A Data-Driven Framework for Koopman Semigroup Estimation in Stochastic Dynamical Systems

Yuanchao Xu,[1] Kaidi Shao,[2] Isao Ishikawa,[3] Yuka Hashimoto,[4] Nikos Logothetis,[2] and Zhongwei Shen[1]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, University Commons 5-140 Edmonton, Alberta, Canada T6G 2N8.*

[2]*International Center for Primate Brain Research, Chinese Academy of Sciences, No. 500 Qiang Ye Road, Songjiang District, Shanghai, 201602, China.*

[3]*Center for Science Adventure and Collaborative Research Advancement, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto-shi, Kyoto 606-8502, Japan.*

[4]*NTT Network Service Systems Laboratories, NTT Corporation, 3-9-11, Midori-cho, Musashinoshi, Tokyo, 180-8585, Japan.*

(*Electronic mail: yuanchao@ualberta.ca.)

(Dated: 3 October 2025)

We present Stochastic Dynamic Mode Decomposition (SDMD), a novel data-driven framework for approximating the Koopman semigroup in stochastic dynamical systems. Unlike existing approaches, SDMD explicitly incorporates sampling time into its formulation to ensure numerical stability and precision in the presence of noise. By directly approximating the Koopman semigroup rather than its generator, SDMD avoids computationally expensive matrix exponential calculation, providing a more practically efficient pathway for analyzing stochastic dynamics. The framework also leverages neural networks for automated basis selection, minimizing manual effort while preserving computational efficiency. We establish SDMD's theoretical foundations through rigorous convergence guarantees across three critical limits in order: large data, infinitesimal sampling time, and increasing dictionary size. Numerical experiments on canonical stochastic systems including oscillatory system, mean-reverting processes, metastable system and a neural mass model demonstrate SDMD's effectiveness in capturing the spectral properties of the Koopman semigroup, even in systems with complex random behavior.

Understanding the long-term behavior of dynamical systems perturbed by random noise is fundamental to many scientific fields, from neuroscience to molecular dynamics, where stochastic effects often drive critical transitions and emergent behaviors. The Koopman operator framework provides a powerful mathematical tool for analyzing such systems by transforming nonlinear dynamics into linear spectral problems. Inspired by the previous successful application such as EDMD and gEDMD methods to stochastic systems, this work introduces Stochastic Dynamic Mode Decomposition (SDMD), a new data-driven approach that directly approximates Koopman semigroups while explicitly incorporating sampling time Δt with controlled truncation of higher-order terms to enhance computational stability. SDMD avoids expensive matrix exponential calculations and demonstrates superior performance in capturing spectral properties and recovering hidden multiscale dynamics across various canonical stochastic systems which is a more efficient and robust framework for analyzing complex random dynamical behaviors.

I. INTRODUCTION

Dynamical systems theory[1] provides a foundational framework for modeling complex phenomena across different areas such as climate science[2], molecular dynamics[3], fluid mechanics[4], finance[5], and neuroscience[6]. In these fields, stochastic dynamical systems naturally arise due to the impor-

tant role of random perturbation. Consideration of stochasticity is essential because real-world systems often feature incomplete data, measurement noise, high degrees of freedom, and complex, unknown mechanisms. Random fluctuation can drive critical transitions like abrupt climate shifts[7] or molecular conformational changes[8] that deterministic models often fail to capture. This is also evident in neuroscience, where stochastic modeling methods are increasingly preferred over deterministic approach[9,10] due to intrinsic variability in neural activity from synaptic release[11], ion-channel gating[12], and fluctuating inputs[13]. Moreover, neural dynamics are inherently multi-scale, ranging from millisecond ion-channel fluctuations to slow population-level oscillations, and stochastic approaches remain one of the few tractable ways to bridge these scales[14]. In general, models accounting for stochasticity usually offer a more realistic and mathematically robust representation of brain dynamics.

Practically speaking, the lack of complete knowledge about underlying first-principles makes direct modeling of complex systems extremely challenging. However, data-driven methods have gained significant attention in recent years for their ability to directly extract insights from observations without requiring detailed prior knowledge or explicit mathematical formulations[15–19]. This approach is particularly valuable for complex stochastic systems where theoretical models are incomplete or computationally infeasible.

Among various data-driven approaches, operator-theoretic methods excel as powerful tools for dynamical system analysis. In particular, the Koopman operator theory[20,21] converts nonlinear dynamics into a linear, though potentially infinite-dimensional, framework through observables. This

conversion allows spectral analysis[22–24], where eigenvalues and eigenfunctions reveal crucial details about a system's stability and long-term behavior. In stochastic systems, the operator becomes a Markov semigroup defined by conditional expectations and captures probabilistic evolution over time. Eigenvalues indicate decay rates and growth patterns, while eigenfunctions help identify invariant structures and coherent features within the dynamics. Such spectral insights are invaluable for understanding system behavior, predicting future evolution, and designing control strategies in both deterministic and stochastic settings. Recent developments in data-driven methods for estimating Koopman semigroup or generator[25–31], particularly Extended Dynamic Mode Decomposition (EDMD)[30], have made significant progress in approximating the Koopman operator directly from data. However, EDMD was originally designed for deterministic systems and does not explicitly account for stochastic effects. To address these limitations, various methods[27,28,32–35] have been proposed over time. For example, in Colbrook *et al.*[32], authors have introduced the concept of variance-pseudospectra as a measure of statistical coherency, which helps in understanding the stochastic system's spectral properties. In Črnjarić-Žic, Maćešić, and Mezić[33], authors introduced stochastic Hankel-DMD (sHankel-DMD) algorithm to approximate the spectral properties of the stochastic Koopman operator. In Klus *et al.*[34] authors generalized Galerkin approximation method as an extension of EDMD (gEDMD) to approximate the infinitesimal generator of the Koopman operator. In Kostic *et al.*[27], authors developed a statistical learning framework to learn Koopman operators in reproducing kernel Hilbert spaces (RKHS). In Noé and Nuske[28], authors proposed a variational approach based on maximizing the Rayleigh coefficient for modeling slow processes in stochastic dynamical systems. In Wanner and Mezic[35], authors introduced a new DMD algorithm that can accurately approximate the stochastic Koopman operator even when both the dynamics are random and the measurements contain noise. It also enables time-delayed observables for random systems using data from a single trajectory.

In this paper, we introduce **Stochastic Dynamic Mode Decomposition (SDMD)**, a novel data-driven framework that estimates the Koopman semigroup in the stochastic system by explicitly incorporating sampling time into the approximation process. The key innovation of directly approximating the Koopman semigroup bypass the need for matrix exponential computations. This design not only enhances computational efficiency but also ensures numerical stability when dealing with a typically unbounded Koopman generator. The main contributions of this work include:

- **Explicit Consideration of Sampling Time (Δt) for Stability**: The explicit inclusion of sampling time (Δt) in the SDMD framework is a key innovation, which ensures numerical stability and precision, and addresses challenges faced by other methods in handling stochastic dynamics.

- **Direct Approximation of the Semigroup**: SDMD directly approximates the Koopman semigroup, which

avoids the computationally expensive matrix exponential calculations required by most methods that return only the generator. This approach reduces computational cost while providing a more practical and efficient pathway for analyzing stochastic systems.

- **Computational Efficiency with Neural Network Integration**: The neural network extension enables adaptive basis selection directly from data without requiring manual intervention. Unlike other methods which may involves resource-intensive computations of Jacobian and Hessian matrices, our method significantly reduces computational resource while maintaining consistency with the stochastic evolution.

- **Rigorous Theoretical Guarantees**: The proposed framework includes comprehensive convergence analysis, covering the large data limit, the zero-limit of sampling time, and the large dictionary size. These guarantees establish the reliability and robustness of SDMD in approximating the Koopman semigroup.

From an application perspective, our SDMD framework aims to provide a practical and efficient tool for analyzing stochastic dynamical systems. By explicitly incorporating sampling time and directly approximating the Koopman semigroup, SDMD provides more reliable spectral analysis for systems where random fluctuations play a crucial role. This enhanced reliability translates to improved identification of dominant modes, more accurate prediction of system evolution, and better characterization of system stability under stochastic perturbations. The method's computational efficiency makes it particularly suitable for analyzing large datasets from experimental observations, where both noise and deterministic dynamics need to be properly accounted for. Additionally, the neural network extension of SDMD offers automated feature extraction from complex data, reducing the need for domain-specific expertise in basis function selection. These practical benefits complement the theoretical guarantees of our approach, making SDMD a powerful tool for researchers seeking to understand the fundamental dynamical properties of stochastic systems across various scientific disciplines.

The rest of this paper is organized as follows: Section 2 provides the mathematical background of stochastic Koopman operators. Section 3 details our computational methodology. Section 4 presents the convergence analysis. Section 5 extends the framework to neural network implementations. Section 6 demonstrates the effectiveness of our approach through experiments. Finally, Section 7 concludes with discussions and future directions.

II. STOCHASTIC KOOPMAN OPERATOR

In dynamical systems, the Koopman operator provides a powerful mathematical framework for analyzing the evolution of observables instead of the system states themselves. For stochastic systems, the Koopman operator forms a Markov

semigroup[36–38] defined through conditional expectations, capturing the probabilistic evolution of observables over time by describing how their expected values change under the influence of both deterministic dynamics and random perturbations.

Let $\mathcal{M} \subseteq \mathbb{R}^d$ be the state space equipped with the Borel σ-algebra, and consider a continuous-time stochastic process $(\mathbf{X}_t)_{t \geq 0}$ on a probability space $(\Omega, \mathbb{P})$ defined by the stochastic differential equation:

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{W}_t, \quad \mathbf{X}_0 = \mathbf{x}, \tag{1}$$

where $\mathbf{b} : \mathcal{M} \to \mathbb{R}^d$ is the drift term, $\boldsymbol{\sigma} : \mathcal{M} \to \mathbb{R}^{d \times d}$ is the diffusion term, and $(\mathbf{W}_t)_{t \geq 0}$ is an d-dimensional Wiener process. We assume that both $\mathbf{b}$ and $\boldsymbol{\sigma}$ satisfy appropriate regularity condition[39].

Let ρ be a probability distribution on $\mathcal{M}$. The space $\mathcal{F}$ of square-integrable functions with respect to ρ is defined as:

$$\mathcal{F} := \left\{ f : \int_{\mathcal{M}} |f|^2 d\rho < \infty \right\},$$

equipped with the inner product $\langle f, g \rangle_\rho := \int_{\mathcal{M}} fg \, d\rho$ and corresponding norm $\|f\|_\rho := \sqrt{\langle f, f \rangle_\rho}$. Notice that ρ is not necessarily a stationary distribution of the underlying dynamical system.

For any observable $f \in \mathcal{F}$, the stochastic Koopman operator family $(\mathcal{K}^t)_{t \geq 0}$ is defined as

$$(\mathcal{K}^t f)(\mathbf{x}) := \mathbb{E}_{\mathbb{P}}[f(\mathbf{X}_t) | \mathbf{X}_0 = \mathbf{x}], \tag{2}$$

where $\mathbb{E}_{\mathbb{P}}$ denotes the expectation with respect to the probability measure $\mathbb{P}$ on Ω, and $\mathbf{X}_t : \Omega \to \mathcal{M}$ is the process starting from $\mathbf{x}$.

Assumption II.1. We assume that $\{\mathcal{K}^t\}_{t \geq 0}$ is a strongly continuous semigroup of bounded linear operators on $\mathcal{F}$, that is,

- $\mathcal{K}^t$ is a bounded linear operator on $\mathcal{F}$ for each $t \geq 0$;
- $\mathcal{K}^0 = \mathrm{I}$, $\mathcal{K}^t \circ \mathcal{K}^s = \mathcal{K}^{t+s}$ for all $t, s \geq 0$;
- $\lim_{t \to 0^+} \|\mathcal{K}^t f - f\|_\rho = 0$ for each $f \in \mathcal{F}$.

The connection between the stochastic process and Koopman operator can be further understood through its infinitesimal generator $\mathcal{A}$, defined as

$$\mathcal{A}f := \lim_{t \to 0} \frac{\mathcal{K}^t f - f}{t} \tag{3}$$

on the domain

$$\mathcal{D}(\mathcal{A}) = \left\{ f \in \mathcal{F} : \lim_{t \to 0} \frac{\mathcal{K}^t f - f}{t} \text{ exists in } \mathcal{F} \right\}.$$

Remark II.2. Assumption II.1 ensures that the domain $\mathcal{D}(\mathcal{A})$ is dense in $\mathcal{F}$ and the generator $\mathcal{A}$ is a closed operator[38]. This assumption is typically satisfied by a broad class of stochastic dynamical systems when the drift $\mathbf{b}(\cdot)$ and diffusion $\boldsymbol{\sigma}(\cdot)$ coefficients satisfy certain regularity conditions and solutions are well-posed, e.g., a gradient potential system where the potential grows sufficiently fast at infinity. We will rely on Assumption II.1 throughout this paper, especially in Section IV C 1.

From Itô's formula[37], we have:

$$\mathcal{A}f = \sum_{i=1}^d \mathbf{b}_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d (\boldsymbol{\sigma}\boldsymbol{\sigma}^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad \forall f \in C_b^2(\mathcal{M}), \tag{4}$$

where $C_b^2(\mathcal{M})$ denotes the space of twice continuously differentiable functions with bounded derivatives on $\mathcal{M}$.

For spectral analysis of the Koopman generator $\mathcal{A}$, we consider the eigenvalue problem:

$$\mathcal{A}\phi = \lambda\phi,$$

where $\lambda \in \mathbb{C}$ and $\phi \in \mathcal{D}(\mathcal{A})$ are the eigenvalue and eigenfunction respectively. The eigenvalue λ of the Koopman generator $\mathcal{A}$ is closely connected to the eigenvalue μ of the stochastic Koopman operator $\mathcal{K}^t$ through the following relationship:

$$\mu = e^{t\lambda}. \tag{5}$$

This relationship provides a practical way to compute the generator's spectrum from discrete-time observations[39].

III. COMPUTATION METHOD IN STOCHASTIC DYNAMICAL SYSTEM

This section presents a computational method for analyzing stochastic dynamical systems through the lens of Koopman operator theory. The core idea involves utilizing the stochastic Taylor expansion[37] within the framework of the Galerkin method[40], which results in an approach tailored for stochastic systems. This approach, referred to as Stochastic Dynamic Mode Decomposition (SDMD), provides a data-driven framework for approximating the Koopman semigroup of stochastic systems. Below, we introduce some relevant background and the necessary notation.

Notation: Let $\{\psi_1, \ldots, \psi_N\} \subset \mathcal{D}(\mathcal{A})$ be a set of dictionary functions defined on the state space $\mathcal{M}$, forming the finite-dimensional space $\mathcal{F}_N := \mathrm{span}\{\psi_1, \ldots, \psi_N\}$. For these functions, we define the following Gram matrices $G, H \in \mathbb{R}^{N \times N}$

$$[G]_{ij} := \langle \psi_i, \psi_j \rangle_\rho, \quad [H]_{ij} := \langle \psi_i, \mathcal{A}\psi_j \rangle_\rho, \tag{6}$$

where the dictionary functions are assumed to be linearly independent in order to satisfy G's invertibility.

In practice, let $\{\mathbf{x}_k\}_{k=1}^m$ be the i.i.d. data sampled from ρ, i.e., each $\mathbf{x}_i$ is drawn independently and identically from the probability distribution ρ. Next, construct the data matrices $\Psi_X, \Psi'_X \in \mathbb{R}^{m \times N}$ in the following:

$$\Psi_X := \begin{bmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_m) & \cdots & \psi_N(\mathbf{x}_m) \end{bmatrix},$$

$$\Psi'_X := \begin{bmatrix} \mathcal{A}\psi_1(\mathbf{x}_1) & \cdots & \mathcal{A}\psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \mathcal{A}\psi_1(\mathbf{x}_m) & \cdots & \mathcal{A}\psi_N(\mathbf{x}_m) \end{bmatrix}. \tag{7}$$

Remark III.1. Since we manually pick up the basis functions, we can directly obtain the Jacobian and Hessian matrices as required for computing each $\mathscr{A}\psi_j(\mathbf{x}_i)$. However, in Neural Network based method, basis functions can be trained from a time series by Automatic Differentiation[41]. More details of such method will be discussed in Section V.

