Estimation of Koopman Operator in Stochastic Dynamical System: A Data-Driven Framework with Perturbation Theory

∗1, †2, ‡3, §4, ¶5, and ‖6

1

2

3

4

Abstract

Write your abstract here. Include your research motivation, methods, main results, and conclusions.

***Keywords*—** stochastic Koopman operator, Markov semigroup, data-driven dynamical system, dynamic mode decomposition, perturbation theory

1 Introduction

Dynamical systems provide a fundamental framework for modeling complex phenomena across various scientific disciplines. Of particular importance are stochastic dynamical systems, which arise naturally in diverse applications such as climate science [7], molecular dynamics [27], fluid systems [18], finance [15], etc., where both deterministic dynamics and random perturbations play crucial roles. In recent years, data-driven methods for analyzing these systems have gained significant attention due to their ability to extract meaningful insights directly from observations without requiring detailed prior knowledge of the underlying mechanisms.

Among various data-driven approaches, operator-theoretic methods have emerged as particularly powerful tools. The Koopman operator theory [2, 17, 25], which lifts nonlinear dynamics into a linear (but infinite-dimensional) setting through its action on observable functions, has proven especially valuable. For stochastic systems, the Koopman operator forms a Markov semigroup defined through conditional expectations, capturing the probabilistic evolution of observables over time.

Recent developments in data-driven methods [4, 9, 11, 20, 29, 32], particularly Extended Dynamic Mode Decomposition (EDMD) [32], 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 [3, 10, 11, 14, 16, 31] have been proposed over time. Here, we propose a novel computational framework that combines perturbation theory with

∗@

†@

‡@

§@

¶@

‖@

data-driven approximation techniques. Our framework introduces several key methodological and theoretical advances. At its core, we develop a perturbation-based approach called Stochastic Dynamic Mode Decomposition (SDMD) that explicitly incorporates sampling time information (Δt) in the approximation process. This represents a crucial advancement over existing methods, as the careful consideration of Δt's role provides essential stability guarantees for numerical computations, particularly when dealing with the typically unbounded Koopman generator, where direct Galerkin approaches often struggle.

Beyond the theoretical foundations, our framework offers substantial practical advantages in computational efficiency. Rather than following the conventional path of approximating the generator and then computing expensive matrix exponentials to recover the semigroup, we develop a direct approximation scheme for the semigroup itself. This approach not only eliminates the computational burden of matrix exponential calculations but also introduces a specially designed loss function that significantly reduces computational resource requirements, particularly beneficial for neural network implementations.

While achieving both stability and computational benefits, our framework maintains strong theoretical guarantees. We establish comprehensive convergence results that include both probabilistic error bounds and finite-dimensional operator approximations. These theoretical foundations provide rigorous justification for the method's reliability and practical applicability.

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.

2 Stochastic Koopman Operator

In dynamical systems, the Koopman operator provides a powerful mathematical framework for analyzing the evolution of observable functions instead of the system states themselves. For stochastic systems, the Koopman operator forms a Markov semigroup [12, 22, 23] 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 $M\mathscr{M}\mathscr{X}$ be a Polish 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:

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)d\boldsymbol{W}_t, \quad \mathbf{X}_0 = \mathbf{x}, \tag{2.1}$$

where $\mathbf{b} : M \to \mathbb{R}^d$ is the drift term, $\sigma : M \to \mathbb{R}^{d\times m}$ is the diffusion term, and $(\mathbf{W}_t)_{t\geq 0}$ is an m-dimensional Wiener process. We assume that both $\mathbf{b}$ and σ satisfy appropriate regularity conditions, i.e., local Lipschitz continuity and sublinear growth rate.

Let ρ be a probability measure on M. The space of square-integrable functions with respect to ρ is defined as:

$$L^2_\rho(M) := \left\{ f : \int_M |f|^2 d\rho < \infty \right\},$$

equipped with the inner product

$$\langle f, g\rangle_\rho := \int_M fg\, d\rho,$$

and the corresponding norm

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

Notice that ρ is not necessarily an invariant measure of the underlying dynamical system. For any observable function $f \in \mathscr{F}$ with $\mathscr{F} := L^2_\rho(M)$, the stochastic Koopman operator family $(\mathscr{K}^t)_{t\geq 0}$ is defined as

$$(\mathscr{K}^t f)(\mathbf{x}) := \mathbb{E}_\mathbb{P}[f(\mathbf{X}_t)|\mathbf{X}_0 = \mathbf{x}], \tag{2.2}$$

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

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

$$\mathscr{A} f \coloneqq \lim_{t \to 0} \frac{\mathscr{K}^t f - f}{t} \tag{2.3}$$

on a suitable domain $\mathscr{D}(\mathscr{A}) \subset \mathscr{F}$. From Itô's formula, for $f \in C_b^2(M)$, we have:

$$\mathscr{A} f = \sum_{i=1}^{d} b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} (\sigma \sigma^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{2.4}$$

Remark 2.1. Eq. (2.4) is also called Kolmogorov backward equation. More details can be found in Appendix A.1.

Assumption 2.2. The Koopman semigroup $\{\mathscr{K}^t\}_{t \geq 0}$ is strongly continuous on $\mathscr{F}$.

This assumption ensures that the generator $\mathscr{A}$ is closed on its domain $\mathscr{D}(\mathscr{A}) \subset \mathscr{F}$, making it possible to approximate $\mathscr{A}$ using finite-dimensional operators. We will rely on Assumption 2.2 throughout this paper, as discussed in Section 4.3.1.

The relationship between the stochastic Koopman semigroup and its generator is characterized by: $\mathscr{K}^t = e^{t\mathscr{A}}$ [5]. For spectral analysis of the Koopman generator, we consider the eigenvalue problem:

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

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

$$\mu = e^{t\lambda}, \tag{2.5}$$

where μ is the eigenvalue of $\mathscr{K}^t$. This relationship stems from the semigroup property $\mathscr{K}^t = e^{t\mathscr{A}}$ and provides a practical way to compute the generator's spectrum from discrete-time observations.

3 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 [22, section 5.2] within the framework of the EDMD method [1], which results in an approach tailored for stochastic systems. This approach, referred as Stochastic Dynamic Mode Decomposition (SDMD), provides a data-driven framework for approximating the Koopman semigroup of stochastic systems. Below, we introduce some relavent background and the necessary notation:

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

$$[G]_{ij} \coloneqq \langle \psi_i, \psi_j \rangle_\rho, \quad [H]_{ij} \coloneqq \langle \psi_i, \mathscr{A} \psi_j \rangle_\rho. \tag{3.1}$$

where ρ is the data distribution.

In practice, let $\{\mathbf{x}_k\}_{k=1}^m$ be the i.i.d. data sampled from ρ. Construct the data matrices $\Psi_X, \mathscr{L}^{\Psi_X} \in \mathbb{R}^{m \times N}$ in the following:

$$\Psi_X \coloneqq \begin{bmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_m) & \cdots & \psi_N(\mathbf{x}_m) \end{bmatrix}, \; \mathscr{L}^{\Psi_X} \coloneqq \begin{bmatrix} \mathscr{A}\psi_1(\mathbf{x}_1) & \cdots & \mathscr{A}\psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \mathscr{A}\psi_1(\mathbf{x}_m) & \cdots & \mathscr{A}\psi_N(\mathbf{x}_m) \end{bmatrix}. \tag{3.2}$$

Remark 3.1. 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, we will be training these basis functions using Automatic Differentiation, so we will also need a time series data, not just only at each i.i.d. data points. More details will be discussed in Section 5.

