, I.: Koopman operator spectrum for random dynamical systems. *J. Nonlinear Sci.* **30**, 2007–2056 (2020)
25. Da Prato, G., Zabczyk, J.: *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge (2014)
26. Das, S., Giannakis, D., Slawinska, J.: Reproducing kernel Hilbert space compactification of unitary evolution groups. *Appl. Comput. Harmon. Anal.* **54**, 75–136 (2021)
27. Dawson, S., Hemati, M.S., Williams, M.O., Rowley, C.W.: Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Exp. Fluids* **57**(3), 1–19 (2016)
28. Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., et al.: Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**(12), 1655–1665 (2020)
29. Gao, Y., Archer, E.W., Paninski, L., Cunningham, J.P.: Linear dynamical neural population models through nonlinear embeddings. *Adv. Neural Inf. Process. Syst.* **29**, 163–171 (2016)
30. Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* **47**(2), 338–396 (2019)
31. Giannakis, D., Kolchinskaya, A., Krasnov, D., Schumacher, J.: Koopman analysis of the long-term evolution in a turbulent convection cell. *J. Fluid Mech.* **847**, 735–767 (2018)
32. Giannakis, D., Majda, A.J.: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.* **109**(7), 2222–2227 (2012)
33. Gikhman, I.I., Skorokhod, A.V.: *The Theory of Stochastic Processes: I*, vol. 210. Springer, Berlin (2004)
34. Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17**(6), R55 (2004)
35. Hemati, M.S., Rowley, C.W., Deem, E.A., Cattafesta, L.N.: De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. *Theor. Comput. Fluid Dyn.* **31**(4), 349–368 (2017)
36. Kachurovskii, A.G.: The rate of convergence in ergodic theorems. *Russ. Math. Sur.* **51**(4), 653–703 (1996)
37. Kaiser, E., Kutz, J.N., Brunton, S.L.: Data-driven discovery of Koopman eigenfunctions for control. *Mach. Learn. Sci. Technol.* **2**(3), 035023 (2021)

38. Klus, S., Koltai, P., Schütte, C.: On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comput. Dyn.* **3**(1), 51–79 (2016)
39. Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F.: Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.* **28**(3), 985–1010 (2018)
40. Kolmogoroff, A.: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 415–458 (1931)
41. Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. *J. Nonlinear Sci.* **28**(2), 687–710 (2018)
42. Korda, M., Putinar, M., Mezić, I.: Data-driven spectral analysis of the Koopman operator. *Appl. Comput. Harmon. Anal.* **48**(2), 599–629 (2020)
43. Kostic, V.R., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., et al.: Learning dynamical systems via Koopman operator regression in reproducing kernel Hilbert spaces. In: *Advances in Neural Information Processing Systems*
44. Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: *Dynamic Mode Decomposition: Data-driven Modeling of Complex Systems*. SIAM, Philadelphia (2016)
45. Leung, H.: Stochastic transient of a noisy van der Pol oscillator. *Physica A* **221**(1–3), 340–347 (1995)
46. Mann, J., Kutz, J.N.: Dynamic mode decomposition for financial trading strategies. *Quant. Finance* **16**(11), 1643–1655 (2016)
47. Marrouch, N., Slawinska, J., Giannakis, D., Read, H.L.: Data-driven Koopman operator approach for computational neuroscience. *Ann. Math. Artif. Intell.* **88**(11–12), 1155–1173 (2020)
48. Mauroy, A., Mezić, I.: On the use of Fourier averages to compute the global isochrons of (quasi) periodic dynamics. *Chaos Interdiscip. J. Nonlinear Sci.* **22**(3), 033112 (2012)
49. McGinley, M.J., David, S.V., McCormick, D.A.: Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* **87**(1), 179–192 (2015)
50. McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., McCormick, D.A.: Waking state: rapid variations modulate neural and behavioral responses. *Neuron* **87**(6), 1143–1161 (2015)
51. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**(1), 309–325 (2005)
52. Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. *Ann. Rev. Fluid Mech.* **45**, 357–378 (2013)
53. Mezić, I.: Koopman operator spectrum and data analysis. arXiv preprint [arXiv:1702.07597](https://arxiv.org/abs/1702.07597) (2017)
54. Mezić, I.: Koopman operator, geometry, and learning of dynamical systems. *Not. Am. Math. Soc.* **68**, 1087–1105 (2021)
55. Mezić, I.: On numerical approximations of the Koopman operator. *Mathematics* **10**(7), 1180 (2022)
56. Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior: spectral methods. In: *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, vol. 2, pp. 1224–1231. IEEE (2000)
57. Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior. *Phys. D: Nonlin. Phen.* **197**(1–2), 101–133 (2004)
58. Mollenhauer, M., Klus, S., Schütte, C., Koltai, P.: Kernel autocovariance operators of stationary processes: estimation and convergence. *J. Mach. Learn. Res.* **23**(327), 1–34 (2022)
59. Nuske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S., Noé, F.: Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **10**(4), 1739–1752 (2014)
60. Nüske, F., Peitz, S., Philipp, F., Schaller, M., Worthmann, K.: Finite-data error bounds for Koopman-based prediction and control. *J. Nonlinear Sci.* **33**(1), 14 (2023)
61. Pandarinath, C., O’Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., et al.: Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* **15**(10), 805–815 (2018)
62. Paninski, L., Cunningham, J.P.: Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.* **50**, 232–241 (2018)
63. Proctor, J.L., Brunton, S.L., Kutz, J.N.: Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.* **15**(1), 142–161 (2016)
64. Proctor, J.L., Eckhoff, P.A.: Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *Int. Health* **7**(2), 139–145 (2015)
65. Röbber, A.: Runge–Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Numer. Anal.* **48**(3), 922–952 (2010)
66. Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Hennings, D.S.: Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127 (2009)
67. Schmid, P.J.: Dynamic mode decomposition of experimental data. In: *8th International Symposium on Particle Image Velocimetry (PIV09)* (2009)
68. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28 (2010)
69. Schwantes, C.R., Pande, V.S.: Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**(4), 2000–2009 (2013)
70. Schwantes, C.R., Pande, V.S.: Modeling molecular kinetics with tICA and the kernel trick. *J. Chem. Theory Comput.* **11**(2), 600–608 (2015)
71. Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Gradis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al.: Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**(7852), 86–92 (2021)
72. Sinha, S., Huang, B., Vaidya, U.: On robust computation of Koopman operator and prediction in random dynamical systems. *J. Nonlinear Sci.* **30**(5), 2057–2090 (2020)
73. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., Harris, K.D.: Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**(6437), eaav7893 (2019)
74. Strogatz, S.H.: *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, Boca Raton (2018)
75. Susuki, Y., Mezić, I.: Nonlinear Koopman modes and coherency identification of coupled swing dynamics. *IEEE Trans. Power Syst.* **26**(4), 1894–1904 (2011)

76. Susuki, Y., Mezić, I., Hikiyama, T.: Coherent swing instability of power grids. *J. Nonlinear Sci.* **21**(3), 403–439 (2011)
77. Takeishi, N., Kawahara, Y., Yairi, T.: Subspace dynamic mode decomposition for stochastic Koopman analysis. *Phys. Rev. E* **96**(3), 033310 (2017)
78. Trefethen, L.N., Embree, M.: *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, Princeton (2005)
79. Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Kutz, J.N.: On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**(2), 391–421 (2014)
80. Ulam, S.M.: *A Collection of Mathematical Problems*, vol. 8. Interscience Publishers, New York (1960)
81. Vitalini, F., Noé, F., Keller, B.: A basis set for peptides for the variational approach to conformational kinetics. *J. Chem. Theory Comput.* **11**(9), 3992–4004 (2015)
82. Wanner, M., Mezić, I.: Robust approximation of the stochastic Koopman operator. *SIAM J. Appl. Dyn. Syst.* **21**(3), 1930–1951 (2022)
83. Webber, R.J., Thiede, E.H., Dow, D., Dinner, A.R., Weare, J.: Error bounds for dynamical spectral estimation. *SIAM J. Math. Data Sci.* **3**(1), 225–252 (2021)
84. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* **25**(6), 1307–1346 (2015)
85. Williams, M.O., Rowley, C.W., Kevrekidis, I.G.: A kernel-based method for data-driven Koopman spectral analysis. *J. Comput. Dyn.* **2**(2), 247 (2015)
86. Yerkes, R.M., Dodson, J.D., et al.: The relation of strength of stimulus to rapidity of habit-formation. *Int. Health* **7**, 139–145 (1908)
87. Zhang, B.J., Sahai, T., Marzouk, Y.M.: A Koopman framework for rare event simulation in stochastic differential equations. *J. Comput. Phys.* **456**, 111025 (2022)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.