Thus, the Gram matrices G and H can be estimated empirically from these data matrices. Specifically, we construct

$$\widehat{G} = \frac{1}{m}\Psi_X^*\Psi_X, \quad \widehat{H} = \frac{1}{m}\Psi_X^*\Psi_X'. \qquad (8)$$

where $*$ denotes conjugate transpose of a matrix. These empirical approximation converges to their theoretical counterparts as the amount of data increases as discussed in[30,34]. Specifically, by the Strong Law of Large Numbers (SLLN), we have

$$\lim_{m\to\infty}[\widehat{G}]_{ij} = [G]_{ij}, \quad \lim_{m\to\infty}[\widehat{H}]_{ij} = [H]_{ij} \quad \text{a.s.} \qquad (9)$$

Building on these definitions, the following section introduces the SDMD method, which combines the Galerkin approximation framework with stochastic dynamics.

Remark III.2. The construction of the matrices $\widehat{G}$ and $\widehat{H}$ in Eq. (8) follows essentially the same procedure as in the gEDMD framework, where $\widehat{G}$ approximates the Gram matrix of the chosen dictionary functions and $\widehat{H}$ approximates the action of the generator on this dictionary. The main distinction lies in that SDMD estimates the semigroup instead of generator, but the underlying matrix assembly is consistent with gEDMD. See Klus *et al.*[34] for more details.

A. Stochastic Dynamic Mode Decomposition (SDMD)

The SDMD method approximates the stochastic Koopman operator by incorporating the stochastic Taylor expansion[37] into the EDMD framework. The method explicitly includes the sampling time Δt and truncates higher-order terms in the expansion. Specifically, SDMD estimates the Koopman operator as follows:

$$\widehat{K}_{N,\Delta t,m} := I + \Delta t\,\widehat{G}^{-1}\widehat{H}, \qquad (10)$$

where $\widehat{G}$ and $\widehat{H}$ are the Gram matrices computed from data as in Eq. (8). This formulation omits higher-order terms for sufficiently small Δt, enabling an efficient approximation of the stochastic Koopman operator. Below, we provide the derivation that leads to this result.

Derivation of SDMD Method

Consider the stochastic system defined in Eq. (1). Let

$$\Psi_N(\mathbf{x}_i) = [\psi_1(\mathbf{x}_i)\ \ldots\ \psi_N(\mathbf{x}_i)]^\top$$

be a vector of manually selected basis functions evaluated at some data point $\mathbf{x}_i$. Suppose $f(\mathbf{x}_i) = \Psi_N(\mathbf{x}_i)^\top\boldsymbol{a}$ for some $\boldsymbol{a} \in$

$\mathbb{R}^N$, then for any $\Delta t > 0$, EDMD[30] approximates the Koopman operator $\mathscr{K}^{\Delta t}$ onto $\mathscr{F}_N$ using Galerkin approximation:

$$\mathscr{K}^{\Delta t}f(\mathbf{x}_i) = \Psi_N(\mathbf{x}_i)^\top\widehat{K}_{N,m}\boldsymbol{a} + r(\mathbf{x}_i), \qquad (11)$$

where $r(\mathbf{x}_i)$ is the residual.

While EDMD provides a framework for approximating the Koopman operator, our SDMD method explicitly addresses the challenges of stochastic systems by incorporating stochastic Taylor expansion[37] in Eq. (3) to account for noise. Specifically, for each basis function ψ_j evaluated at $\mathbf{x}_i$

$$\mathscr{A}\psi_j(\mathbf{x}_i) \approx \frac{\mathscr{K}^{\Delta t}\psi_j(\mathbf{x}_i) - \psi_j(\mathbf{x}_i)}{\Delta t}.$$

After rearrange, we have

$$\mathscr{K}^{\Delta t}\psi_j(\mathbf{x}_i) \approx \psi_j(\mathbf{x}_i) + \Delta t\,\mathscr{A}\psi_j(\mathbf{x}_i) + o_{i,j}(\Delta t), \qquad (12)$$

where each $\mathscr{A}\psi_j(\mathbf{x}_i)$ is computed by Eq. (4) and $o_{i,j}(\Delta t)$ is the asymptotic term $o(\Delta t)$ corresponding to $\mathscr{K}^{\Delta t}\psi_j$ expanded at data point $\mathbf{x}_i$.

While SDMD truncates the stochastic Taylor expansion at the first-order term $\Delta t\,\mathscr{A}\psi_j$, the expansion can in principle be extended to include the second-order term $\frac{\Delta t^2}{2}\mathscr{A}^2\psi_j$. This would compute a higher-order approximation of the Koopman semigroup that potentially improves accuracy for larger sampling intervals Δt. However, computing $\mathscr{A}^2\psi_j$ requires higher-order derivatives of both the basis functions and the SDE coefficients, which greatly increases computational cost. The explicit form of $\mathscr{A}^2\psi_j$ for general SDEs is provided in Appendix A.

Remark III.3. The asymptotic term $o(\Delta t)$ represents the remainder terms in the Taylor expansion that decay faster than Δt as $\Delta t \to 0$. Specifically, for each basis function ψ_j evaluated at data point $\mathbf{x}_i$, we have:

$$\lim_{\Delta t \to 0}\frac{o_{i,j}(\Delta t)}{\Delta t} = 0.$$

This notation is used to indicate that these terms become negligible compared to the linear term Δt for sufficiently small sampling time steps, which justifies their omission in the approximation Eq. (10) when Δt is small.

Next, we use Eq. (12) for each basis function ψ_j in the expansion of the stochastic Koopman operator to approximate the expected value of f in the stochastic dynamical system starting at $\mathbf{x}_i$, which gives:

$$\begin{aligned}
\mathscr{K}^{\Delta t}f(\mathbf{x}_i) &= \sum_{j=1}^n [\psi_j(\mathbf{x}_i) + \Delta t\,\mathscr{A}\psi_j(\mathbf{x}_i) + o_{i,j}(\Delta t)]\,a_j \\
&= \left[\Psi_N(\mathbf{x}_i)^\top + \Delta t\,\mathscr{A}_N^{\Psi}(\mathbf{x}_i)^\top + \boldsymbol{o}_{i,N}(\Delta t)^\top\right]\boldsymbol{a},
\end{aligned} \qquad (13)$$

where $\boldsymbol{o}_{i,N}(\Delta t) = [o_{i,1}(\Delta t)\ \ldots\ o_{i,N}(\Delta t)]^\top$ and $\mathscr{A}_N^{\Psi}(\mathbf{x}_i) = [\mathscr{A}\psi_1(\mathbf{x}_i)\ \ldots\ \mathscr{A}\psi_N(\mathbf{x}_i)]^\top$.

Now, equating both Eq. (11) and Eq. (13), then evaluating over all data points $\{\mathbf{x}_i\}_{i=1}^m$, we can have the following minimization problem:

$$\min_{\tilde{K}_{N,\Delta t,m}\in\mathbb{R}^{N\times N}}\sum_{i=1}^m |r(\mathbf{x}_i)|^2$$
$$= \min_{\tilde{K}_{N,\Delta t,m}\in\mathbb{R}^{N\times N}}\sum_{i=1}^m \Big|\big[\mathbf{\Psi}_N(\mathbf{x}_i)^\top + \Delta t\,\mathscr{A}_N^{\mathbf{\Psi}}(\mathbf{x}_i)^\top + \mathbf{o}_{i,N}(\Delta t)^\top$$
$$- \mathbf{\Psi}_N(\mathbf{x}_i)^\top \tilde{K}_{N,\Delta t,m}\big]\mathbf{a}\Big|^2,$$

which is equivalent to:

$$\min_{\tilde{K}_{N,\Delta t,m}\in\mathbb{R}^{N\times N}}\|\mathbf{\Psi}_X + \Delta t\,\mathbf{\Psi}_X' + \mathbf{o}_{m,N}(\Delta t) - \mathbf{\Psi}_X\tilde{K}_{N,\Delta t,m}\|_F^2, \quad (14)$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm and the higher order matrix $\mathbf{o}_{m,N}(\Delta t)$ is $[\mathbf{o}_{m,N}(\Delta t)]_{ij} = o_{i,j}(\Delta t)$. More specifically,

$$\mathbf{o}_{m,N}(\Delta t) = \begin{bmatrix} o_{1,1}(\Delta t) & \cdots & o_{1,N}(\Delta t) \\ \vdots & \ddots & \vdots \\ o_{m,1}(\Delta t) & \cdots & o_{m,N}(\Delta t) \end{bmatrix}.$$

Thus, the minimal $\tilde{K}_{N,\Delta t,m}$ is

$$\tilde{K}_{N,\Delta t,m} = \mathbf{\Psi}_X^\dagger(\mathbf{\Psi}_X + \Delta t\,\mathbf{\Psi}_X' + \mathbf{o}_{m,N}(\Delta t))$$
$$= I + \Delta t\,(\mathbf{\Psi}_X^*\mathbf{\Psi}_X)^{-1}\left(\mathbf{\Psi}_X^*\mathbf{\Psi}_X'\right) + \mathbf{\Psi}_X^\dagger\mathbf{o}_{m,N}(\Delta t)$$
$$= \widehat{K}_{N,\Delta t,m} + \mathbf{\Psi}_X^\dagger\mathbf{o}_{m,N}(\Delta t), \quad (15)$$

where † denotes the pseudoinverse. When Δt is very small, we omit $\mathbf{o}_{m,N}(\Delta t)$ and keep $\widehat{K}_{N,\Delta t,m}$ as in the Eq. (10).

Remark III.4. We typically compute $(\widehat{G}+\gamma I)^{-1}$ instead of $\widehat{G}^{-1}$ for some small number $\gamma > 0$ to avoid singularity.

B. Computing Algorithm

This section presents the algorithmic implementation of our method. To provide a clear understanding of the computational procedure, we first present a flow chart in Figure 1 that illustrates the key steps of our approach. Following the flow chart, we provide a detailed pseudocode in Algorithm 1 that formalizes the computational steps. The algorithm takes as input the time series data and system parameters, and outputs the approximated Koopman operator. Each step in the algorithm corresponds to the theoretical framework developed in Section III A, ensuring a complete implementation of our method.

The coefficients $\mathbf{b}(x)$ and $\sigma(x)$ of the stochastic differential equation (SDE) can either be assumed as known for predefined models or estimated from sampled time-series data (see D). Specifically, for each independently and identically distributed (i.i.d.) initial point x_k, these coefficients can be approximated using discrete-time methods based on observed
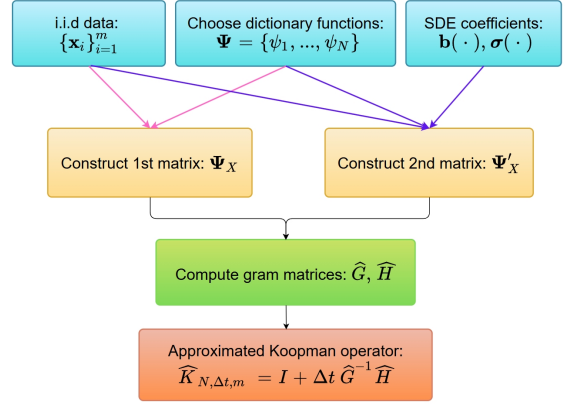

FIG. 1: A flow chart for SDMD method.

trajectories. For instance, $\mathbf{b}(x)$ can be derived from finite differences to approximate the derivative, while $\sigma(x)$ can be inferred from the covariance of the increments.

Algorithm 1 Estimation of stochastic Koopman operator

Require: i.i.d. data $\{\mathbf{x}_k\}_{k=1}^m$, dictionary functions $\{\psi_1,\ldots,\psi_N\}$, SDE coefficients $\mathbf{b}(\cdot),\sigma(\cdot)$, sampling time step Δt, regularization parameter γ.
1: Construct matrices: $\mathbf{\Psi}_X, \mathbf{\Psi}_X'$ (See Eq. (7)).
2: Construct empirical Gram matrices:

$$\widehat{G} = (1/m)\mathbf{\Psi}_X^*\mathbf{\Psi}_X, \quad \widehat{H} = (1/m)\mathbf{\Psi}_X^*\mathbf{\Psi}_X' \text{ (See Eq. (8)).}$$

3: Compute: $\widehat{K}_{N,\Delta t,m} = I + \Delta t\,\widehat{G}^{-1}\widehat{H}$ (See Eq. (10)).
Ensure: Approximated Koopman operator $\widehat{K}_{N,\Delta t,m}$.

Remark III.5. Although in our implementation the inverse of $\hat{G}+\gamma I$ in Eq. (10) is computed via Cholesky factorization for numerical stability, the spectral computation can also be formulated as a generalized eigenvalue problem. Specifically, we have

$$\widehat{H}v = \frac{\mu-1}{\Delta t}\widehat{G}v, \quad \lambda = \frac{\mu-1}{\Delta t},$$

where μ and v are the approximated eigenpair of $\mathscr{K}$, and λ corresponds to the approximated eigenvalue of the generator $\mathscr{A}$.

C. Computing in Deterministic Continuous-Time Systems

While the SDMD framework is designed with stochastic systems in mind, the core idea of approximating the Koopman semigroup using a truncated Taylor expansion can also be applied to deterministic continuous-time systems. In such cases, the underlying dynamics are governed by an ODE:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{b}(\mathbf{x}),$$

and the corresponding Koopman generator becomes a Lie derivative along the vector field $\mathbf{b}(\mathbf{x})$:

$$\mathscr{A}f(\mathbf{x}) = \mathbf{b}(\mathbf{x}) \cdot \nabla f(\mathbf{x}).$$

Then, the Taylor expansion of the semigroup remains: $\mathscr{K}^{\Delta t} f(\mathbf{x}) \approx f(\mathbf{x}) + \Delta t \mathscr{A}f(\mathbf{x})$, which leads to a similar approximation formula as in Eq.(10) where $\mathscr{A}\psi_j(\mathbf{x}) = \mathbf{b}(\mathbf{x}) \cdot \nabla \psi_j(\mathbf{x})$. This can be viewed as a continuous-time version to the standard EDMD method[30], which is formulated for discrete-time snapshot pairs, and potentially bridges the gap between EDMD and generator-based methods like gEDMD[34].

Remark III.6. It should be noted that the time-difference approach of EDMD and gEDMD may produce significantly different results even in the limit of high sampling rate ($\Delta t \to 0$), particularly when the data sampling measure does not satisfy certain regularity conditions. However, SDMD additionally truncated the higher order term $o(\Delta t)$ from stochastic Taylor expansion to enhance the numerical stability. See Section 4.3 of Bramburger and Fantuzzi[42], Section 4.2 of Mauroy[43] and Section V B for more details.

IV. CONVERGENCE ANALYSIS

This section analyzes the convergence of our stochastic Koopman operator approximation scheme. The analysis follows a sequential framework, where three fundamental regimes are examined in order:

Large Data Convergence: We first establish the convergence of the empirical approximation by deriving probabilistic error bounds that explicitly quantify the convergence rate as the sample size $m \to \infty$.

Zero-Limit of Sampling Time: Next we prove that time-discretized approximation converges to the true Koopman generator as the sampling interval $\Delta t \to 0$. This limit builds on large data convergence and connects discrete-time computations to continuous dynamics.

Large Dictionary Size Convergence: We finally demonstrate that finite-dimensional approximation of the Koopman generator and semigroups converges to their infinite-dimensional counterparts as dictionary size $N \to \infty$. This step completes the convergence analysis.

This structured approach $\lim_{N \to \infty} \lim_{\Delta t \to 0} \lim_{m \to \infty}$ highlights the critical interplay between these limits and ensures rigorous convergence results through appropriate function spaces.