The Gram matrices G and H can be approximated empirically from these data matrices. Specifically, we construct

$$\widehat{G} = \frac{1}{m}\Psi_X^\top \Psi_X, \quad \widehat{H} = \frac{1}{m}\Psi_X^\top \mathscr{L}^{\Psi_X} \tag{3.3}$$

These empirical approximations converge to their theoretical counterparts as the amount of data increases as discussed in [10]. Specifically, by the Strong Law of Large Numbers (SLLN), we have

$$\lim_{m\to\infty}[\widehat{G}]_{ij} = [G]_{ij}, \quad \lim_{m\to\infty}[\widehat{H}]_{ij} = [H]_{ij} \quad \text{a.e.}$$

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

3.1 Stochastic Dynamic Mode Decomposition (SDMD)

The SDMD method provides a framework for approximating the stochastic Koopman operator by incorporating the stochastic Taylor expansion into the EDMD framework. This method offers several key **advantages**: (1) explicitly accounts for stochastic effects through perturbation theory, which provides better stability in numerical computations; (2) directly approximates the semigroup rather than just the generator, which avoids expensive matrix exponential calculations; and (3) maintains theoretical convergence guarantees while being computationally efficient. The main result of this method is the following matrix approximation:

$$\widehat{K}_{N,\Delta t,m} \coloneqq I + \Delta t \cdot \widehat{G}^{-1}\widehat{H}, \tag{3.4}$$

where $\widehat{G}$ and $\widehat{H}$ are the Gram matrices computed from data. 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 Approximation

Consider the stochastic system defined in Eq. (2.1). Let $\mathbf{\Psi}_N(\mathbf{x}_i) = [\psi_1(\mathbf{x}_i)\ \dots\ \psi_N(\mathbf{x}_i)]^\top$ be a vector of manually selected basis functions evaluated at some data point $\mathbf{x}_i$. Suppose $f(\mathbf{x}_i) = \mathbf{\Psi}_N(\mathbf{x}_i)^\top \boldsymbol{a}$ for some $\boldsymbol{a} \in \mathbb{R}^N$, then for any $\Delta t > 0$, EDMD (A.2) approximates the stochastic Koopman operator $\mathscr{K}^{\Delta t}$ onto $\mathscr{F}_N$ using Galerkin approximation:

$$\begin{aligned}\mathscr{K}^{\Delta t} f(\mathbf{x}_i) &= \mathbb{E}_{\mathbb{P}}[f \circ \Phi^{\Delta t}(\mathbf{x}_i)] \\ &= \mathbf{\Psi}_N(\mathbf{x}_i)^\top \widehat{K}_{N,\Delta t,m}\boldsymbol{a} + r(\mathbf{x}_i)\end{aligned} \tag{3.5}$$

where $\Phi^{\Delta t}(\mathbf{x}_i)$ is the stochastic flow from $\mathbf{x}_i$ after time Δt and $r(\mathbf{x}_i)$ 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 [22] in Eq. (2.3) to account for noise. Specifically, for each basis function ψ_j at $\mathbf{x}_i$,

$$\mathscr{A}\psi_j(\mathbf{x}_i) \approx \frac{\mathscr{K}^{\Delta t}\psi_j(\mathbf{x}_i) - \psi_j(\mathbf{x}_i)}{\Delta t} = \frac{\mathbb{E}_{\mathbb{P}}[\psi_j(\Phi^{\Delta t}(\mathbf{x}_i))] - \psi_j(\mathbf{x}_i)}{\Delta t}$$

$$\mathscr{K}^{\Delta t}\psi_j(\mathbf{x}_i) = \mathbb{E}_{\mathbb{P}}[\psi_j(\Phi^{\Delta t}(\mathbf{x}_i))] \approx \psi_j(\mathbf{x}_i) + \Delta t \cdot \mathscr{A}\psi_j(\mathbf{x}_i) + o_{i,j}(\Delta t) \tag{3.6}$$

where each $\mathscr{A}\psi_j(\mathbf{x}_i)$ is computed by Eq. (2.4) 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 $\mathbf{x}_i$.

Remark 3.2. 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. (3.4) when Δt is small.

Next, we use Eq. (3.6) for each basis function ψ_j in the expansion of the stochastic Koopman operator to approximate the expected value of f under the stochastic flow map $\Phi^{\Delta t}$, which gives us:

$$\begin{aligned}
\mathscr{K}^{\Delta t} f(\mathbf{x}_i) &= \mathbb{E}_{\mathbb{P}}[f \circ \Phi^{\Delta t}(\mathbf{x}_i)] \\
&= \mathbb{E}_{\mathbb{P}}[\boldsymbol{\Psi}_N(\Phi^{\Delta t}(\mathbf{x}_i))^\top]\boldsymbol{a} \\
&= \sum_{j=1}^{n} \mathbb{E}_{\mathbb{P}}[\psi_j(\Phi^{\Delta t}(\mathbf{x}_i))]\boldsymbol{a}_j \\
&= \sum_{j=1}^{n} [\psi_j(\mathbf{x}_i) + \Delta t \cdot \mathscr{A}\psi_j(\mathbf{x}_i) + o_{i,j}(\Delta t)]\,\boldsymbol{a}_j \\
&= \left[\boldsymbol{\Psi}_N(\mathbf{x}_i)^\top + \Delta t \cdot \mathscr{L}_N^{\boldsymbol{\Psi}}(\mathbf{x}_i)^\top + \boldsymbol{o}_{i,N}(\Delta t)^\top\right]\boldsymbol{a}
\end{aligned} \tag{3.7}$$

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

Now, equating both Eq. (3.5) and Eq. (3.7), then evaluating over all data points, we can have the following minimization problem:

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

which is equivalent to:

$$\min_{\tilde{K}_{N,\Delta t,m} \in \mathbb{R}^{N\times N}} \|\Psi_X + \Delta t \cdot \mathscr{L}^{\Psi_X} + \boldsymbol{o}_{m,N}(\Delta t) - \Psi_X \tilde{K}_{N,\Delta t,m}\|_F^2 \tag{3.8}$$

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

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

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

$$\begin{aligned}
\tilde{K}_{N,\Delta t,m} &= \Psi_X^\dagger(\Psi_X + \Delta t \cdot \mathscr{L}^{\Psi_X} + \boldsymbol{o}_{m,N}(\Delta t)) \\
&= I + \Delta t \cdot \left(\Psi_X^\top \Psi_X\right)^{-1} \left(\Psi_X^\top \mathscr{L}^{\Psi_X}\right) + \Psi_X^\dagger \boldsymbol{o}_{m,N}(\Delta t) \\
&= I + \Delta t \cdot \widehat{G}^{-1}\widehat{H} + \Psi_X^\dagger \boldsymbol{o}_{m,N}(\Delta t)
\end{aligned} \tag{3.9}$$

where $\dagger$ denotes the pseudoinverse. If Δt is very small, we can update the Koopman matrix approximation $\widehat{K}_{N,\Delta t,m}$ with $\boldsymbol{o}_{m,N}(\Delta t)$ omitted as in the Eq. (3.4):

$$\tilde{K}_{N,\Delta t,m} \approx \widehat{K}_{N,\Delta t,m} \coloneqq \Psi_X^\dagger(\Psi_X + \Delta t \cdot \mathscr{L}^{\Psi_X}) = I + \Delta t \cdot \widehat{G}^{-1}\widehat{H}.$$

Remark 3.3. We typically add some small number γ for $\widehat{G}$ to avoid singularity, i.e., compute $(\widehat{G} + \gamma I)^{-1}$ instead of $\widehat{G}^{-1}$.