The convergence analysis will be based on the following assumption:

Assumption IV.1. Let $\mathscr{O} \subset \mathscr{M}$ be a compact subset that contains all sampled data points $\mathbf{x}_1, \ldots, \mathbf{x}_m$. We assume that for each basis function ψ_i, both ψ_i and $\mathscr{A}\psi_i$ are uniformly bounded on $\mathscr{O}$, i.e., there exist constants $C = C_{N,\mathscr{O}}, L = L_{N,\mathscr{O}} > 0$ such that:

$$|\psi_i(x)| \le C, \quad |\mathscr{A}\psi_i(x)| \le L \quad \text{for all } x \in \mathscr{O}, \ i = 1, \ldots, N.$$

A. Convergence in the Limit of Large Data

Fix dictionary size N and sampling time step Δt. Denote $K_{N,\Delta t}$ by the matrix that represents the large data limit of $\tilde{K}_{N,\Delta t,m}$ given in Eq. (15):

$$
\begin{aligned}
K_{N,\Delta t} &= \lim_{m \to \infty} \tilde{K}_{N,\Delta t,m} \\
&= \lim_{m \to \infty} \left(\widehat{K}_{N,\Delta t,m} + \Psi_X^\dagger \boldsymbol{o}_{m,N}(\Delta t) \right) \\
&= I + \Delta t \, G^{-1} H + \boldsymbol{o}_N(\Delta t), \quad (16)
\end{aligned}
$$

where $\boldsymbol{o}_N(\Delta t) = \lim_{m \to \infty} \Psi_X^\dagger \boldsymbol{o}_{m,N}(\Delta t)$. Note that, $\boldsymbol{o}_N(\Delta t)$ exists since both $K_{N,\Delta t}$ exists from EDMD theory[30,44] and $\lim_{m \to \infty} \widehat{K}_{N,\Delta t,m} = I + \Delta t \, G^{-1} H$ exists due to Eq. (9).

Remark IV.2. Each element in the matrix $\Psi_X^\top \boldsymbol{o}_{m,N}(\Delta t)$ is $o(\Delta t)$. After taking the large data limit $m \to \infty$, each element in $\boldsymbol{o}_N(\Delta t)$ is still $o(\Delta t)$.

Now we aim to prove that the empirical Koopman matrix $\widehat{K}_{N,\Delta t,m}$ converges to its large-data limit $K_{N,\Delta t}$ as $m \to \infty$; more specifically, we shall prove a concentration inequality that quantifies the probability of their difference exceeding any given threshold $\varepsilon > 0$. Such a probabilistic bound will demonstrate that the empirical approximation becomes increasingly accurate as the sample size m grows. Specifically, we aim to bound the following probabilistic error bound in the following theorem:

Theorem IV.3. Let $\varepsilon > 0$. Define $\tilde{\varepsilon} = (\varepsilon - \|\boldsymbol{o}_N(\Delta t)\|_F)/\Delta t$. Then, with same notation and conditions defined in Lemma IV.5 and Lemma IV.6, we have

$$
\begin{aligned}
\mathbb{P}\Big(\|\widehat{K}_{N,\Delta t,m} - K_{N,\Delta t}\|_F > \varepsilon \Big) \le 2N^2 &\left[\exp\Big(-\frac{m}{8} \big(\frac{\tilde{\varepsilon}\|G^{-1}\|^{-1}}{N^{3/2}C^2\tau_\varepsilon} \big)^2 \Big) \right. \\
&\left. + \exp\Big(-\frac{m}{8} \big(\frac{\tilde{\varepsilon}\|H\|}{N^{3/2}C\tau_\varepsilon L} \big)^2 \Big) \right],
\end{aligned}
$$

where $\|\cdot\|$ denotes the matrix operator norm.

We will introduce the following lemmas in order to prove Theorem IV.3. Note that, Lemma IV.4 is a result of McDiarmid inequality B.2.

Lemma IV.4 (Hoeffding's Inequality[45]). Assume $X_1, \ldots, X_n$ are i.i.d. with each $X_i \in [a_i, b_i]$ for all $1 \le i \le n$. Then, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i \right] \right| \ge \varepsilon \right) \le 2e^{-\frac{2\varepsilon^2}{\Sigma_{i=1}^n (b_i - a_i)^2}}.$$

Lemma IV.5. Suppose Assumption IV.1 holds with uniform bounds $C, L > 0$. Then, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left\| \widehat{G} - G \right\|_F \ge \varepsilon \right) \le 2N^2 \exp\left(-\frac{m\varepsilon^2}{8N^2C^4} \right),$$

$$\mathbb{P}\left(\left\| \widehat{H} - H \right\|_F \ge \varepsilon \right) \le 2N^2 \exp\left(-\frac{m\varepsilon^2}{8N^2C^2L^2} \right).$$

Proof. Let $G, \widehat{G}$ be defined as in Section 2. Define $\eta_{ij}(\mathbf{x}_k) := \psi_i(\mathbf{x}_k)\psi_j(\mathbf{x}_k)$. Then,

$$\|\widehat{G} - G\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\widehat{G}_{ij} - G_{ij}\right|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{m}\sum_{k=1}^{m}\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2,$$

where $\tilde{\eta}_{ij}(\mathbf{x}_k) := \eta_{ij}(\mathbf{x}_k) - \mathbb{E}\left[\eta_{ij}(\mathbf{x}_k)\right] = \eta_{ij}(\mathbf{x}_k) - \mathbb{E}\left[\eta_{ij}(\mathbf{x}_1)\right]$ with $\{\mathbf{x}_k\}_{k=1}^{m}$ being i.i.d. and thus $\left|\tilde{\eta}_{ij}(\mathbf{x}_k)\right| \le 2C^2$ for all $1 \le i, j \le N$. Next, applying Hoeffding's inequality from Lemma IV.4, we have

$$\mathbb{P}\left(\left\|\widehat{G} - G\right\|_F \ge \varepsilon\right) \le \mathbb{P}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{m}\sum_{k=1}^{m}\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2 \ge \varepsilon^2\right)$$

$$\le \sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{P}\left(\left|\frac{1}{m}\sum_{k=1}^{m}\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2 \ge \varepsilon^2/N^2\right)$$

$$\le N^2\mathbb{P}\left(\frac{1}{m}\sum_{k=1}^{m}\max_{1\le i,j\le N}\left|\tilde{\eta}_{ij}(\mathbf{x}_k)\right| \ge \varepsilon/N\right)$$

$$\le 2N^2\exp\left(-\frac{m\varepsilon^2}{8N^2C^4}\right).$$

Similarly, since $\left|\psi_i(\mathbf{x}_k)\mathscr{A}\psi_j(\mathbf{x}_k)\right| \le CL$, we have

$$\mathbb{P}\left(\|\widehat{H} - H\|_F \ge \varepsilon\right) \le 2N^2\exp\left(-\frac{m\varepsilon^2}{8N^2C^2L^2}\right).$$

$\square$

Lemma IV.6 (Lemma C.5[46]). Let $G, H \in \mathbb{R}^{N\times N}$ be such that G is invertible and $H \ne 0$. Let $\widehat{G}, \widehat{H} \in \mathbb{R}^{N\times N}$ be random matrices such that $\widehat{G}$ is invertible a.s. Then for any $\varepsilon > 0$ we have

$$\mathbb{P}\left(\|G^{-1}H - \widehat{G}^{-1}\widehat{H}\| > \varepsilon\right) \le \mathbb{P}\left(\|H - \widehat{H}\| > \frac{\varepsilon}{\tau_\varepsilon}\|H\|\right)$$
$$+ \mathbb{P}\left(\|G - \widehat{G}\| > \frac{\varepsilon}{\tau_\varepsilon}\|G^{-1}\|^{-1}\right),$$

where $\tau_\varepsilon = 2\|G^{-1}\|\|H\| + \varepsilon$.

Remark IV.7. We choose operator norm $\|\cdot\|$ here since it is submultiplicative, i.e., $\|GH\| \le \|G\|\|H\|$ while Frobenious norm $\|\cdot\|_F$ is not.

Proof of Theorem IV.3. Recall that $\tilde{\varepsilon} = \left(\varepsilon - \|\mathbf{o}_N(\Delta t)\|_F\right)/\Delta t$

as in the statement of Theorem IV.3,

$$\mathbb{P}\left(\|\widehat{K}_{N,\Delta t,m} - K_{N,\Delta t}\|_F > \varepsilon\right)$$
$$= \mathbb{P}\left(\|\Delta t\ \widehat{G}^{-1}\widehat{H} - \Delta t\ G^{-1}H + \mathbf{o}_N(\Delta t)\|_F > \varepsilon\right)$$
$$\le \mathbb{P}\left(\Delta t\ \|\widehat{G}^{-1}\widehat{H} - G^{-1}H\|_F + \|\mathbf{o}_N(\Delta t)\|_F > \varepsilon\right)$$
$$= \mathbb{P}\left(\|\widehat{G}^{-1}\widehat{H} - G^{-1}H\|_F > \tilde{\varepsilon}\right)$$
$$\le \mathbb{P}\left(\|\widehat{G}^{-1}\widehat{H} - G^{-1}H\| > \tilde{\varepsilon}/\sqrt{N}\right)$$
$$\le \mathbb{P}\left(\|G - \widehat{G}\| > \frac{\tilde{\varepsilon}}{\sqrt{N}\tau_\varepsilon}\|G^{-1}\|^{-1}\right)$$
$$\qquad + \mathbb{P}\left(\|H - \widehat{H}\| > \frac{\tilde{\varepsilon}}{\sqrt{N}\tau_\varepsilon}\|H\|\right)$$
$$\le 2N^2\left[\exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}\|G^{-1}\|^{-1}}{N^{3/2}C^2\tau_\varepsilon}\right)^2\right)\right.$$
$$\left. + \exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}\|H\|}{N^{3/2}C\tau_\varepsilon L}\right)^2\right)\right].$$

$\square$

B. Convergence in the Zero-Limit of Sampling Time

Define the matrix $A_{N,\Delta t}$ as

$$A_{N,\Delta t} := \frac{K_{N,\Delta t} - I}{\Delta t} = G^{-1}H + \frac{\mathbf{o}_N(\Delta t)}{\Delta t}, \qquad (17)$$

where we use Eq. (16) for $K_{N,\Delta t}$ in the last equality; and let $\mathscr{A}_{N,\Delta t} : \mathscr{F}_N \to \mathscr{F}_N$ be the linear map defined by

$$(\psi_1, \ldots, \psi_N)\mathbf{a} \mapsto (\psi_1, \ldots, \psi_N)A_{N,\Delta t}\mathbf{a},$$

for any coefficient vector $\mathbf{a} \in \mathbb{R}^N$.

Let $\mathscr{A}_N := \mathscr{P}_N\mathscr{A}\mathscr{P}_N$ be the finite dimensional approximation of Koopman generator $\mathscr{A}$ where we denote $\mathscr{P}_N$ by the orthorgnal projection from $\mathscr{F}$ to $\mathscr{F}_N$ and $\mathscr{F}_N = \text{span}\{\psi_1, \ldots, \psi_N\}$ as mentioned in last section.

In the following Theorem IV.8, we will show that this time-discretized approximant $\mathscr{A}_{N,\Delta t}$ converges to $\mathscr{A}_N$ in operator norm as $\Delta t \to 0$.

Theorem IV.8. Let $\mathscr{A}_{N,\Delta t}$ and $\mathscr{A}_N$ be defined as above. For each dictionary size $N > 0$, we have

$$\lim_{\Delta t \to 0}\|\mathscr{A}_{N,\Delta t} - \mathscr{A}_N\| = 0.$$

Proof of Theorem IV.8. Define $A_N := G^{-1}H$ where matrices G, H are given in Eq. (9). Applying Galerkin approximation for the Koopman generator $\mathscr{A}$, we have that A_N is the matrix representation of $\mathscr{A}_N$ on $\mathscr{F}_N$ as shown in Proposition 3.5 of Klus *et al.*[34], i.e.,

$$\mathscr{A}_N\big((\psi_1, \ldots, \psi_N)\mathbf{a}\big) = (\psi_1, \ldots, \psi_N)A_N\mathbf{a},$$

for any coefficient vector $\mathbf{a} \in \mathbb{R}^N$. From Eq. (17), we know that

$$A_{N,\Delta t} = A_N + o_N(\Delta t)/\Delta t.$$

Since each element in $o_N(\Delta t)$ is $o(\Delta t)$ as pointed out in Remark IV.2, we have $\lim_{\Delta t \to 0} \|A_{N,\Delta t} - A_N\| = \lim_{\Delta t \to 0} \|o_N(\Delta t)\|/\Delta t = 0$. The conclusion of the theorem follows. $\square$

Remark IV.9. In finite dimensional space, the uniform operator topology is equivalent to strong operator topology.

C. Convergence in the Limit of Large Dictionary Size

This section establishes the following two convergence results of our approximation scheme for stochastic Koopman operator as the dictionary size $N \to \infty$.

Generator Convergence Under Assumption IV.10, the finite dimensional generator $\mathscr{A}_N = \mathscr{P}_N \mathscr{A} \mathscr{P}_N$ strongly converges to the true generator $\mathscr{A}$ strongly as $N \to \infty$.

Semigroup Convergence Under Assumption IV.13, the approximated semigroups $(e^{t\mathscr{A}_N})_{t \geq 0}$ strongly converge to the true Koopman semigroup $(e^{t\mathscr{A}})_{t \geq 0}$ uniformly over compact time interval as $N \to \infty$.

The proof builds upon the established framework in Section 4 of Llamazares-Elias *et al.*[47] and utilizes the Trotter-Kato Approximation theorem[39], which will be introduced systematically in the following subsections.

1. Convergence of Finite Dimensional Koopman Generator

By Assumption II.1 in Section II, we define the inner product

$$\langle f, g \rangle_{\mathscr{A}} := \langle f, g \rangle_\rho + \langle \mathscr{A}f, \mathscr{A}g \rangle_\rho, \quad \forall f, g \in \mathscr{D}(\mathscr{A}),$$

and the corresponding graph norm

$$\|f\|_{\mathscr{A}} := \sqrt{\langle f, f \rangle_{\mathscr{A}}}.$$

Clearly, the generator satisfies $\|\mathscr{A}\|_{\mathscr{D}(\mathscr{A}) \to \mathscr{F}} := \sup_{\|f\|_{\mathscr{A}} = 1} \|\mathscr{A}f\|_\rho \leq 1$.

Recall that $\mathscr{P}_N$ is projection of $\mathscr{F}$ onto $\mathscr{F}_N$ equipped with $\|\cdot\|_\rho$. Denote by $\widetilde{\mathscr{P}}_N$ the projection of $\mathscr{D}(\mathscr{A})$ onto $\mathscr{F}_N$. We now introduce the following assumptions:

Assumption IV.10. We assume that the dictionary Ψ satisfies the following conditions:

- $\lim_{N \to \infty} \|(\mathscr{P}_N - I)f\|_\rho = 0, \quad \forall f \in \mathscr{F}.$

- $\lim_{N \to \infty} \|(\widetilde{\mathscr{P}}_N - I)f\|_{\mathscr{A}} = 0, \quad \forall f \in \mathscr{D}(\mathscr{A}),$

Remark IV.11. The second part of Assumption IV.10 is moderately more restrictive than first part, as it requires convergence in the graph norm of the generator. This assumption is standard in the Galerkin approximation of unbounded operators (e.g., see Assumption 4 of Llamazares-Elias *et al.*[47])

but may not be satisfied for arbitrary dictionaries. However, in practice, data are typically collected in a compact subset $\mathscr{O} \subset \mathscr{M}$, so the restriction to $\mathscr{O}$ makes this assumption much less limiting. On such compact domain, this condition can often be approximated by selecting a sufficiently smooth dictionary.

The next result shows the strong convergence of $\mathscr{A}_N$ to $\mathscr{A}$ as $N \to \infty$.

Theorem IV.12. Suppose Assumption IV.10 holds. Then for all $f \in \mathscr{D}(\mathscr{A})$,

$$\lim_{N \to \infty} \|\mathscr{A}_N f - \mathscr{A}f\|_\rho = 0.$$

Proof of Theorem IV.12. For any $f \in \mathscr{D}(\mathscr{A})$, let $\widetilde{f}_N := \widetilde{\mathscr{P}}_N f$. Then,

$$\begin{aligned}
&\|\mathscr{A}_N f - \mathscr{A}f\|_\rho \\
&= \|\mathscr{P}_N \mathscr{A} \mathscr{P}_N f - \mathscr{A}f\|_\rho \\
&= \|\mathscr{P}_N \mathscr{A} \mathscr{P}_N (f - \widetilde{f}_N) + \mathscr{P}_N \mathscr{A} \mathscr{P}_N \widetilde{f}_N - \mathscr{A}f\|_\rho \\
&\leq \underbrace{\|\mathscr{P}_N \mathscr{A} \mathscr{P}_N (f - \widetilde{f}_N)\|_\rho}_{\text{first term}} + \underbrace{\|\mathscr{P}_N \mathscr{A} \mathscr{P}_N \widetilde{f}_N - \mathscr{A}f\|_\rho}_{\text{second term}}.
\end{aligned}$$

The first term $\|\mathscr{P}_N \mathscr{A} \mathscr{P}_N (f - \widetilde{f}_N)\|_\rho \to 0$ as $N \to \infty$ follows immediately from the second part of Assumption IV.10. The convergence of the second term is given by the following:

$$\begin{aligned}
&\|\mathscr{P}_N \mathscr{A} \mathscr{P}_N \widetilde{f}_N - \mathscr{A}f\|_\rho \\
&= \|\mathscr{P}_N \mathscr{A} \widetilde{\mathscr{P}}_N f - \mathscr{A}f\|_\rho \\
&= \|(\mathscr{P}_N - I + I)\mathscr{A} \widetilde{\mathscr{P}}_N f - \mathscr{A}f\|_\rho \\
&\leq \|(\mathscr{P}_N - I)\mathscr{A} \widetilde{\mathscr{P}}_N f\|_\rho + \|\mathscr{A}\|_{\mathscr{D}(\mathscr{A}) \to \mathscr{F}} \|\widetilde{\mathscr{P}}_N f - f\|_{\mathscr{A}} \\
&\to 0 \quad \text{as } N \to \infty,
\end{aligned}$$

where we use Assumption IV.10 and $\|\mathscr{A}\|_{\mathscr{D}(\mathscr{A}) \to \mathscr{F}} \leq 1$ in the last limit. $\square$

2. Convergence of Finite Dimensional Koopman Semigroups

Semigroups generated by $\mathscr{A}_N$ and $\mathscr{A}$ are denoted by: $\mathscr{K}_N^t = e^{t\mathscr{A}_N}$ and $\mathscr{K}^t = e^{t\mathscr{A}}$ respectively. To establish the convergence of $\mathscr{K}_N^t$ to $\mathscr{K}^t$ as $N \to \infty$, we first introduce the following assumptions:

Assumption IV.13. There exists a core $\mathscr{D}[39]$ for $\mathscr{A}$, that is a linear subspace $\mathscr{D} \subseteq \mathscr{D}(\mathscr{A})$ and dense in $\mathscr{D}(\mathscr{A})$ with respect to $\|\cdot\|_{\mathscr{A}}$ such that:

- $\mathscr{D} \subseteq \mathscr{D}(\mathscr{A}_N)$, for all $N \in \mathbb{N}$,

- $\forall f \in \mathscr{D}, \mathscr{A}_N f \to \mathscr{A}f$ in $\mathscr{F}$ as $N \to \infty$.

Now we establish the strong convergence of $\mathscr{K}_N^t$ to $\mathscr{K}^t$ uniformly over compact time interval as $N \to \infty$ based on Theorem IV.12 and Assumption IV.13 in the following result:

Theorem IV.14. There exists some constants $D \geq 1$ and $\omega \in \mathbb{R}$ such that the semigroups satisfy

$$\|\mathscr{K}_N^t\|, \|\mathscr{K}^t\| \leq De^{\omega t} \quad \text{for all } t \geq 0, N \in \mathbb{N}.$$

Furthermore, for every $f \in \mathscr{F}$ and every $T > 0$, we have

$$\lim_{N \to \infty} \sup_{t \in [0,T]} \|\mathscr{K}_N^t f - \mathscr{K}^t f\|_\rho = 0.$$

Proof of Theorem IV.14. Since $\mathscr{K}_N^t$ are generated by a finite dimensional operator $\mathscr{A}_N$ for each $N \in \mathbb{N}$, they are strongly continuous; and also $\mathscr{K}^t$ is strongly continuous by Assumption II.1. Thus, the exponential boundedness in the first part is proved according to Theorem 2.2 of Pazy[38]. The second part of the theorem is an immediate consequence of Trotter–Kato Approximation Theorem[39]. $\qquad\square$

Remark IV.15. The Trotter–Kato Approximation Theorem[39] ensures uniform semigroup convergence on compact time intervals once strong resolvent convergence of the approximate generators is established, i.e., $(\lambda - \mathscr{A}_N)^{-1}f \to (\lambda - \mathscr{A})^{-1}f$ for all $f \in \mathscr{F}$, $\lambda > \omega$, together with other properties and assumptions. From our generator convergence result in the first part of the theorem, the required resolvent convergence follows directly, hence the semigroup convergence in the second part. See[39] (Section 4) for details.

Remark IV.16. In our setting, $\mathscr{K}^{\Delta t}$ approximates the one-step Koopman operator $e^{\Delta t \mathscr{A}}$. Iterating this operator computes a discrete-time semigroup $(\mathscr{K}^{\Delta t})^n$, which converges to the continuous semigroup $e^{t\mathscr{A}}$ as $\Delta t \to 0$. This convergence can be formally justified by the Trotter–Kato theorem under standard assumptions on the generator approximation.

V. SDMD WITH DICTIONARY LEARNING (SDMD-DL)

Manual selection of basis functions often requires domain expertise and prior knowledge of the system's dynamics, which can be challenging for complex stochastic systems. Additionally, hand-crafted dictionaries may not optimally capture the underlying state space geometry and lead to suboptimal approximation of the Koopman operator. To address these limitation, SDMD can be integrated with dictionary learning technique[48,49] that automatically extracts optimal basis functions directly from data[31,49–52].

A. Methodology and Discussion

Overall Framework and Design In SDMD-DL framework, the matrix $\Psi_X + \Delta t \; \Psi_X' + o_{m,N}(\Delta t)$ in Eq. (14) is replaced by the data matrix Ψ_Y and have the following minimization problem

$$\min_{\tilde{K}_{N,\Delta t,m} \in \mathbb{R}^{N \times N}} \|\Psi_Y - \Psi_X \tilde{K}_{N,\Delta t,m}\|_F^2, \tag{18}$$

where the matrix Ψ_Y is defined as

$$\Psi_Y := \begin{bmatrix} \psi_1(\mathbf{y}_1) & \cdots & \psi_N(\mathbf{y}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{y}_m) & \cdots & \psi_N(\mathbf{y}_m) \end{bmatrix},$$

and $\mathbf{y}_i$ represents the evolved state Δt time after $\mathbf{x}_i$ under stochastic dynamics for each $1 \leq i \leq m$, in other words, we can say that the state $\mathbf{y}_i$ is the realization of this stochastic evolution as given in Eq. (1) starting from the initial state $\mathbf{x}_i$ over Δt time. In SDMD-DL, a neural network is integrated to parameterize dictionary functions $\Psi(\mathbf{x}; \theta)$ for data-driven learning. This approach alternates between optimizing the approximated Koopman operator $\widehat{K}_{N,\Delta t,m}(\theta)$ by Eq. (10) and updating the neural network parameters θ using stochastic gradient descent for Eq. (18), ensuring that both the dictionary functions and the operator approximation improve iteratively.
Parameterization We now show more details of how to paramterize the basis functions and the training scheme. Denote $\Psi(\mathbf{x}; \theta) = [\psi_1(\mathbf{x}; \theta), \ldots, \psi_N(\mathbf{x}; \theta)]^\top$ by a dictionary parameterized by a neural network, where θ represents the trainable parameters. Both $\Psi_X = \Psi_X(\theta)$ and $\Psi_Y = \Psi_Y(\theta)$ are computed directly from the neural network and the Koopman operator $\widehat{K}_{N,\Delta t,m}(\theta)$ is then computed by Eq. (10) as following

$$\widehat{K}_{N,\Delta t,m}(\theta) = I + \Delta t \; \widehat{G}(\theta)^{-1}\widehat{H}(\theta).$$

The training process involves minimizing a loss function that balances the approximation quality of $\widehat{K}_{N,\Delta t,m}$ with regularization. Here the loss function is defined as:

$$J(\theta, \widehat{K}_{N,\Delta t,m}) = \|\Psi_Y(\theta) - \Psi_X(\theta)\widehat{K}_{N,\Delta t,m}(\theta)\|_F^2 + \gamma R(\widehat{K}_{N,\Delta t,m}),$$

where $\gamma > 0$ is some small positive number and $R(\widehat{K}_{N,\Delta t,m})$ is the regularization term. In this work, we use Tikhonov regularization $R(\widehat{K}_{N,\Delta t,m}) = \|\widehat{K}_{N,\Delta t,m}(\theta)\|_F^2$. We show the pseudocode in Algorithm 2.

Remark V.1. For the estimation of SDE coefficients, see Appendix D.

B. Comparison with Other Methods

First we compare SDMD-DL with EDMD-DL[49], which extends traditional EDMD by replacing manually selected dictionaries with ones learned from data using a simple feedforward neural network. While SDMD-DL shares similarities with EDMD-DL in the loss function computing formula, a key distinction lies in how $\widehat{K}_{N,\Delta t,m}$ is updated. SDMD-DL employs Eq. (10), leveraging the stochastic Taylor expansion and incorporating the sampling time Δt. In contrast, EDMD-DL uses $\widehat{K}_{N,\Delta t,m} = (\Psi_X^*\Psi_X)^{-1}(\Psi_X^*\Psi_Y)$, which is better suited for deterministic systems without consideration of stochasticity.

Next, we can also compare our method to gEDMD with dictionary learning setting, i.e., gEDMD-DL, which is almost same as the EDMD-DL's framework except that the loss function in gEDMD-DL is defined to minimize the linear regression error associated with the generator $\mathscr{A}$ instead of semigroup $\mathscr{K}^{\Delta t}$ as in EDMD-DL. This contains the computation

of $\Psi_X^* \Psi_X$ and $\Psi_X^* \Psi_X'$, where Ψ_X' involves evaluating Jacobian and Hessian matrices as given in Eq. (7). However, calculation of Ψ_X' can be extremely computationally expensive, especially for large datasets, as they require repeatedly computing higher-order derivatives during the Automatic Differentiation process and cross validation process. In contrast, SDMD-DL defines its loss function directly using Ψ_Y, the parameterized dictionary applied to the evolved data, instead of Ψ_X'. By directly utilizing Ψ_Y, SDMD-DL avoids the heavy evaluation of Ψ_X', thus significantly reducing computational overhead. Furthermore, the use of Ψ_Y ensures consistency with the stochastic Koopman operator's evolution, which can enable the approximation of the Koopman operator in stochastic systems more accurately.

The following Table I gives a comparison of each method's distinction in the aspects of both loss function and updating formula:

Algorithm 2 Estimation of stochastic Koopman operator with dictionary learning

---

Require: i.i.d. initial data points $\{\mathbf{x}_k\}_{k=1}^m$ and their n-step time series trajectories $\{\{\mathbf{y}_k^{(j)}\}_{j=1}^n\}_{k=1}^m$, dictionary size N, sampling time step Δt, regularization parameter $\gamma > 0$, learning rate $\eta > 0$, number of epochs $T > 0$.

1: Estimate SDE coefficients $\mathbf{b}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ from the time series data (See Appendix D).
2: Initialize neural network parameters θ.
3: Initialize Koopman operator $\widehat{K}_{N,\Delta t,m}$.
4: **for** epoch = 1 to T **do**
5:     Compute: $\Psi_X(\theta), \Psi_Y(\theta), \Psi_X'(\theta)$ using $\{(\mathbf{x}_k, \mathbf{y}_k^{(j)})\}_{k=1,j=1}^{m,n}$, $\mathbf{b}(\cdot)$, and $\boldsymbol{\sigma}(\cdot)$.
6:     Compute Gram matrices:

$$\widehat{G}(\theta) = \frac{1}{m} \Psi_X(\theta)^* \Psi_X(\theta),$$

$$\widehat{H}(\theta) = \frac{1}{m} \Psi_X(\theta)^* \Psi_X'(\theta).$$

7:     Compute loss:

$$J(\theta, \widehat{K}_{N,\Delta t,m}) = \|\Psi_Y(\theta) - \Psi_X(\theta)\widehat{K}_{N,\Delta t,m}\|_F^2 + \gamma \|\widehat{K}_{N,\Delta t,m}\|_F^2.$$

8:     Update neural network parameters:

$$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta, \widehat{K}_{N,\Delta t,m}).$$

9:     Update approximated Koopman operator:

$$\widehat{K}_{N,\Delta t,m} = I + \Delta t \left(\widehat{G}(\theta) + \gamma I\right)^{-1} \widehat{H}(\theta).$$

10: **end for**
Ensure: Approximated Koopman operator $\widehat{K}_{N,\Delta t,m}(\theta)$.

VI. APPLICATION

This section evaluates the proposed Stochastic Dynamic Mode Decomposition (SDMD) framework through four representative experiments, each selected to highlight unique

---

EDMD-DL
Loss function:
$J(\theta, \widehat{K}_{N,m}) = \|\Psi_Y(\theta) - \Psi_X(\theta)\widehat{K}_{N,m}(\theta)\|_F^2 + \gamma R(\widehat{K}_{N,m})$
Updating formula:
$\widehat{K}_{N,m}(\theta) = \left(\Psi_X(\theta)^*\Psi_X(\theta) + \gamma I\right)^{-1}\left(\Psi_X(\theta)^*\Psi_Y(\theta)\right)$
gEDMD-DL
Loss function:
$J(\theta, \widehat{K}_{N,m}) = \|\Psi_X'(\theta) - \Psi_X(\theta)\widehat{A}_{N,m}(\theta)\|_F^2 + \gamma R(\widehat{A}_{N,m})$
Updating formula:
$\widehat{A}_{N,m}(\theta) = \left(\Psi_X(\theta)^*\Psi_X(\theta) + \gamma I\right)^{-1}\left(\Psi_X(\theta)^*\Psi_X'(\theta)\right)$
SDMD-DL
Loss function:
$J(\theta, \widehat{K}_{N,\Delta t,m}) = \|\Psi_Y(\theta) - \Psi_X(\theta)\widehat{K}_{N,\Delta t,m}(\theta)\|_F^2 + \gamma R(\widehat{K}_{N,\Delta t,m})$
Updating formula:
$\widehat{K}_{N,\Delta t,m}(\theta) = I + \Delta t \left(\Psi_X(\theta)^*\Psi_X(\theta) + \gamma I\right)^{-1}\left(\Psi_X(\theta)^*\Psi_X'(\theta)\right)$

TABLE I: Comparison test between EDMD-DL, gEDMD-DL and SDMD-DL.

features of stochastic dynamical systems and demonstrate SDMD's effectiveness. We first provide a brief overview of these examples and explain our objectives and accomplishments in applying SDMD. These experiments showcase the method's ability to approximate the Koopman semigroup and their spectral properties which offers insights into system behaviors such as oscillatory dynamics, decay rates, and metastable transitions, while also comparing SDMD's performance to established methods like EDMD and gEDMD.

The 2D Stuart-Landau system[53] is a well-known model in nonlinear dynamics, often used to study oscillatory behaviors perturbed by stochastic noise. It behaves as a limit cycle under some conditions whose radius and phase evolve under both deterministic and random influences, which makes it an excellent testbed for analyzing spectral properties in the presence of noise. In this experiment, we apply SDMD to capture the eigenvalues of the Koopman semigroup, aiming to accurately reflect the system's oscillatory patterns and phase diffusion. Our results demonstrate SDMD's superior ability to handle stochastic perturbation compared to traditional method like EDMD.

The 1D Ornstein-Uhlenbeck (OU) process[54] is a time-reversible system characterized by mean reversion driven by

a linear drift and Gaussian noise, with closed-form solution for its spectral properties. Here, we use SDMD to estimate the leading eigenpairs of the Koopman semigroup and achieve high precision that aligns with theoretical expectation.