3.2 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 3.1, 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. 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 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.

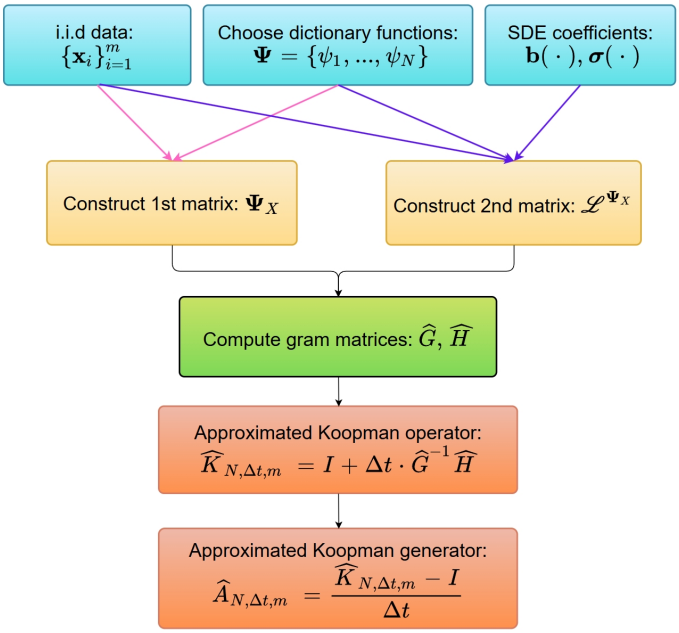


Figure 1: A flow chart for SDMD method.

Algorithm 1 Estimation of stochastic Koopman operator

Input: 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 $\Psi_X, \mathscr{L}^{\Psi_X}$ (See Eq. (3.2)).
2: Construct empirical gram matrices $\widehat{G} = (1/m)\Psi_X^\top \Psi_X, \widehat{H} = (1/m)\Psi_X^\top \mathscr{L}^{\Psi_X}$ (See Eq. (3.3)) .
3: Compute $\widehat{K}_{N,\Delta t,m} = I + \Delta t \cdot \widehat{G}^{-1}\widehat{H}$ (See Eq. (3.4)).
4: Compute $\widehat{A}_{N,\Delta t,m} = \frac{\widehat{K}_{N,\Delta t,m} - I}{\Delta t}$.

Output: Approximated Koopman operator $\widehat{K}_{N,\Delta t,m}$ and approximated Koopman generator $\widehat{A}_{N,\Delta t,m}$.

4 Convergence Analysis

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

Large Data Convergence (First Limit): Establishes that empirical approximations converge to theoretical values as sample size increases, supported by probabilistic error bounds. This step ensures the reliability of approximations before considering other limits.

Zero-Limit of Sampling Time (Second Limit): Proves that time-discretized approximations converge to the true Koopman generator as the sampling interval decreases. This limit builds on large data convergence and connects discrete-time computations to continuous dynamics.

Large Dictionary Size Convergence (Final Limit): Demonstrates that finite-dimensional approximations of the Koopman generator and semigroups converge to their infinite-dimensional counterparts. This step completes the convergence analysis, relying on the stability established in the previous limits.

This structured approach highlights the critical interplay between these limits and ensures rigorous convergence results through appropriate function spaces and operator topologies.

4.1 Convergence in the Limit of Large Data

Given the dictionary size N and sampling time step Δt. Let $K_{N,\Delta t}$ denote the matrix that represents the large data limit of $\tilde{K}_{N,\Delta t,m}$ given in Eq. (3.9):

$$\begin{aligned} K_{N,\Delta t} &= \lim_{m\to\infty} \Psi_X^{\dagger}\left(\Psi_X + \Delta t\cdot \mathscr{L}^{\Psi_X} + \boldsymbol{o}_{m,N}(\Delta t)\right) \\ &= \lim_{m\to\infty}\left[I + \Delta t\cdot\left(\Psi_X^{\top}\Psi_X\right)^{-1}\left(\Psi_X^{\top}\mathscr{L}^{\Psi_X}\right)\right] + \lim_{m\to\infty}\Psi_X^{\dagger}\boldsymbol{o}_{m,N}(\Delta t) \\ &= I + \Delta t\cdot G^{-1}H + \boldsymbol{o}_N(\Delta t) \end{aligned} \tag{4.1}$$

where $\boldsymbol{o}_N(\Delta t) = \lim_{m\to\infty}\Psi_X^{\dagger}\boldsymbol{o}_{m,N}(\Delta t) = G^{-1}\lim_{m\to\infty}\Psi_X^{\top}\boldsymbol{o}_{m,N}(\Delta t) \in \mathbb{R}^{N\times N}$.

Remark 4.1. Each element in the matrix $\Psi_X^{\top}\boldsymbol{o}_{m,N}(\Delta t)$ is $o(\Delta t)$ and the matrix shape will be kept $N\times N$ as $m\to\infty$.

Recall that $\widehat{K}_{N,\Delta t,m} = I + \Delta t\cdot\widehat{G}^{-1}\widehat{H}$ from Eq. (3.4). To establish the convergence of the empirical Koopman matrix $\widehat{K}_{N,\Delta t,m}$ to the limiting matrix $K_{N,\Delta t}$ in large data, 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 4.2. Let $\varepsilon > 0$. Define $\tilde{\varepsilon} = (\varepsilon - \boldsymbol{o}_N(\Delta t))/\Delta t$. Then, with same notation and conditions defined in Lemma 4.5 and Lemma 4.6, we have

$$\mathbb{P}\left(\|\widehat{K}_{N,\Delta t,m} - K_{N,\Delta t}\|_F > \varepsilon\right) \le 2N^2\left[\exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}}{NC^2\tau\|G^{-1}\|}\right)^2\right) + \exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}\|H\|}{NC^2\tau L}\right)^2\right)\right].$$

For deriving such a probabilistic bound, we will employ *Hoeffding's Inequality* [30], which is a powerful concentration inequality for functions of independent random variables and is essential for its application.

Lemma 4.3 (Hoeffding's Inequality)**.** Assume $X_i \in [a_i, b_i]$ for all $1 \le i \le n$. Then, for any $\varepsilon > 0$,

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

Proof. Let $f(X_1,\dots,X_n)=\sum_{i=1}^n X_i$ satisfying the *Bounded Differences Property* given in A.3, we can apply the McDiarmid's Inequality [26] and immediately obtain the result. □

Remark 4.4. Refer to A.3 for more details on Bounded Differences Property and McDiarmid's Inequality.

Lemma 4.5. Let $L,C\in\mathbb{R}$ such that $|\psi_i(\mathbf{x}_k)|\le C, |\mathscr{A}\psi_i(\mathbf{x}_k)|\le LC$ for all $1\le i\le N$ and $1\le k\le m$. Then, for any $\varepsilon>0$,

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

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

$$\begin{aligned}\|\widehat{G}-G\|_F^2&=\sum_{i=1}^N\sum_{j=1}^N\left|\widehat{G}_{ij}-G_{ij}\right|^2\\&=\sum_{i=1}^N\sum_{j=1}^N\left|\frac{1}{m}\sum_{k=1}^m\eta_{ij}(\mathbf{x}_k)-\mathbb{E}\left[\eta_{ij}(\mathbf{x}_1)\right]\right|^2\\&=\sum_{i=1}^N\sum_{j=1}^N\left|\frac{1}{m}\sum_{k=1}^m\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2\end{aligned}$$

where $\tilde{\eta}_{ij}(\mathbf{x}_k)\coloneqq\eta_{ij}(\mathbf{x}_k)-\mathbb{E}\left[\eta_{ij}(\mathbf{x}_1)\right]$ 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 Eq. (4.3) and union bound, we have

$$\begin{aligned}\mathbb{P}\left(\left\|\widehat{G}-G\right\|_F\ge\varepsilon\right)&=\mathbb{P}\left(\left\|\widehat{G}-G\right\|_F^2\ge\varepsilon^2\right)\\&\le\mathbb{P}\left(\sum_{i=1}^N\sum_{j=1}^N\left|\frac{1}{m}\sum_{k=1}^m\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2\ge\varepsilon^2\right)\\&\le N^2\mathbb{P}\left(\left|\frac{1}{m}\sum_{k=1}^m\tilde{\eta}_{ij}(\mathbf{x}_k)\right|^2\ge\varepsilon^2/N^2\right)\\&=N^2\mathbb{P}\left(\left|\frac{1}{m}\sum_{k=1}^m\tilde{\eta}_{ij}(\mathbf{x}_k)\right|\ge\varepsilon/N\right)\\&\le 2N^2\exp\left(-\frac{2(m\varepsilon/N)^2}{m(4C^2)^2}\right)\\&=2N^2\exp\left(-\frac{m\varepsilon^2}{8N^2C^4}\right)\end{aligned}$$

Similarly, since $|\psi_i(\mathbf{x}_k)\mathscr{L}\psi_j(\mathbf{x}_k)|\le C^2L$, we have

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

□

Lemma 4.6. Let $G,H\in\mathbb{R}^{N\times N}$ be such that G is invertible and $H\neq 0$. 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 sub-multiplicative matrix norm $\|\cdot\|$ on $\mathbb{R}^{N\times N}$ and any $\varepsilon>0$ we have

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

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

Proof. See [24, Lemma C.5]. □

Proof of Theorem 4.2.

$$\begin{aligned}
\mathbb{P}\left(\|\widehat{K}_{N,\Delta t,m} - K_{N,\Delta t}\|_F > \varepsilon\right) &= \mathbb{P}\left(\|\Delta t \cdot \widehat{G}^{-1}\widehat{H} - \Delta t \cdot G^{-1}H + \boldsymbol{o}_N(\Delta t)\|_F > \varepsilon\right) \\
&\leq \mathbb{P}\left(\Delta t \cdot \|\widehat{G}^{-1}\widehat{H} - G^{-1}H\|_F + \boldsymbol{o}_N(\Delta t) > \varepsilon\right) \\
&= \mathbb{P}\left(\|\widehat{G}^{-1}\widehat{H} - G^{-1}H\|_F > \tilde{\varepsilon}\right) \\
&\leq \mathbb{P}\left(\|G - \widehat{G}\| > \frac{\tilde{\varepsilon}}{\tau}\|G^{-1}\|^{-1}\right) + \mathbb{P}\left(\|H - \widehat{H}\| > \frac{\tilde{\varepsilon}}{\tau}\|H\|\right) \\
&\leq 2N^2 \exp\left(-\frac{m(\frac{\tilde{\varepsilon}}{\tau}\|G^{-1}\|^{-1})^2}{8N^2C^4}\right) + 2N^2 \exp\left(-\frac{m(\frac{\tilde{\varepsilon}}{\tau}\|H\|)^2}{8N^2C^4L^2}\right) \\
&= 2N^2\left[\exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}\|G^{-1}\|^{-1}}{NC^2\tau}\right)^2\right) + \exp\left(-\frac{m}{8}\left(\frac{\tilde{\varepsilon}\|H\|}{NC^2\tau L}\right)^2\right)\right]
\end{aligned}$$

□

4.2 Convergence in the Zero-Limit of Sampling Time

Now we have Koopman matrix approximation $K_{N,\Delta t}$ for some $\Delta t > 0$. We will use it to construct the matrix $A_{N,\Delta t}$ and show that it is the matrix representation of the time-discretized approximant $\mathscr{A}_{N,\Delta t}$ which converges to projected Koopman generator $\mathscr{A}_N$ in strong operator topology as $\Delta t \to 0$. This convergence is shown in the following theorem:

Theorem 4.7. For each dictionary size $N > 0$, construct the following matrix:

$$A_{N,\Delta t} := \frac{K_{N,\Delta t} - I}{\Delta t}.$$

where $K_{N,\Delta t}$ is given in Eq. (4.1). Suppose $A_{N,\Delta t}$ is the matrix representation of some time-discretized operator $\mathscr{A}_{N,\Delta t}$. Let $\mathscr{A}_N = \mathscr{P}_N \mathscr{A}|_{\mathscr{F}_N}$ be the projected Koopman generator on $\mathscr{F}_N$ with $\mathscr{A}$ being the Koopman generator. Then, $\mathscr{A}_{N,\Delta t} \to \mathscr{A}_N$ in strong operator topology as $\Delta t \to 0$.

Lemma 4.8. Show that the matrix representation of the projected Koopman generator $\mathscr{A}_N := \mathscr{P}_N \mathscr{L}|_{\mathscr{F}_N}$ on $\mathscr{F}_N$ is $G^{-1}H = \lim_{m\to\infty} \Psi_X^\dagger \mathscr{L} \Psi_X =: A_N$ where G, H are given in Eq. (3.1).

Proof. We can obtain this result by applying Galerkin approximation for the Koopman generator $\mathscr{A}$. A similar proof is given in [10, Proposition 3.5]. □

Proof of Theorem 4.7. First, given $N > 0$, we have

$$\begin{aligned}
A_{N,\Delta t} &= \frac{K_{N,\Delta t} - I}{\Delta t} \\
&= \frac{I + \Delta t \cdot G^{-1}A + \boldsymbol{o}_N(\Delta t) - I}{\Delta t} \\
&= G^{-1}A + \boldsymbol{o}_N(\Delta t)/\Delta t \\
&= A_N + \boldsymbol{o}_N(\Delta t)/\Delta t
\end{aligned}$$

where $A_N = G^{-1}H$ by Lemma 4.8.