The 2D Triple-Well system[55] models a metastable stochastic process with a potential landscape and is commonly studied in chemical physics and molecular dynamics. Its dynamics exhibit slow transitions between basins alongside rapid fluctuation within them, which arised a challenge for capturing long-term behaviors. With SDMD, we identify the slow timescales of basin transitions through the Koopman semigroup's eigenvalues and eigenfunctions, which successfully reveals the system's metastable structure. This experiment shows SDMD's strength in analyzing complex, time-reversible metastable systems and its advantage over methods like gEDMD in computational efficiency and accuracy.

The final example involves a minimal two-dimensional neural-mass model[56] that exhibits multiscale dynamics, including a slow-modulating latent input, an intermediate transient decay, and fast oscillations. Using only the observable state variables, SDMD accurately recovers eigenfunctions corresponding to these latent timescales, even in the presence of stochastic perturbation. In particular, SDMD reveals one eigenfunction that reflects the slow switching of input-modulated regimes, and another that captures transient dynamics during transitions. In contrast, EDMD with dictionary learning fails to isolate these features, which produces eigenfunctions corrupted by fast oscillations and noise. This example highlights SDMD's capability to infer hidden multiscale structures from limited, noisy measurements, which can demonstrate its robustness and interpretability in realistic settings.

The four experiments use the Euler-Maruyama (EM) method for numerical integration. In the two examples of OU-process and triple-well system, we observe that the approximated eigenvalues are purely real, consistent with their time-reversible property. We also note that the drift and diffusion coefficients used in the 2D Stuart-Landau system are analytical values in polar coordinates, whereas those in the 1D OU process and 2D Triple-Well system are estimated from collected data using a simple neural network (NN) instead of manually selecting basis functions. The error bound analysis of this coefficient estimation method is provided in[57]. The following subsections detail the setups and results. All experiments are available on our GitHub repository.

A. 2D Stuart-Landau Equation

The 2D stochastic Stuart-Landau (SL) equation[53] serves as a canonical example to study nonlinear oscillatory dynamics under stochastic perturbation. It is frequently used to validate numerical methods for estimating the stochastic Koopman operator since it has analytical Koopman eigenpairs expression. The equation in standard Cartesian coordinates is the following

$$dx = \left[\left(\delta - \kappa(x^2+y^2)\right)x - \left(\gamma - \beta(x^2+y^2)\right)y\right]dt + \varepsilon\,dW_x,$$
$$dy = \left[\left(\gamma - \beta(x^2+y^2)\right)x + \left(\delta - \kappa(x^2+y^2)\right)y\right]dt + \varepsilon\,dW_y,$$

where $R := \sqrt{\delta/\kappa}$ is the radius of the limit cycle, $\gamma > 0$ is the rotation frequency parameter that controls the linear rotational motion of the system and $\beta > 0$ is the nonlinear frequency parameter that controls amplitude-related frequency changes. The noise terms dW_x and dW_y are independent Wiener processes with intensity ε. Note that, a positive parameter $\delta > 0$ indicates that the system exhibit a stable limit cycle behavior. When $\delta < 0$, the system experiences negative radial growth, with trajectories spiraling inward toward the origin and the origin becomes a stable focus.

To simplify the analysis, the system is often transformed into polar coordinates

$$dr = \left(\delta r - \kappa r^3 + \frac{\varepsilon^2}{2r}\right)dt + \varepsilon\,dW_r,$$
$$d\theta = \left(\gamma - \beta r^2\right)dt + \frac{\varepsilon}{r}\,dW_\theta,$$

where r and θ represent the radius and angular position, respectively, and dW_r and dW_θ are derived Wiener processes. The analytical eigenvalues of the stochastic Stuart-Landau system's Kolmogorov operator, i.e., the Koopman generator, are given as:

$$\lambda_{ln} = \begin{cases} -\frac{n^2\varepsilon^2}{2R^2} + in(1-\delta) + O(\varepsilon^3), & l=0, n\in\mathbb{Z}, \\ -2l\delta + in(1-\delta) + O(\varepsilon), & l\in\mathbb{Z}^+, n\in\mathbb{Z}. \end{cases} \quad (19)$$

Here l and n are indices that label the radial and angular modes of the system's dynamics, respectively. We will now talk about the two separate cases of the eigenvalues.

For $l=0$, the focus is on angular dynamics along the limit cycle, with no radial nodes in the eigenfunctions. Here, $n\in\mathbb{Z}$ indexes the angular frequency, representing the azimuthal mode number, e.g., $e^{in\theta}$, where θ is the phase angle. The real part $-\frac{n^2\varepsilon^2}{2R^2}$ of the eigenvalues describes the phase diffusion rate induced by noise (controlled by ε). This negative term, proportional to n^2, indicates that higher angular modes (larger $|n|$) decay faster due to stochastic dynamics. The imaginary part $in(1-\delta)$ corresponds to the angular oscillation frequency around the limit cycle, where $1-\delta$ reflects the deterministic rotational frequency controlled by the damping parameter δ. For $n=0$, there's no oscillation, only slow noise-driven decay; for $n=\pm1$, it's the fundamental frequency, and higher $|n|$ gives faster oscillations, which are multiples (harmonics) of the basic rotation frequency $1-\delta$.

For $l\in\mathbb{Z}^+$, it represents the radial mode number, often associated with the degree of Hermite polynomials in the eigenfunction expression (See C). Each l corresponds to a distinct radial mode with l nodes, describing perturbations away from the limit cycle radius. The real part $-2l\delta$ reflects the decay rate of radial perturbations. Assuming $\delta > 0$ (i.e., stable limit cycle case), this term $-2l\delta$ is negative, and the decay rate increases linearly with l, indicating that higher radial modes

relax more quickly back to the limit cycle due to deterministic damping. The imaginary part $in(1-\delta)$, which is identical to the $l=0$ case, captures the oscillatory behavior tied to angular harmonics.

Experiment Design and Result: In our experiments, we aim to evaluate the accuracy of SDMD and EDMD in estimating eigenvalues of the Koopman generator of the stochastic Stuart-Landau system. The Fourier basis is selected as it aligns well with the periodic nature of the system. We specifically compare eigenvalues corresponding to the $l=0$ case in Eq. (19) from the system with parameter settings: $\delta=0.25$, $\kappa=1$, $\varepsilon=0.05$, $\gamma=1$, and $\beta=1$. The system's state space is discretized over 20 points in both r and θ which forms a uniform grid. The radii are uniformly sampled in the range $r \in [0.4, 0.8]$, and the angles are uniformly sampled from $\theta \in [-\pi, \pi]$. These points serve as initial conditions for the simulations. For the numerical integration, we let the internal integration step size of $h = 1 \times 10^{-5}$ and $n_{\text{steps}} = 10000$ steps; so that the data is collected with $\Delta t = 0.1$. The true eigenvalues of the Koopman generator are computed for the case of $l=0$. For this mode, we focus on eigenvalues corresponding to azimuthal harmonics, represented by $n \in \{-10, -9, \dots, 10\}$ excluding $n=0$, as this eigenvalue corresponds to a trivial mode. These analytical eigenvalues serve as the benchmark to evaluate the accuracy of SDMD and EDMD. In Figure 2, we show comparison of eigenvalues obtained from both methods. More discussion and comparison tests on eigenfunctions will be left in the Appendix C.

B. 1D Ornstein–Uhlenbeck Process

The Ornstein–Uhlenbeck (OU) process is one of the few stochastic processes for which closed-form expressions can be derived for its transition probability density and the spectral properties of its Koopman generator. The one-dimensional OU process is described by the following SDE:

$$dX_t = \theta(\mu_0 - X_t)dt + \sigma dW_t,$$

where $\theta > 0$ is the mean reversion rate, μ_0 is the long-term mean, $\sigma > 0$ is the volatility parameter, and W_t is a standard Wiener process. The generator of this system acting on twice differentiable functions f is[37,58]:

$$(\mathscr{A}f)(x) = \theta(\mu_0 - x)f'(x) + \frac{\sigma^2}{2}f''(x).$$

The eigenvalues and eigenfunctions of this generator have explicit forms, that is, for any non-negative integer n, the n-th eigenpair is:
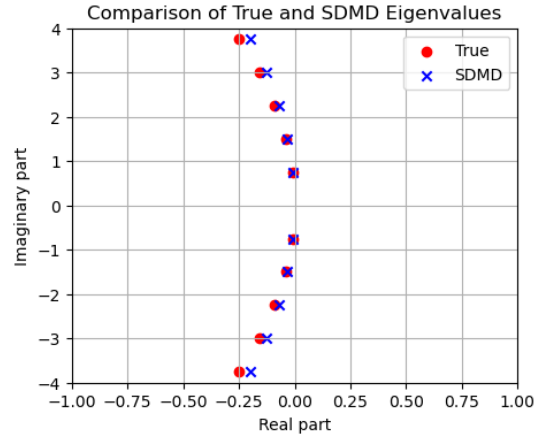
$$\lambda_n = -n\theta, \quad \phi_n(x) = H_n\left(\frac{x - \mu_0}{\sqrt{\sigma^2/(2\theta)}}\right), \quad (20)$$

where $H_n(x)$ is the n-th order Hermite polynomial. The transition density solves the Fokker-Planck equation and has a Gaussian stationary distribution $\rho_\infty(x)$ with mean μ_0 and variance $\sigma^2/(2\theta)$, specifically,

$$\rho_\infty(x) = \left(2\pi\sigma^2/(2\theta)\right)^{-1/2} \exp\left(-(x - \mu_0)^2/(\sigma^2/\theta)\right).$$
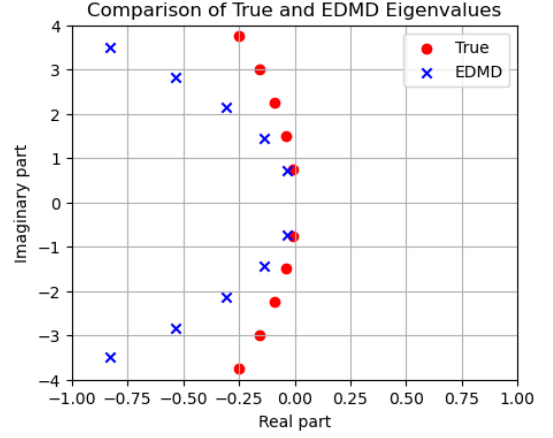
(a) Eigenvalues by EDMD with Fourier basis

(b) Eigenvalues by SDMD with Fourier basis

FIG. 2: Comparison of eigenvalues of Koopman generator estimated by EDMD and SDMD of the stochastic Stuart-Landau equation using Fourier basis. Figure (2a) shows the eigenvalues obtained from EDMD. Figure (2b) shows those obtained from SDMD.

These analytical solutions provide rigorous benchmarks for testing numerical approximation methods like SDMD and gEDMD with neural network settings.

Experiment Design: The parameters are set as $\theta = 1$, $\mu_0 = 0$, and $\sigma = 0.1$. The EM method approximates the solution using discrete time steps $h = 10^{-4}$, with the update rule:

$$X_{t+h} = X_t + \theta(\mu_0 - X_t)h + \sigma\sqrt{h}\,\xi_t,$$

where $\xi_t \sim \mathcal{N}(0,1)$. In our setup, $m = 10$ initial points are chosen uniformly from the domain $[-2, 2]$, and for each initial point, the process is simulated over $n_{\text{eval}} = 200$ evaluations. We select sampling time interval $\Delta = 0.1$. These simulations generate the time series data required for learning the Koopman generator. To approximate the Koopman generator, we apply SDMD and gEDMD methods with a simple NN with dictionary size $N = 20$.

Experiment Result: Figure 3, we show the comparison of

eigenpairs of Koopman generator $\mathscr{A}$ obtained from SDMD-DL, gEDMD-DL and EDMD-DL, respectively. Notice that our SDMD-DL method computes the approximated eigenvalues of the Koopman semigroup $\mathscr{K}^{\Delta t}$, not of the Koopman generator $\mathscr{A}$, so the eigenvalues plotted are by Remark III.5. In Table II, we show the leading four approximated eigenvalues of the Koopman generator, which is first obtained by SDMD-DL and then computed by Eq. (5). The Table II exhibits two different test results and shows that SDMD-DL can successfully approximate the leading four eigenvalues of the generator since they are close to $0, -1, -2, -3$, as discussed in Eq. (20). However, for eigenvalues corresponding to faster-decaying modes, the accuracy also diminishes. This limitation is likely due to insufficient data and an inadequate number of basis functions, constrained by the local computing resource. In Figure 4, we show that the eigenfunctions computed by SDMD-DL exhibit a polynomial structure in correct order, which also highlights the method's consistency with theoretical expectation as in Eq. (5). To provide a clearer, baseline comparison of the core algorithms, we conducted another set of experiments under more idealized conditions. In this case, we used a fixed dictionary of monomial basis functions up to degree 5 and the analytical values for the SDE coefficients. The results are shown in Figure 5. As expected, with a well-suited basis and perfect model knowledge, both SDMD and gEDMD achieve nearly perfect results, accurately capturing the theoretical spectrum. In contrast, standard EDMD still performs less accurate, as it does not correctly account for the stochastic term in its formulation.aware methods (SDMD and gEDMD) for stochastic systems.

Index	Test 1 Eigenvalues	Test 2 Eigenvalues
λ_1	$-5.75892737 \times 10^{-5}$	$-4.93801632 \times 10^{-5}$
λ_2	-1.03753149	-1.15851652
λ_3	-2.23424145	-2.61595495
λ_4	-3.88774292	-3.59187336

TABLE II: OU process: First four eigenvalues of the Koopman generator estimated using SDMD.

C. 2D Triple-Well System

The 2D triple-well potential system[55] provides a rich setting for studying stochastic dynamics due to its intricate energy landscape, which features three distinct basins of attraction separated by potential barriers. This system is inherently metastable, meaning that it tends to remain in one basin for long period, while random fluctuation will also cause slow transitions between basins. Meanwhile, within each basin the system experiences much quicker, small-scale fluctuation. Studying such metastable systems is valuable because it helps us understand long-term behavior, the occurrence of rare events, and the mechanism driving state changes. The dynamics of the system are governed by the following stochastic differential equation

$$d\mathbf{X}_t = -\nabla V(\mathbf{X}_t)dt + \boldsymbol{\sigma} d\mathbf{W}_t,$$

where $\mathbf{X}_t = (x_t, y_t)$ represents the system's state, $V(x,y)$ is the potential function defining the energy landscape, σ is the diffusion coefficient that characterizes the intensity of the stochastic noise, and $\mathbf{W}_t$ is a standard Wiener process.

Experiment design: In this experiment, the potential landscape $V(x,y)$ is defined by

$$V(x,y) := 3e^{-x^2-\left(y-\frac{1}{3}\right)^2} - 3e^{-x^2-\left(y-\frac{5}{3}\right)^2} - 5e^{-(x-1)^2-y^2}$$
$$- 5e^{-(x+1)^2-y^2} + 0.2x^4 + 0.2\left(y-\frac{1}{3}\right)^4,$$

which is depicted in Figure 6. For this system, the noise coefficient matrix σ is defined as a diagonal matrix, where each diagonal element is set to 1.09, that reflects the noise intensity in each spatial direction. The spatial domain is defined by $x \in [-2,2]$ and $y \in [-1,2]$, with $m = 35$ points uniformly sampled along each dimension to create a grid of m^2 initial conditions. The trajectories are generated using the Euler-Maruyama (EM) method for numerical integration, with an integration step size of $h = 1 \times 10^{-3}$ and $n_{\text{steps}} = 100$ steps simulated for each trajectory. In this case, the collected data snapshots have time interval $\Delta t = 0.1$. We apply SDMD and gEDMD methods with dictionary size $N = 10$.

Index	Test 1 Eigenvalues	Test 2 Eigenvalues
λ_1	0.999999135	1.00000083
λ_2	0.993172638	0.99359029
λ_3	0.892089359	0.86772991

TABLE III: Triple-well system: First three eigenvalues of the Koopman generator estimated using SDMD.

Experiment Result: Figure 7 displays the approximated eigenvalues of the Koopman semigroup $\mathscr{K}^{\Delta t}$ computed by SDMD. These eigenvalues quantify the rates of transitions between basins, with smaller eigenvalues corresponding to slower transitions that reflect the metastable behavior of the system. Here's a more detailed analysis for Figure 7 and Figure 8: The eigenvalue $\lambda = 1$ corresponds to the system's steady state. The associated eigenfunction in this case is simply the constant function, which does not provide additional information about the system. It encodes features linked to the long-term behavior of the system. The eigenvalues close to 1 correspond to the slowest timescales in the system, indicating transitions between metastable states. In this triple well potential system, it represents transitions between the two deeper wells, as their similar depths create slow dynamics governed by the energy barrier between them. The slightly smaller eigenvalue reflects faster dynamics, particularly involving transitions with the shallower well. Since the shallower well is less stable, transitions involving this well occur more rapidly, resulting in a larger separation from 1. The smaller eigenvalues capture other transient behaviors that decay quickly. Table III shows the eigenvalues of Koopman generator obtained by Eq. (5). These eigenvalues highlight the system's faster timescales and play a lesser role in describing long-term dynamics. This interpretation aligns with the expected behavior of metastable systems: the spectrum reflects
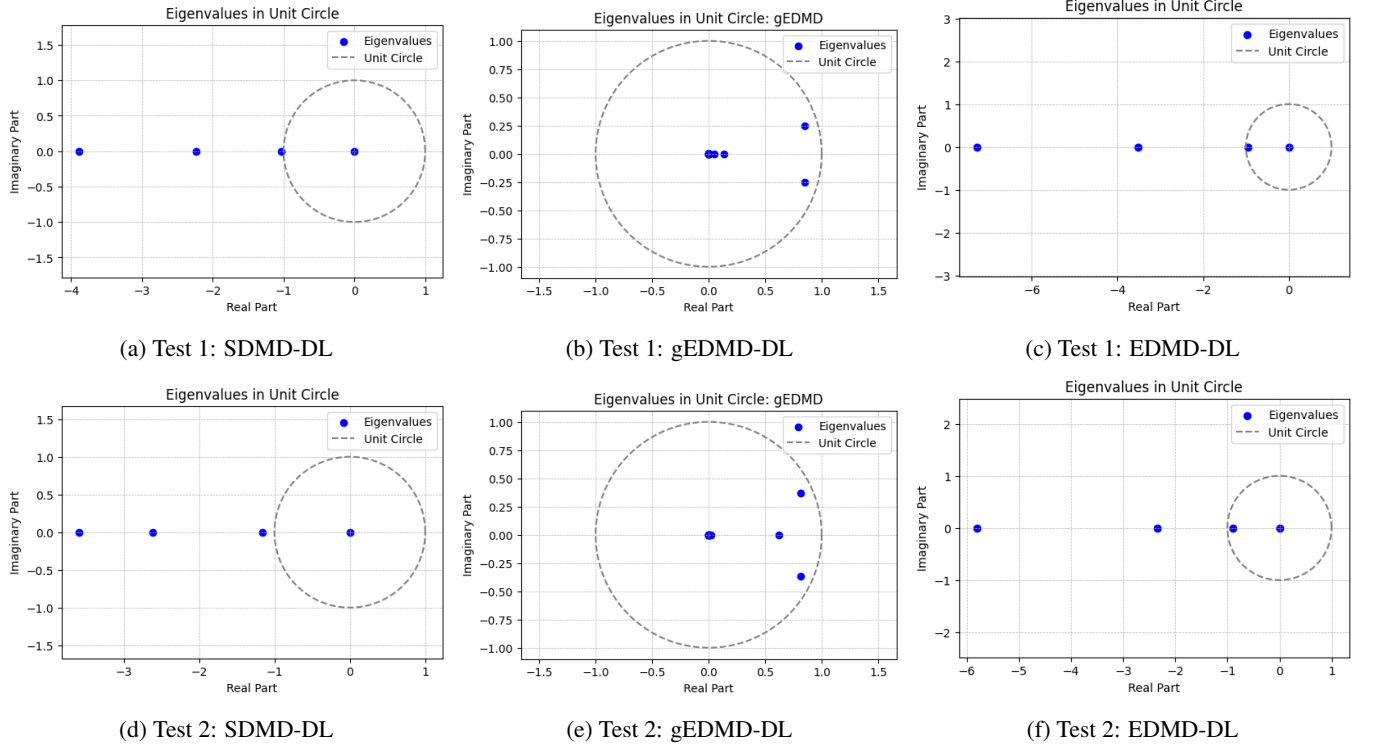
(a) Test 1: SDMD-DL   (b) Test 1: gEDMD-DL   (c) Test 1: EDMD-DL

(d) Test 2: SDMD-DL   (e) Test 2: gEDMD-DL   (f) Test 2: EDMD-DL

FIG. 3: OU process: Two tests on computing eigenvalues of Koopman generator $\mathscr{A}$ obtained from SDMD-DL, gEDMD-DL and EDMD-DL.

D. Recovering Latent Timescales in a Neural Mass Model

Our previous examples (Stuart-Landau oscillator, Ornstein-Uhlenbeck process, triple-well potential system) demonstrated SDMD's accuracy in recovering known spectral structures. However, practical scenarios often involve latent dynamics, such as slow modulations or transient intermediate states that are not directly observable in measured variables. To illustrate SDMD's capacity to efficiently uncover these latent timescales with minimal observables and a compact dictionary, we present here a minimal neural-mass model[56]. This example emphasizes SDMD's effectiveness in extracting hidden multiscale structures from limited measurements, further underscoring the method's general applicability.

Specifically, we consider the following neural-mass model, a minimal two-dimensional reduction of a large population of quadratic integrate-and-fire neurons derived by Montbrio et. al.[56]:

$$dr = \left(\frac{\Delta}{\pi} + 2rv \right) dt + \sigma_r^2 \, dW_r(t),$$

$$dv = \left(v^2 + Jr + I(t) - \pi^2 r^2 \right) dt + \sigma_v^2 \, dW_v(t),$$

where σ_V and σ_r are the standard deviation of addictive Gaussian noise. This system intrinsically possesses two distinct

dynamical regimes: a low-activity fixed-point regime and a damped-oscillatory (spiral) regime (see Figure 9A). The external input $I(t)$ can be set to two different values, each stabilizing one of these regimes (Figure 9B, top). Thus, the input acts as a latent slow-modulating state. Switching between input values induces a transient intermediate timescale associated with decay towards the spiral attractor. This provides a clear scenario to test the capability of SDMD to efficiently recover multiple latent timescales, including the slow modulation, transient decay, and fast oscillation from only the two observables (r, v).

Experiment design: We numerically integrated the above equations using the standard Euler method with a fixed time step ($\Delta t = 0.01$ s). The parameters were set as $\Delta = 1$, $J = 15$, $\sigma_V = \sigma_r = 0.01$ and the input $I(t)$ alternated between two constant values ($I = -2$ and $I = -10$) driven by a Markov switching process. At each time step, the probability of maintaining the same input state was set to 0.9999. Additionally, independent Gaussian noise with zero mean and standard deviation 0.01 was added to both state variables. The total simulation duration was 6000 s (600,000 data points), providing ample data to clearly identify latent slow modulations, intermediate transient dynamics, and fast oscillatory modes using SDMD. Both methods are trained with 300 outer epochs and 2 inner epochs, where the procedures are elaborated in Algorithm 2 and in[49].

Experiment results: To examine the efficacy of SDMD relative to classical dictionary-learning approaches, we compared
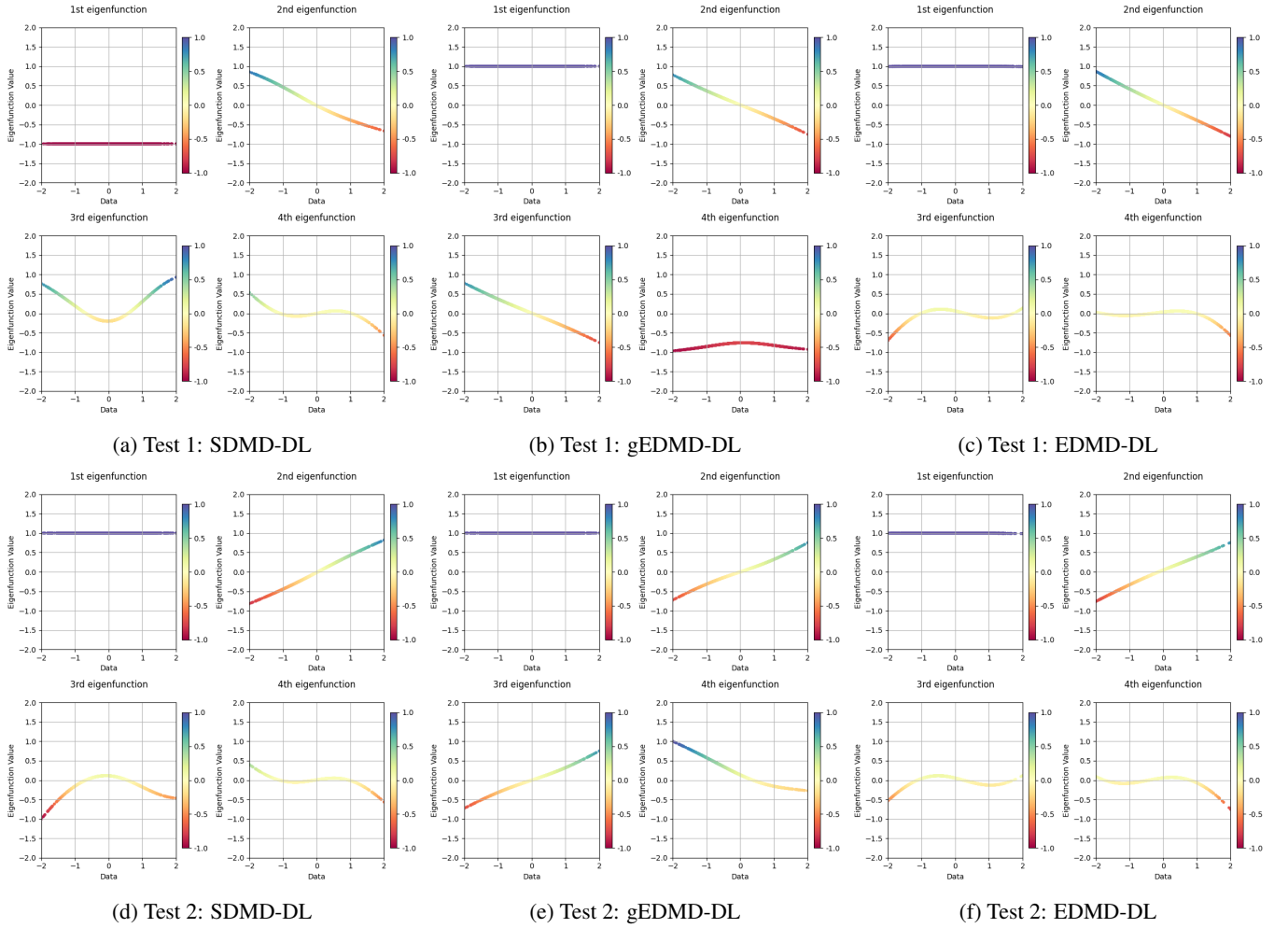
FIG. 4: OU process: two tests on computing eigenfunctions obtained from SDMD-DL, gEDMD-DL and EDMD-DL.

SDMD to Extended Dynamic Mode Decomposition with Dictionary Learning (EDMD-DL). Both methods used identical observable variables (r and v) from the full 6000 s simulated dataset. Crucially, no direct information about the latent input $I(t)$ was provided during the training process. To ensure a fair and consistent comparison, both SDMD and EDMD-DL employed dictionaries constructed via an identical three-layer neural network (50 neurons per hidden layer). The final dictionary consisted of 25 neural-network-trained basis functions, supplemented by 2 first-order polynomial observables (r, v) and a constant term, yielding a total of 28 dictionary elements. This setup allowed for an objective evaluation of SDMD's capability in recovering latent dynamical timescales and hidden states relative to standard EDMD-DL approaches. It is worth mentioning that while EDMD-DL converges after 300 epochs, SDMD-DL does not converge. However, we measured the similarity between the normalized second Koopman eigenfunction (see following) and the ground truth latent input $I(t)$ with the Pearson correlation coefficient and picked the training result from the outer epoch with the highest similarity.

Figure 9B(middle, bottom) shows the first two dominant eigenfunctions identified by SDMD-DL and EDMD-DL from the neural-mass model data. SDMD successfully recovers a latent slow state (represented by eigenfunction ϕ_2) that aligns well with the ground-truth partitions induced by the input modulation (Fig. 9B, top). Moreover, SDMD reveals another eigenfunction (ϕ_3) clearly illustrating the transient dynamics during regime transitions. In contrast, EDMD-DL fails to capture these two latent states despite using the identical neural-network dictionary structure. Specifically, EDMD-DL's eigenfunction ϕ_2 does exhibit baseline shifts between different slow states, but it remains significantly corrupted by high-frequency oscillations from the spiral regime, failing to clearly isolate the slow modulation. Additionally, EDMD-DL's eigenfunction ϕ_3 does not provide interpretable transient dynamics.

Although we selected the SDMD epoch solely by maximizing the similarity between the normalized slow eigenfunction $\phi_2(t)$ and the ground-truth input $I(t)$, this choice also quite remarkably yields an accurate recovery of the transient switching dynamics in the third eigenfunction $\phi_3(t)$. In the selected epoch, $\phi_3(t)$ cleanly isolates the brief regime transitions, whereas in epochs with lower $\phi_2(t) - I(t)$ correla-
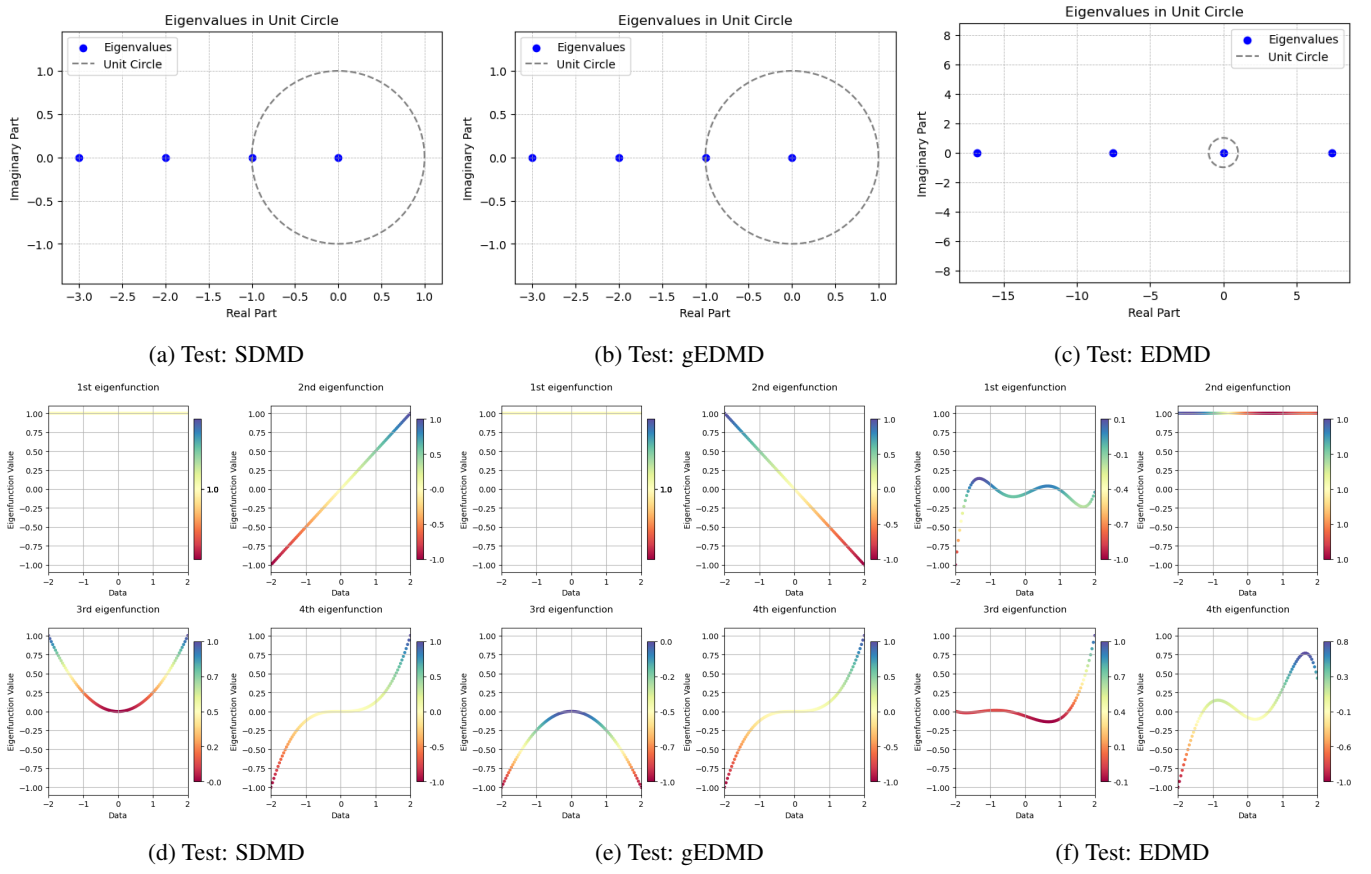
FIG. 5: OU process: test on computing eigenvalues and eigenfunctions of Koopman generator $\mathscr{A}$ obtained from SDMD, gEDMD and EDMD with monomial dictionary.
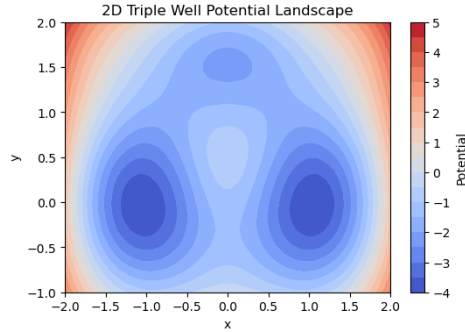