Next, choose $f \in \text{span}\{\psi_i\}_{i=1}^N$ so that $f = \mathbf{\Psi}_N \boldsymbol{a}$ where $\boldsymbol{a} \in \mathbb{R}^N$. Then we have

$$\begin{aligned}
\mathscr{A}_{N,\Delta t} f &= \mathbf{\Psi}_N A_{N,\Delta t} \boldsymbol{a} \\
&= \mathbf{\Psi}_N \left(A_N + \boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a} \\
&= \mathbf{\Psi}_N A_N \boldsymbol{a} + \mathbf{\Psi}_N \left(\boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a} \\
&= \mathscr{A}_N f + \mathbf{\Psi}_N \left(\boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a}
\end{aligned}$$

Notice that, the first term $\mathscr{A}_N f = \mathbf{\Psi}_N A_N \boldsymbol{a}$ is independent of Δt since we choose the basis functions manually. Now, we consider the second term $\mathbf{\Psi}_N \left(\boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a}$:

$$\begin{aligned}
&\left\|\mathbf{\Psi}_N \left(\boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a}\right\| \\
&= \left\| \begin{bmatrix} \psi_1 & \cdots & \psi_N \end{bmatrix} \begin{bmatrix} \frac{o_{11}(\Delta t)}{\Delta t} & \cdots & \frac{o_{1N}(\Delta t)}{\Delta t} \\ \vdots & \ddots & \vdots \\ \frac{o_{N1}(\Delta t)}{\Delta t} & \cdots & \frac{o_{NN}(\Delta t)}{\Delta t} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} \right\| \\
&= \left\| \sum_{i=1}^N \psi_i \left(\sum_{j=1}^N \left[\boldsymbol{o}_N(\Delta t)/\Delta t\right]_{ij} a_j \right) \right\| \\
&\leq \left\| \sum_{i=1}^N \psi_i \left(N \cdot \max_{1 \leq i,j \leq N} \left(\frac{o_{i,j}(\Delta t)}{\Delta t} \right) \sum_{j=1}^N a_j \right) \right\| \\
&= N \cdot \left| \max_{1 \leq i,j \leq N} \left(\frac{o_{i,j}(\Delta t)}{\Delta t} \right) \right| \left\| \sum_{i=1}^N \psi_i \left(\sum_{j=1}^N a_j \right) \right\| \to 0 \text{ as } \Delta t \to 0.
\end{aligned}$$

Therefore, as $\Delta t \to 0$, we have

$$\left\|\mathscr{A}_{N,\Delta t} f - \mathscr{A}_N f\right\| = \left\|\mathbf{\Psi}_N \left(\boldsymbol{o}_N(\Delta t)/\Delta t\right) \boldsymbol{a}\right\| \to 0,$$

In other words,

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

□

4.3 Convergence in the Limit of Large Dictionary Size

This section establishes the theoretical convergence properties of our approximation scheme for stochastic Koopman operator as the dictionary size increases. Our analysis consists of two main convergence results:

Generator Convergence The projected generator $\mathscr{A}_N = \mathscr{P}_N \mathscr{A}|_{\mathscr{F}_N}$ converges to the true generator $\mathscr{A}$ in strong operator topology as $N \to \infty$.

Semigroup Convergence Under suitable boundedness conditions on $(e^{t\mathscr{A}_N})_{t\geq 0}$ and $(e^{t\mathscr{A}})_{t\geq 0}$, the approximated semigroups converge to the true Koopman semigroups uniformly over compact time interval.

Our analysis builds upon the framework in [14, Section 4] and uses the Trotter-Kato Approximation theorem [5]. The proofs require careful consideration of function spaces and several technical assumptions, which we will introduce systematically building on those stated in Section 2. These ensure the necessary topological and continuity properties of the true Koopman generator $\mathscr{A}$ and semigroups $(e^{t\mathscr{A}})_{t\geq 0}$.

4.3.1 Convergence of Projected Koopman Generator

Recall that given the state space $(M, \mathscr{B}, \rho)$ and the set of observables $\mathscr{F} = L^2_\rho(M)$ from previous sections, the domain $\mathscr{D}(\mathscr{A})$ of the Koopman generator $\mathscr{A}$ is defined as:

$$\mathscr{D}(\mathscr{A}) := \{f \in \mathscr{F} : \mathscr{A} f \in \mathscr{F}\}.$$

Given the assumptions in Section 2, we know that the Koopman generator $\mathscr{A}$ is closed, in other words, the graph of $\mathscr{A}$, defined by:

$$\{(f, \mathscr{A}f) \in \mathscr{F} \times \mathscr{F} : f \in \mathscr{D}(\mathscr{A})\},$$

is closed in $\mathscr{F} \times \mathscr{F}$. Therefore, $\mathscr{D}(\mathscr{A})$ becomes a Hilbert space equipped with the inner product:

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

and the corresponding graph norm is defined as:

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

Assumption 4.9. We assume that the dictionary Ψ satisfies the following conditions:

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

where $\mathscr{P}_N$ and $\mathscr{P}_{\mathscr{D}_N}$ are projections of $\mathscr{F}$ and $\mathscr{D}(\mathscr{A})$ onto $\mathscr{F}_N$ using inner products on $\mathscr{F}$ and $\mathscr{D}(\mathscr{A})$ respectively.

With this framework in place, we proceed to specify the crucial approximation properties of the projection operator as follows:

Theorem 4.10. Suppose Assumption 4.9 holds. Then,

$$\lim_{N\to\infty} \mathscr{A}_N = \mathscr{A} \quad \text{S.O.T.}$$

where S.O.T. denotes strong operator topology.

Lemma 4.11. $\mathscr{D}(\mathscr{A})$ is dense in $\mathscr{F}$. Moreover, the generator $\mathscr{A}$ satisfies $\|\mathscr{A}\| \leq 1$.

Proof. The first statement can be proved immediately since $\mathscr{A}$ is the generator of a strongly continuous semigroup $\mathscr{K}^t$ [23, Corollary 2.5]. As for the second statement, for any $f \in \mathscr{D}(\mathscr{A})$:

$$\|\mathscr{A}f\|_{\mathscr{F}}^2 \leq \|f\|_{\mathscr{F}}^2 + \|\mathscr{A}f\|_{\mathscr{F}}^2 = \|f\|_{\mathscr{A}}^2.$$

Therefore $\|\mathscr{A}\| \leq 1$. □

Proof of Theorem 4.10. Recall that $\mathscr{A}_N = \mathscr{P}_N \mathscr{A}|_{\mathscr{F}_N}$. For any $f \in \mathscr{D}(\mathscr{A})$, we can write:

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

by Assumption 4.9 and Lemma 4.11. Therefore, $\mathscr{A}_N$ converges to $\mathscr{A}$ in strong operator topology. □

Remark 4.12. A similar result is given in [14, Theorem 4.2].