FIG. 6: 2D Triple Well Potential Landscape

tion it remains buried under high-frequency oscillations. This demonstrates that calibrating the Koopman operator on a single, dominant timescale can effectively tune the full spectral approximation, allowing both the slow modulations and the intermediate transients to emerge clearly. In summary, this neural mass model example effectively demonstrates SDMD-DL's advantage over EDMD-DL in extracting underlying multiscale dynamics from a stochastic system. Under identical conditions, i.e., limited observables, parametrized dictionaries, and neural-network architectures, SDMD accurately

isolates latent slow and transient states, despite the presence of noise. The inherent robustness of SDMD to noise is highlighted by its capability to extract meaningful eigenfunctions even from noisy, partially observed data, whereas EDMD-DL fails to decouple latent structures from the stochastic fluctuation. These results illustrate SDMD's suitability and superior performance for analyzing stochastic dynamical systems with hidden multiscale features. Moreover, our findings show that by selecting the model that best aligns the dominant slow eigenmode with the true latent input, one simultaneously cali-
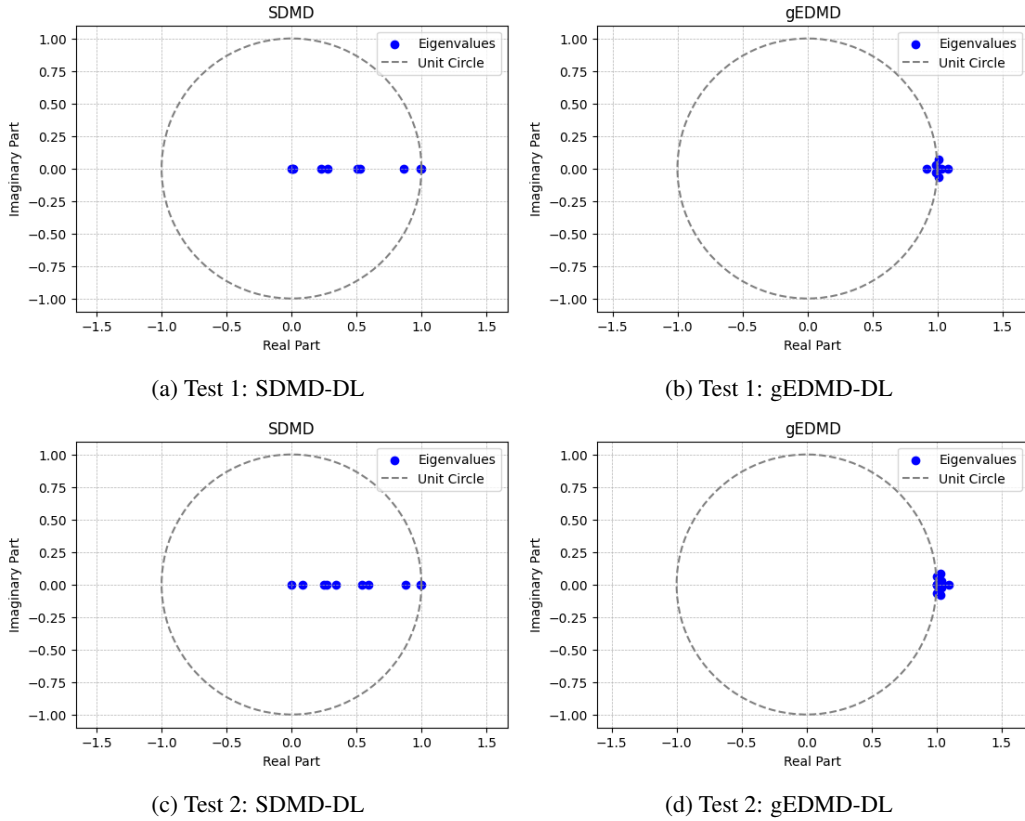
(a) Test 1: SDMD-DL

(b) Test 1: gEDMD-DL

(c) Test 2: SDMD-DL

(d) Test 2: gEDMD-DL

FIG. 7: Triple-well: Two tests on computing eigenvalues of Koopman operator $\mathscr{K}^{\Delta t}$ obtained from SDMD-DL and gEDMD-DL.

brates the full Koopman operator, which enables clean recovery of both slow modulations and rapid switching dynamics.

VII. CONCLUSION

In this paper, we presented a novel computational framework (SDMD) for estimating Koopman semigroups in stochastic dynamical systems. Our approach addresses several challenges in the field and provides a robust foundation for analyzing spectral properties in stochastic dynamics. By directly approximating the semigroup, it eliminates the need for expensive matrix exponential computations, significantly improving computational efficiency. These features, combined with a specially designed loss function and updating rule, make the framework particularly suitable for neural network implementations. Rigorous convergence analysis further supports the method's reliability, which provides probabilistic error bounds and finite-dimensional approximations. Numerical experiments on the Stuart-Landau oscillator, OU process, triple-well potential, and neural mass model confirm that the framework not only estimates accurate Koopman eigenvalues and eigenfunctions but also recovers hidden slow-fast dynamics from limited noisy observations.

Future work will focus on three primary directions. First, we plan to extend the SDMD framework to high-dimensional systems. This extension will tackle both theoretical and computational challenges, including investigating how dimensionality influences convergence rates. Second, we aim to apply SDMD to real-world data to study multiscale phenomena, thereby bridging the gap between theory and practice in fields such as brain dynamics. Third, we intend to incorporate the resolvent operator into our analysis to derive explicit error bounds for the estimated Koopman semigroup, Koopman generator, and their spectral components. In addition, calibrating the approximation to emphasize a specific timescale by explicitly optimizing for a chosen eigenmode or weighting particular spectral components could offer a powerful new strategy for improving Koopman operator estimation and targeting distinct dynamical regimes. By analyzing the spectral characteristics of the resolvent operator, we hope to establish the theoretical framework of stability analysis for our SDMD method in stochastic dynamical systems.

(a) Test 1: SDMD-DL

(b) Test 1: gEDMD-DL

(c) Test 2: SDMD-DL
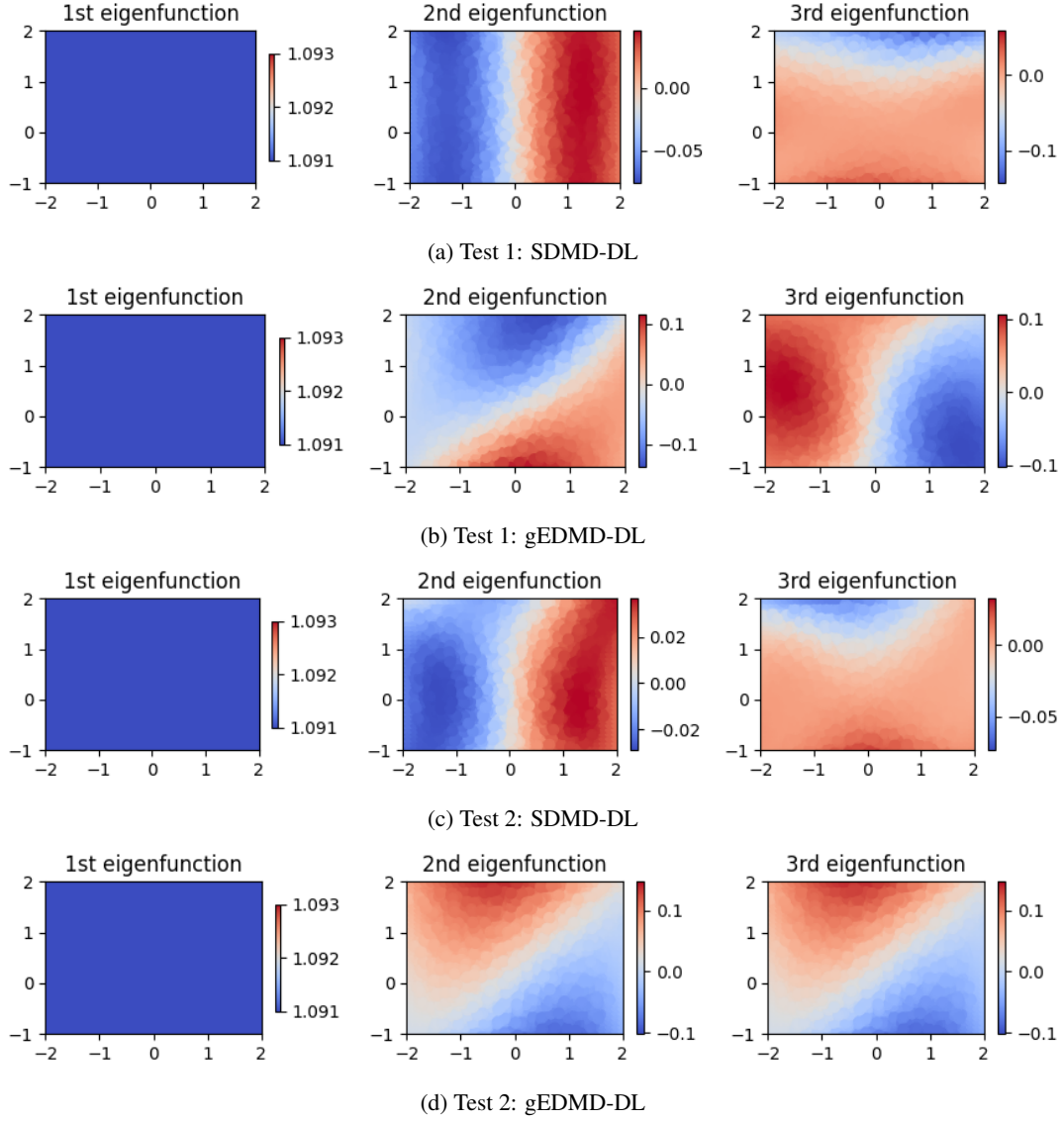
(d) Test 2: gEDMD-DL

FIG. 8: Triple-well: two tests on computing eigenfunctions obtained from SDMD-DL and gEDMD-DL.

results and for reviewing the technical part during the early stages of this work.

Appendix A: Second-order Stochastic Taylor Expansion in SDMD

For $f \in C_b^4(\mathcal{M})$, the Koopman semigroup admits the 2nd order stochastic Taylor expansion in the following

$$\mathcal{K}^{\Delta t} f = f + \Delta t \, \mathscr{A} f + \frac{\Delta t^2}{2} \mathscr{A}^2 f + o(\Delta t^2),$$

where the generator $\mathscr{A}$ acts as

$$\mathscr{A} f = \sum_{i=1}^{d} b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} (\sigma \sigma^\top)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$
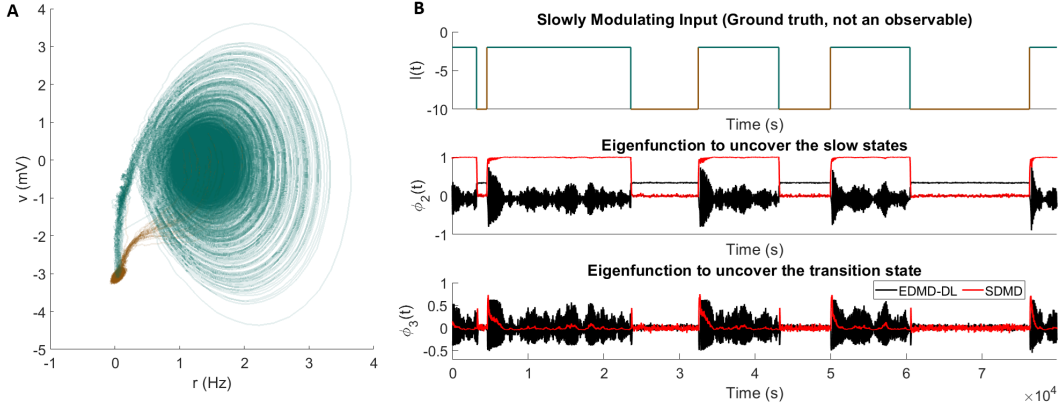
Applying $\mathscr{A}$ again yields

$$\mathscr{A}^2 f = \sum_{k=1}^{d} b_k \frac{\partial}{\partial x_k} (\mathscr{A} f) + \frac{1}{2} \sum_{k,\ell=1}^{d} (\sigma \sigma^\top)_{k\ell} \frac{\partial^2}{\partial x_k \partial x_\ell} (\mathscr{A} f),$$

which expands into a combination of terms involving derivatives of b and σ up to second order, and derivatives of f up to fourth order. Specifically,

$$\mathscr{A}^2 f = \sum_{i,k} b_k \frac{\partial b_i}{\partial x_k} \frac{\partial f}{\partial x_i} + \cdots + \frac{1}{4} \sum_{i,j,k,\ell} (\sigma \sigma^\top)_{k\ell} \frac{\partial^2 (\sigma \sigma^\top)_{ij}}{\partial x_k \partial x_\ell} \frac{\partial^2 f}{\partial x_i \partial x_j} + \cdots,$$

FIG. 9: **SDMD recovers latent states and transient dynamics from a stochastic neural-mass model.** (**A**) Phase-space trajectory of the neural-mass model illustrating two distinct dynamical regimes: a low-activity fixed-point (brown) and a spiral oscillatory state (green), modulated by a slowly alternating latent input (not included as an observable or feature in either method). (**B, top**) Ground truth of the slowly modulating input $I(t)$, switching between two values (-2 and -10), where the color code of the two states matches (A). This latent input state was not directly observable during analysis. (**B, middle and bottom**) Dominant eigenfunctions estimated by SDMD (red) and EDMD-DL (black). SDMD clearly isolates the latent slow input modulation (ϕ_2) and the transient dynamics during regime switches (ϕ_3). By contrast, EDMD-DL fails to effectively recover either latent variable, with its eigenfunctions significantly influenced by high-frequency oscillations and stochastic noise.

where the omitted terms include all mixed derivatives up to order four of f multiplied by appropriate derivatives of b and σ. In the SDMD context, incorporating this term for each basis function ψ_j would modify Eq. (12) to

$$\mathcal{K}^{\Delta t}\psi_j(x) \approx \psi_j(x) + \Delta t\,(\mathcal{A}\psi_j)(x) + \frac{\Delta t^2}{2}(\mathcal{A}^2\psi_j)(x),$$

which leads to a higher-order approximation but with significantly greater computational cost due to the evaluation of $\mathcal{A}^2\psi_j$.

Appendix B: McDiarmid's Inequality

Definition B.1 (Bounded Differences Property). A function $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the *bounded differences property* if substituting the value of the i-th coordinate x_i changes the value of f by at most c_i. More formally, if there are constants $c_1, c_2, \ldots, c_n$ such that for all $i \in [n]$, and all $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \ldots, x_n \in \mathcal{X}_n$,

$$\sup_{x_i' \in \mathcal{X}_i} \Big| f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$$
$$- f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \Big| \le c_i.$$

Lemma B.2 (McDiarmid's Inequality[59]). Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfy the bounded differences property with bounds $c_1, c_2, \ldots, c_n$ as in Definition B.1. Consider independent random variables $X_1, X_2, \ldots, X_n$ where $X_i \in \mathcal{X}_i$ for all i. Then, for any $\varepsilon > 0$,

$$\mathbb{P}\big(|f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)]| \ge \varepsilon\big)$$
$$\le 2\exp\Big(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\Big).$$

Appendix C: 2D Stuart-Landau equation: Phase Diffusion Equation

The SL equation not only has the Cartesian and polar coordinates form, but also has phase coordinates form[53] given in the following:

$$dr = \Big(\delta r - \kappa r^3 + \frac{\varepsilon^2}{2r}\Big)dt + \varepsilon\,dW_r,$$
$$d\phi = \omega_f\,dt + \frac{\varepsilon}{r}\,dW_\theta - \tilde{\beta}\frac{\varepsilon}{r}\,dW_r,$$

where $\phi = \theta - \tilde{\beta}\log(r/R)$ with $\tilde{\beta} = \beta/\kappa$ being the twist factor.

For $\delta > 0$ and $\varepsilon\sqrt{\kappa}/\delta \ll 1$, the eigenfunctions are given by:

- $l = 0$:

$$\psi_{0n} = \exp(i(n(\theta - \tilde{\beta}\log\frac{r}{R}))),$$

- $l > 0$:

$$\psi_{ln} \propto H_l(\sqrt{2\delta}\frac{r-R}{\varepsilon})\exp(i(n(\theta - \tilde{\beta}\log\frac{r}{R}))),$$

where H_l is the l-th order Hermite polynomial.

Figure 10 displays a comparison of the analytical eigenfunctions with those obtained from EDMD and SDMD for various modes. The analytical eigenfunctions are normalized by a factor of $1/\sqrt{2^{|n|}|n|!}$ and combine the phase dynamics, expressed by $\exp\Big(in\Big(\theta - \tilde{\beta}\log\big(\frac{r}{R}\big)\Big)\Big)$, with the radial structure given by Hermite polynomials $H_l\Big(\frac{\sqrt{2\delta}(r-R)}{\varepsilon}\Big)$. Both

EDMD and SDMD utilize Fourier basis functions; however, SDMD demonstrates superior accuracy in capturing the phase structure. Note that our dataset covers the range $r \in [0.4, 0.8]$, so there is no information available for the region $r < 0.4$. Within the considered range, the eigenfunctions computed by both methods exhibit the rotational behavior observed in the true eigenfunctions, as shown in Figure 10.

Appendix D: Estimation of Drift and Diffusion Coefficients in SDE

We start with SDE given in Eq. (1), discretized via the Euler-Maruyama (EM) method over a small time step Δt:

$$\mathbf{X}_{t+\Delta t} - \mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)\Delta t + \sigma(\mathbf{X}_t)\sqrt{\Delta t}\xi_t.$$

Here, $\xi_t \sim \mathcal{N}(0, I)$ introduces noise, and this generates our time series data pairs, $\mathbf{X}_t$ and $\mathbf{X}_{t+\Delta t}$. A single neural network is trained to predict the next state, taking $\mathbf{X}_t$ as input and outputting $\widehat{\mathbf{X}}_{t+\Delta t}$. The loss function driving this training is the mean squared error(SE):

$$\text{MSE} = \frac{1}{N}\sum_t (\mathbf{X}_{t+\Delta t} - \widehat{\mathbf{X}}_{t+\Delta t})^2.$$

This measures the average squared difference between the actual next state $\mathbf{X}_{t+\Delta t}$ from the data and the NN's prediction $\widehat{\mathbf{X}}_{t+\Delta t}$, over N data pairs. By minimizing this loss, the NN learns to approximate the deterministic shift, $\widehat{\mathbf{X}}_{t+\Delta t} \approx \mathbf{X}_t + \mathbf{b}(\mathbf{X}_t)\Delta t$, as the noise term's mean is zero.

From this NN, the drift is estimated as

$$\mathbf{b}(\mathbf{X}_t) \approx \frac{\widehat{\mathbf{X}}_{t+\Delta t} - \mathbf{X}_t}{\Delta t},$$

using $\mathbf{X}_t$ from the data and $\widehat{\mathbf{X}}_{t+\Delta t}$ from the NN. For diffusion, the residual $\mathbf{r}_t = \mathbf{X}_{t+\Delta t} - \widehat{\mathbf{X}}_{t+\Delta t}$ captures the noise, with variance defined as $variance = (\mathbf{X}_{t+\Delta t} - \widehat{\mathbf{X}}_{t+\Delta t})^2 \approx \sigma(\mathbf{X}_t)^2\Delta t$. Thus,

$$\sigma(\mathbf{X}_t) \approx \sqrt{\frac{variance}{\Delta t}}.$$

REFERENCES

[1] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering* (Chapman and Hall/CRC, 2024).

[2] D. Giannakis, G. Froyland, B. Lintner, M. Pike, and J. Slawinska, "Spectral analysis of climate dynamics with operator-theoretic approaches," in *AGU Fall Meeting Abstracts*, Vol. 2021 (2021) pp. A15E–1677.

[3] C. Schütte, S. Klus, and C. Hartmann, "Overcoming the timescale barrier in molecular dynamics: Transfer operators, variational principles and machine learning," Acta Numerica **32**, 517–673 (2023).

[4] J. D. Anderson and J. Wendt, *Computational fluid dynamics*, Vol. 206 (Springer, 1995).

[5] J. Mann and J. N. Kutz, "Dynamic mode decomposition for financial trading strategies," Quantitative Finance **16**, 1643–1655 (2016), https://doi.org/10.1080/14697688.2016.1170194.

(a) Eigenfunctions by EDMD

(b) Eigenfunctions by SDMD
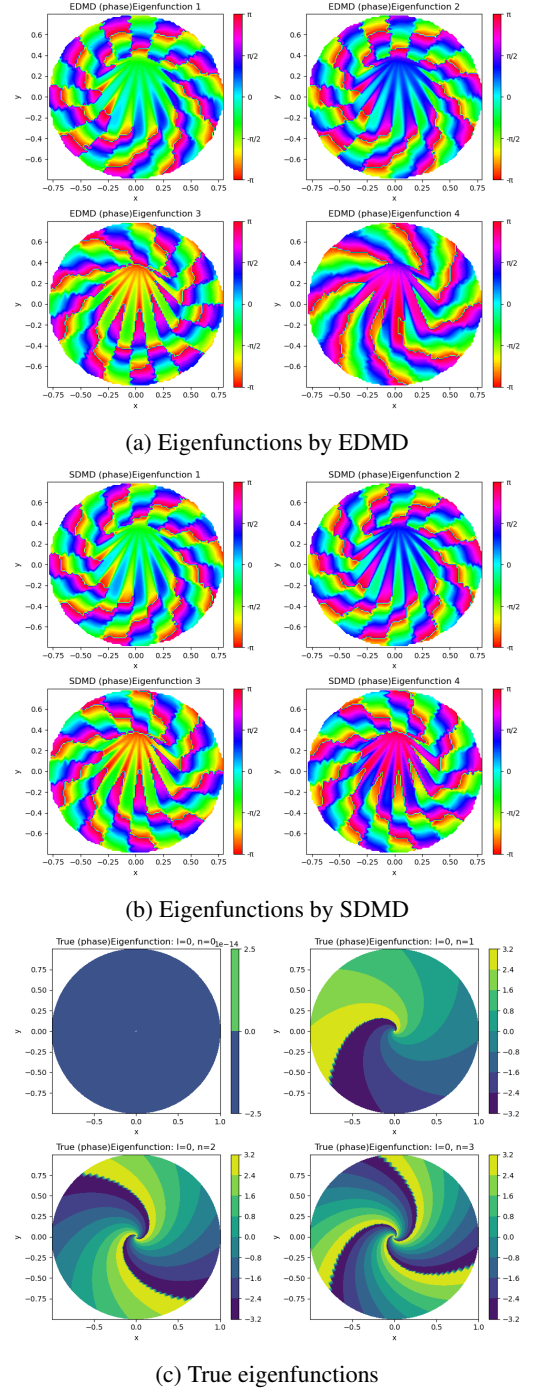

(c) True eigenfunctions

FIG. 10: Comparison of eigenfunctions estimated by EDMD and SDMD for stochastic Stuart-Landau system with Fourier basis. Figure (10a) shows the eigenfunctions obtained from EDMD. Figure (10b) shows those obtained from SDMD.

[6] E. M. Izhikevich, *Dynamical systems in neuroscience* (MIT press, 2007).

[7] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic resonance in climatic change," Tellus **34**, 10–16 (1982).

[8] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: Throwing ropes over rough mountain passes, in the dark," Annual review of physical chemistry **53**, 291–318 (2002).

[9] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system,"

Nature reviews neuroscience **9**, 292–303 (2008).

[10] H. Tuckell, "Introduction to theoretical neurobiology: Volume 2, nonlinear and stochastic theories," (1988).

[11] A. Manwani and C. Koch, "Detecting and estimating signals in noisy cable structures, i: Neuronal noise sources," Neural computation **11**, 1797–1829 (1999).

[12] A. Hawkes, "Stochastic modelling of single ion channels," Computational Neuroscience: a comprehensive approach , 131–158 (2004).

[13] A. Destexhe and D. Paré, "Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo," Journal of neurophysiology **81**, 1531–1547 (1999).

[14] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, "The dynamic brain: from spiking neurons to neural masses and cortical fields," PLoS computational biology **4**, e1000092 (2008).

[15] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proceedings of the national academy of sciences **113**, 3932–3937 (2016).

[16] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," Advances in neural information processing systems **31** (2018).

[17] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," arXiv preprint arXiv:2010.08895 (2020).

[18] L. Lu, P. Jin, and G. E. Karniadakis, "Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators," arXiv preprint arXiv:1910.03193 (2019).

[19] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," Science advances **3**, e1602614 (2017).

[20] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," Proceedings of the National Academy of Sciences **17**, 315 (1931).

[21] B. O. Koopman and J. v. Neumann, "Dynamical systems of continuous spectra," Proceedings of the National Academy of Sciences **18**, 255–263 (1932).

[22] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," Nonlinear Dynamics **41**, 309–325 (2005).

[23] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," Annual review of fluid mechanics **45**, 357–378 (2013).

[24] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," Journal of fluid mechanics **641**, 115–127 (2009).

[25] M. J. Colbrook and A. Townsend, "Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems," Communications on Pure and Applied Mathematics **77**, 221–283 (2024).

[26] I. Ishikawa, Y. Hashimoto, M. Ikeda, and Y. Kawahara, "Koopman operators with intrinsic observables in rigged reproducing kernel Hilbert spaces," arXiv preprint arXiv:2403.02524 (2024).

[27] V. Kostic, P. Novelli, A. Maurer, C. Ciliberto, L. Rosasco, and M. Pontil, "Learning dynamical systems via Koopman operator regression in reproducing kernel Hilbert spaces," (2022), arXiv:2205.14027 [cs.LG].

[28] F. Noé and F. Nuske, "A variational approach to modeling slow processes in stochastic dynamical systems," Multiscale Modeling & Simulation **11**, 635–655 (2013).

[29] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," Journal of Computational Dynamics **1**, 391–421 (2014).

[30] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition," Journal of Nonlinear Science **25**, 1307–1346 (2015).

[31] Y. Xu, K. Shao, N. Logothetis, and Z. Shen, "Reskoopnet: Learning Koopman representations for complex dynamics with spectral residuals," (2025), arXiv:2501.00701 [cs.LG].

[32] M. J. Colbrook, Q. Li, R. V. Raut, and A. Townsend, "Beyond expectations: Residual dynamic mode decomposition and variance for stochastic dynamical systems," ArXiv **abs/2308.10697** (2023).

[33] N. Črnjarić-Žic, S. Maćešić, and I. Mezić, "Koopman operator spectrum for random dynamical systems," Journal of Nonlinear Science **30**, 2007–2056 (2020).

[34] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, and C. Schütte, "Data-driven approximation of the Koopman generator: Model reduction,

[35] M. Wanner and I. Mezic, "Robust approximation of the stochastic Koopman operator," SIAM Journal on Applied Dynamical Systems **21**, 1930–1951 (2022).

[36] A. Lasota and M. C. Mackey, *Chaos, fractals, and noise: stochastic aspects of dynamics*, Vol. 97 (Springer Science & Business Media, 2013).

[37] G. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in Applied Mathematics (Springer New York, 2016).

[38] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences (Springer New York, 2012).

[39] K. Engel, S. Brendle, R. Nagel, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, *et al.*, *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics (Springer New York, 1999).

[40] J. Boyd, *Chebyshev and Fourier Spectral Methods: Second Revised Edition*, Dover Books on Mathematics (Dover Publications, 2013).

[41] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," Journal of Machine Learning Research **18**, 1–43 (2018).

[42] J. J. Bramburger and G. Fantuzzi, "Auxiliary functions as Koopman observables: Data-driven analysis of dynamical systems via polynomial optimization," Journal of Nonlinear Science **34**, 8 (2024).

[43] A. Mauroy, "Koopman operator framework for spectral analysis and identification of infinite-dimensional systems," Mathematics **9** (2021), 10.3390/math9192495.

[44] M. Korda and I. Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," Journal of Nonlinear Science **28**, 687–710 (2017).

[45] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 2018).

[46] F. M. Philipp, M. Schaller, S. Boshoff, S. Peitz, F. Nüske, and K. Worthmann, "Variance representations and convergence rates for data-driven approximations of Koopman operators," (2024), arXiv:2402.02494 [math.DS].

[47] L. Llamazares-Elias, S. Llamazares-Elias, J. Latz, and S. Klus, "Data-driven approximation of Koopman operators and generators: Convergence rates and error bounds," (2024), arXiv:2405.00539 [math.NA].

[48] D. Donoho, "Compressed sensing," IEEE Transactions on Information Theory **52**, 1289–1306 (2006).

[49] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis, "Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator," Chaos: An Interdisciplinary Journal of Nonlinear Science **27** (2017).

[50] Y. Liu, A. Sholokhov, H. Mansour, and S. Nabi, "Physics-informed Koopman network for time-series prediction of dynamical systems," in *ICLR 2024 Workshop on AI4DifferentialEquations In Science* (2024).

[51] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," Nature communications **9**, 4950 (2018).

[52] W. Xiong, M. Ma, X. Huang, Z. Zhang, P. Sun, and Y. Tian, "Koopmanlab: Machine learning for solving complex physics equations," APL Machine Learning **1**, 036110 (2023), https://pubs.aip.org/aip/aml/article-pdf/doi/10.1063/5.0157763/18121014/036110_1_5.0157763.pdf.

[53] A. Tantet, M. D. Chekroun, H. A. Dijkstra, and J. D. Neelin, "Ruelle-pollicott resonances of stochastic systems in reduced state space. part ii: Stochastic hopf bifurcation," (2020), arXiv:1705.07573 [math-ph].

[54] D. T. Gillespie, "Exact numerical simulation of the ornstein-uhlenbeck process and its integral," Physical review E **54**, 2084 (1996).

[55] C. Schütte and M. Sarich, *Metastability and Markov state models in molecular dynamics*, Vol. 24 (American Mathematical Soc., 2013).

[56] E. Montbrió, D. Pazó, and A. Roxin, "Macroscopic description for networks of spiking neurons," Physical Review X **5**, 021028 (2015).

[57] Y. Gu, J. Harlim, S. Liang, and H. Yang, "Stationary density estimation of itô diffusions using deep learning," SIAM Journal on Numerical Analysis **61**, 45–82 (2023).

[58] B. Øksendal, *Stochastic Differential Equations: An Introduction with Ap-

plications, Universitext (Springer Berlin Heidelberg, 2010).

[59]C. Sammut and G. I. Webb, eds., "Mcdiarmid's inequality," in *Encyclopedia of Machine Learning* (Springer US, Boston, MA, 2010) pp. 651–652.