4.3.2 Convergence of Projected Koopman Semigroups

Denote the semigroups generated by $\{\mathscr{A}_N\}_{N\geq 1}$ and $\mathscr{A}$ by:

$$\mathscr{K}_N^t = e^{t\mathscr{A}_N}, \quad \mathscr{K}^t = e^{t\mathscr{A}}.$$

To establish the convergence of the finite-dimensional approximations to their infinite-dimensional counterparts, we first introduce the following concept:

Definition 4.13 (Core of the Koopman Generator)**.** A linear subspace $\mathscr{D} \subseteq \mathscr{D}(\mathscr{A})$ is called a core for the Koopman generator $\mathscr{A}$ if $\mathscr{D}$ is dense in $\mathscr{D}(\mathscr{A})$ with respect to the graph norm.

Assumption 4.14. Let $\mathscr{D}$ be a core for $\mathscr{A}$. We assume:

- $\mathscr{D} \subseteq \mathscr{D}(\mathscr{A}_N)$ for all $N \in \mathbb{N}$
- For any $f \in \mathscr{D}$, $\mathscr{A}_N f \to \mathscr{A} f$ in $\mathscr{F}$ as $N \to \infty$

With the assumptions given above, we show that under suitable boundedness conditions, the projected semigroups $\mathscr{K}_N^t$ converge to the true Koopman semigroup $\mathscr{K}^t$ in the strong operator topology uniformly over compact time interval. The main result is presented in the following Theorem 4.15:

Theorem 4.15. There exists some constants $D \geq 1$, $\omega \in \mathbb{R}$, 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 all $f \in \mathscr{D}$ we have:

$$\mathscr{K}_N^t f \to \mathscr{K}^t f$$

strongly, uniformly for t in compact interval.

Lemma 4.16. $(\mathscr{K}_N^t)_{t\geq 0}$ are strongly continuous semigroups.

Proof. For each $N \geq \mathbb{N}$, let

$$\mathscr{K}_N^t = e^{t\mathscr{A}_N} = \sum_{n=0}^{\infty} \frac{(t\mathscr{A}_N)^n}{n!}$$

This series converges because:

$$\|e^{t\mathscr{A}_N}\| \leq \sum_{n=0}^{\infty} \frac{\|t\mathscr{A}_N\|^n}{n!} = \sum_{n=0}^{\infty} \frac{|t|^n \|\mathscr{A}_N\|^n}{n!} = e^{|t|\|\mathscr{A}_N\|}$$

Now we check two properties below:

1. Semigroup property:
$$\mathscr{K}_N^t \mathscr{K}_N^s = e^{t\mathscr{A}_N} e^{s\mathscr{A}_N} = e^{(t+s)\mathscr{A}_N} = \mathscr{K}_N^{t+s}$$
$$\mathscr{K}_N^0 = e^0 = I$$

2. Strong continuity: for any $x \in \mathscr{D}$,
$$\|\mathscr{K}_N^t x - x\| = \|e^{t\mathscr{A}_N} x - x\| = \|\sum_{n=1}^{\infty} \frac{(t\mathscr{A}_N)^n}{n!} x\|$$
$$\leq |t|\|\mathscr{A}_N\| e^{|t|\|\mathscr{A}_N\|} \|x\| \to 0 \quad \text{as } t \to 0.$$

Therefore, $(\mathscr{K}_N^t)_{t\geq 0}$ are indeed a strongly continuous semigroups generated by $\mathscr{A}_N$ for all $N \in \mathbb{N}$. □

Proof of Theorem 4.15. The exponential boundedness in the first part of the theorem is true by Lemma 4.16 and the strong continuity assumption for $\mathscr{K}^t$ in Section 2 [23, Theorem 2.2]. The second part of the theorem is an immediate consequence of First Trotter-Kato Approximation Theorem. More details of the theorem can be referred to A.4. □

5 Neural Network for SDMD (NN-SDMD)

For the case of a trained basis parameterized by a neural network, we replace $\Psi_X + \Delta t \cdot \mathscr{L}^{\Psi_X} + \boldsymbol{o}_{m,N}(\Delta t)$ in Eq. (3.8) with Ψ_Y, in other words, the loss function will be as follows:

$$\min_{\tilde{K}_{N,\Delta t,m} \in \mathbb{R}^{N\times N}} \|\Psi_Y - \Psi_X \tilde{K}_{N,\Delta t,m}\|_F^2 \tag{5.1}$$

where 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 sampled state at time $t + \Delta t$ via stochastic flow map $\Phi^{\Delta t}(\mathbf{x}_i)$. Starting from the initial condition $\mathbf{x}_i$ at time t, the stochastic flow map $\Phi^{\Delta t}(\cdot)$ captures the evolution of the system under the stochastic dynamics defined in Eq. (2.1). The sampled states $\{\mathbf{y}_i\}_{i=1}^m$ are thus realizations of this stochastic evolution over the time interval Δt.

5.1 Methodology and Discussion

Overall Framework and Design The NN-SDMD framework integrates a neural network to parameterize the dictionary functions $\boldsymbol{\Psi}(\mathbf{x};\theta)$, allowing them to be learned directly from the data. This eliminates the need for manual selection of basis functions and enables a more adaptive representation to the dynamics. The approach alternates between optimizing the approximated Koopman operator $\hat{K}_{N,\Delta t,m}$ and updating the neural network parameters θ, ensuring that both the dictionary functions and the operator approximation improve iteratively. By leveraging parameterized dictionaries Ψ_X and Ψ_Y, this framework retains computational efficiency while ensuring a consistent approximation of the stochastic Koopman semigroup.

Parameterization The dictionary functions $\Psi(\mathbf{x};\theta) = [\psi_1(\mathbf{x};\theta),\ldots,\psi_N(\mathbf{x};\theta)]^\top$ are 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, enabling a flexible and adaptive representation. The Koopman operator $\hat{K}_{N,\Delta t,m}(\theta)$ is also parameterized and computed using:

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

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

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

where γ is some small positive number and $R(\hat{K}_{N,\Delta t,m})$ is a regularization term. In this work, we use Tikhonov regularization $R(\hat{K}_{N,\Delta t,m}) = \|\hat{K}_{N,\Delta t,m}(\theta)\|_F^2$.

Comparison with other methods In gEDMD, the loss function is typically defined to minimize the error associated with the generator $\mathscr{A}$. This requires the computation of $\Psi_X^\top \Psi_X$ and $\Psi_X^\top \mathscr{L}^{\Psi_X}$, where $\mathscr{L}^{\Psi_X}$ involves evaluating Jacobian and Hessian matrices. These calculations can be very computationally expensive, especially for large datasets, as they require repeatedly computing higher-order derivatives during the stochastic gradient descent process and evaluating on data for validation. In contrast, NN-SDMD defines its loss function using Ψ_Y, the parameterized dictionary applied to the evolved data, instead of $\mathscr{L}^{\Psi_X}$. By directly utilizing Ψ_Y, NN-SDMD avoids the explicit evaluation of $\mathscr{L}^{\Psi_X}$, thus significantly reducing computational overhead. This approach is particularly advantageous for large datasets, as it ensures computational efficiency without compromising accuracy. Furthermore, the use of Ψ_Y ensures consistency with the stochastic semigroup's evolution, enabling a more effective approximation of the Koopman operator in stochastic systems.

NN-SDMD can also be compared to EDMD-DL (Extended Dynamic Mode Decomposition with Dictionary Learning) [13], which extends traditional EDMD by replacing manually selected dictionaries with ones learned

from data using simple feedforward neural networks. While NN-SDMD shares similarities with EDMD-DL in its use of neural network-based dictionary learning, a key distinction lies in how $\widehat{K}_{N,\Delta t,m}$ is updated. NN-SDMD employs Eq. (3.4), leveraging the stochastic Taylor expansion and directly incorporating the parameterized Ψ_Y. In contrast, EDMD-DL uses $\widehat{K}_{N,\Delta t,m} = (\Psi_X^\top \Psi_X)^{-1}(\Psi_X^\top \Psi_Y)$, which is better suited for deterministic systems but lacks the stochastic adaptations provided by NN-SDMD.

In summary, NN-SDMD offers significant advantages over both gEDMD and EDMD-DL. By leveraging the stochastic Taylor expansion and utilizing Ψ_Y in the loss function, NN-SDMD achieves a balance between computational efficiency and accuracy, which makes it well-suited to capture the underlying dynamics of stochastic systems.

5.2 Computing Algorithm

Algorithm 2 Estimation of stochastic Koopman operator with NN

Input: 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 γ, learning rate η, number of epochs T.

Estimate SDE coefficients $\mathbf{b}(\cdot)$ and $\sigma(\cdot)$ from the time series data.
Initialize neural network parameters θ.
Initialize Koopman operator $\widehat{K}_{N,\Delta t,m}$.
for epoch = 1 to T **do**
    Compute $\Psi_X(\theta), \Psi_Y(\theta), \mathscr{L}^{\Psi_X}(\theta)$ using $\{(\mathbf{x}_k, \mathbf{y}_k^{(j)})\}_{k=1,j=1}^{m,n}$, $\mathbf{b}(\cdot)$, and $\sigma(\cdot)$
    Compute Gram matrices:
        $\widehat{G}(\theta) = \frac{1}{mn}\Psi_X(\theta)^\top \Psi_X(\theta)$
        $\widehat{H}(\theta) = \frac{1}{mn}\Psi_X(\theta)^\top \mathscr{L}^{\Psi_X}(\theta)$
    Compute loss:
        $J(\theta, \widehat{K}_{N,\Delta t,m}) = \|\Psi_Y(\theta) - \Psi_X(\theta)\widehat{K}_{N,\Delta t,m}\|_F^2 + \gamma\|\widehat{K}_{N,\Delta t,m}\|_F^2$
    Update neural network parameters:
        $\theta \leftarrow \theta - \eta \nabla_\theta J(\theta, \widehat{K}_{N,\Delta t,m})$
    Update approximated Koopman operator:
        $\widehat{K}_{N,\Delta t,m} = I + \Delta t \cdot \left(\widehat{G}(\theta) + \gamma I\right)^{-1} \widehat{H}(\theta)$
end for

Output: Approximated Koopman operator $\widehat{K}_{N,\Delta t,m}(\theta)$.

6 Experiments

This section evaluates the proposed Stochastic Dynamic Mode Decomposition (SDMD) framework through three representative experiments. The 2D Stuart-Landau system demonstrates SDMD's accuracy in capturing eigenvalues of nonlinear oscillatory dynamics under stochastic perturbations. The 1D Ornstein–Uhlenbeck process highlights SDMD's precision in approximating eigenvalues for this system, which correctly recovers several leading eigenvalues as expected from theory. The 2D Triple-Well system explores a time-reversible metastable system, where SDMD accurately identifies transitions between basins and confirms that the approximated eigenvalues are purely real, consistent with theoretical expectations. These experiments collectively show SDMD's advantage, particularly in handling stochasticity and achieving numerical stability with reduced computational cost. The following subsections detail the setups and results. Additional experimental results are available on our GitHub repository for further reference.

6.1 2D Stuart-Landau equation

The 2D stochastic Stuart-Landau (SL) equation [28] serves as a canonical example to study nonlinear oscillatory dynamics under stochastic perturbations. It is frequently used to validate numerical methods for estimating the Koopman operator in stochastic systems.

The equation is expressed in its standard Cartesian coordinates as:

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

where $\delta > 0$ indicates the system is in the limit cycle regime, and $R = \sqrt{\delta/\kappa}$ is the radius of the stable limit cycle. The noise terms dW_x and dW_y are independent Wiener processes with intensity ε.

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

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

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:

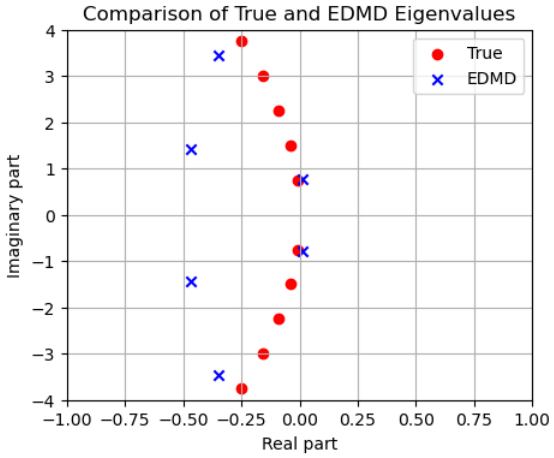
$$\lambda_{ln} = \begin{cases} -\frac{n^2\varepsilon^2}{2R^2} + in(1-\delta) + O(\varepsilon^3), & l = 0,\, n \in \mathbb{Z}, \\ -2l\delta + in(1-\delta) + O(\varepsilon), & l \in \mathbb{Z}^+. \end{cases}$$

Here, $l = 0$ refers to the angular dynamics along the limit cycle, where n indexes the angular harmonics. The real part of the eigenvalues for $l = 0$ describes the phase diffusion rate, proportional to $-n^2\varepsilon^2/R^2$, caused by noise. The imaginary part, proportional to $n(1-\delta)$, corresponds to the angular oscillations around the limit cycle. For $l \in \mathbb{Z}^+$, l represents the degree of the Hermite polynomial in the eigenfunction associated with the eigenvalues. Each l corresponds to a distinct radial mode, with eigenvalues having a real part proportional to $-2l\delta$ and reflecting the decay rates of radial perturbations. $n \in \mathbb{Z}$ represents the azimuthal mode number, capturing angular harmonics. The imaginary part of the eigenvalues, proportional to $n(1-\delta)$, reflects the oscillatory behavior influenced by the system's angular frequency. These eigenvalues provide a rigorous benchmark to validate numerical methods for the Koopman generator, demonstrating both radial and angular contributions to the system's dynamics.

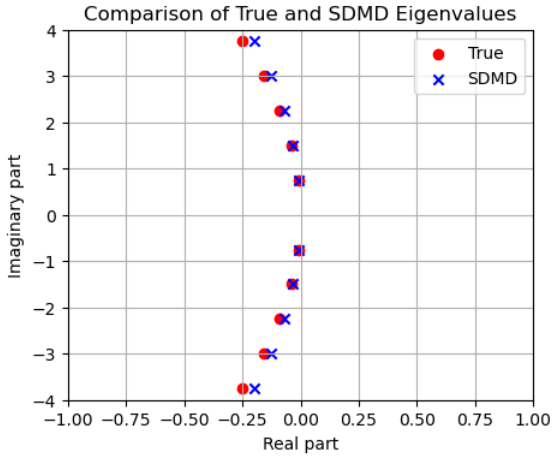
Experiment Design and Result: In our experiments, we aim to evaluate the accuracy of SDMD and EDMD in estimating eigenvalues of the Koopman generator of the stochastic Stuart-Landau system. The Fourier basis is selected as it aligns well with the periodic nature of the system. We specifically compare eigenvalues for selected mode numbers l and n under the following 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, resulting in a uniform grid. The radii are sampled in the range $r \in [0.4, 0.8]$, and the angles are sampled from $\theta \in [-\pi, \pi]$. These points serve as initial conditions for the simulations. For the numerical integration, we use $n_{\text{steps}} = 1000$, with a time step $h = 10^{-6}$. The true eigenvalues of the Koopman generator are computed analytically for $l = 0$. For this mode, we focus on eigenvalues corresponding to azimuthal harmonics, represented by $n \in \{-10, -9, \ldots, 10\}$, explicitly 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 B.1.

6.2 1D Ornstein–Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process is one of the few stochastic processes that has analytical solutions for both its transition probability density and its Koopman generator. The one-dimensional OU process is described



(a) Eigenvalues by EDMD with Fourier basis



(b) Eigenvalues by SDMD with Fourier basis

Figure 2: Comparison of eigenvalues estimated using EDMD and SDMD for the stochastic Stuart-Landau system with Fourier basis. (2a) shows the eigenvalues obtained from EDMD, while (2b) shows those obtained from SDMD.

by the following SDE:

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

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 Koopman generator acting on twice differentiable functions f is [21, 22]:

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

The eigenvalues and eigenfunctions of this generator have explicit forms. For any non-negative integer n, the n-th eigenvalue-eigenfunction pair is:

$$\lambda_n = -n\theta, \quad \phi_n(x) = H_n\left(\frac{x-\mu_0}{\sqrt{\sigma^2/(2\theta)}}\right), \tag{6.1}$$

where $H_n(x)$ is the n-th 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 $\sigma^2/(2\theta)$:

$$\rho_\infty(x) = \left(2\pi\sigma^2/(2\theta)\right)^{-1/2} \exp\left(-(x-\mu_0)^2/(\sigma^2/\theta)\right).$$

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

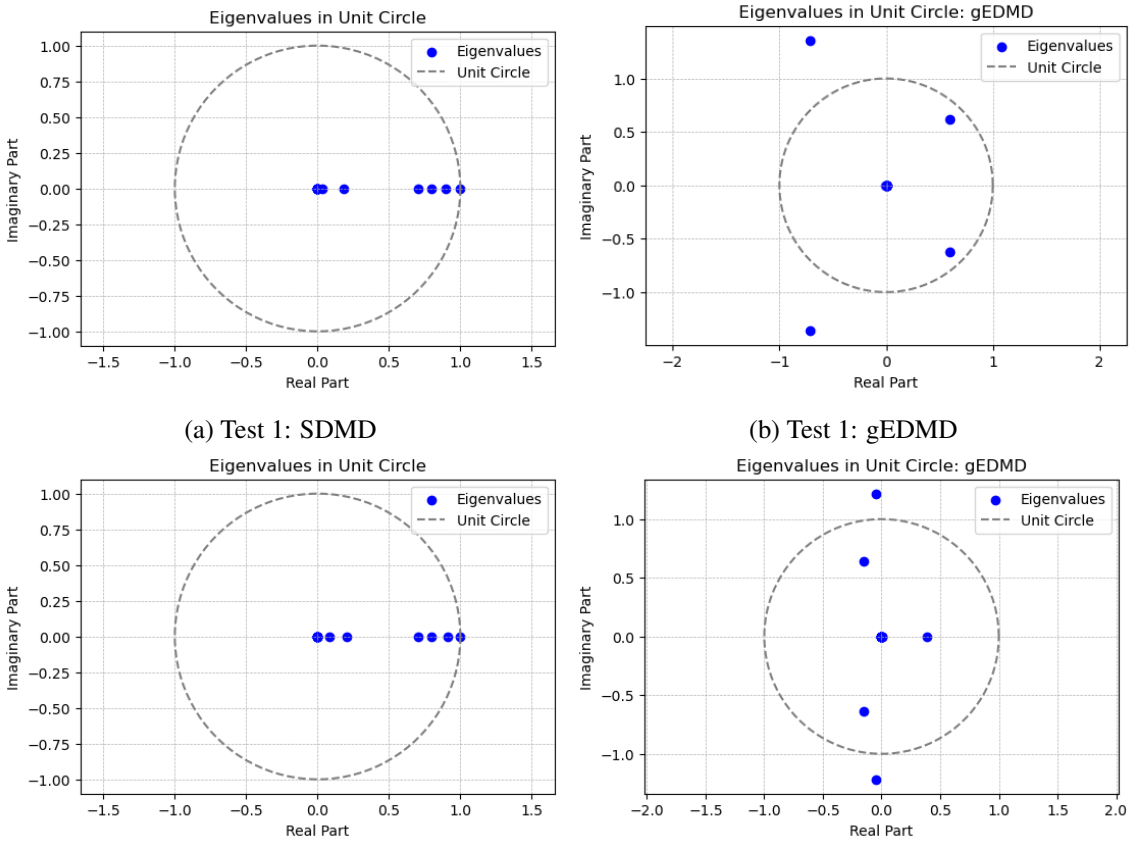
Experiment Design: Our experiments simulate the Ornstein–Uhlenbeck (OU) process using the Euler–Maruyama (EM) method for numerical integration. The parameters are set as $\theta = 1$, $\mu_0 = 0$, and $\sigma = 0.1$. The EM method approximates the solution using discrete time steps $h = 10^{-4}$, with 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. 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. While the SDE (drift and diffusion) coefficients are predefined for simulation, they are independently estimated from the generated time series data using a separate neural network. This network learns the drift coefficient and the diffusion coefficient, which enables the Koopman generator learning process to utilize estimated coefficients rather than true values. This separation isolates the Koopman generator learning task from the system identification step.

Experiment Result: Figure 3 shows the comparison of eigenvalues obtained from NN-SDMD, gEDMD. Figure 4 presents the corresponding eigenfunctions. The results demonstrate that NN-SDMD achieves better accuracy than gEDMD in approximating both the eigenvalues and eigenfunctions, while offering better numerical stability and computational efficiency. Notice that Our SDMD method computes the approximated eigenvalues of the Koopman semigroup $\mathscr{K}^{\Delta t}$ in Figure 3, **NOT** of the generator directly. The approximated eigenvalues of the generator is next computed according to Eq. (2.5). The Table 1 below lists the first four eigenvalues of the generator estimated using SDMD for Test 1 and Test 2. The results show that SDMD successfully approximates the leading four eigenvalues since they are close to $0, -1, -2, -3$, as discussed in Eq. (6.1). However, for eigenvalues corresponding to faster-decaying modes, the accuracy diminishes. This limitation is likely due to insufficient data and an inadequate number of basis functions, constrained by the computational power of the local PC. The eigenfunctions computed by SDMD exhibit polynomial structures that closely match the analytical Hermite polynomial forms, which highlights the method's consistency with theoretical expectations.



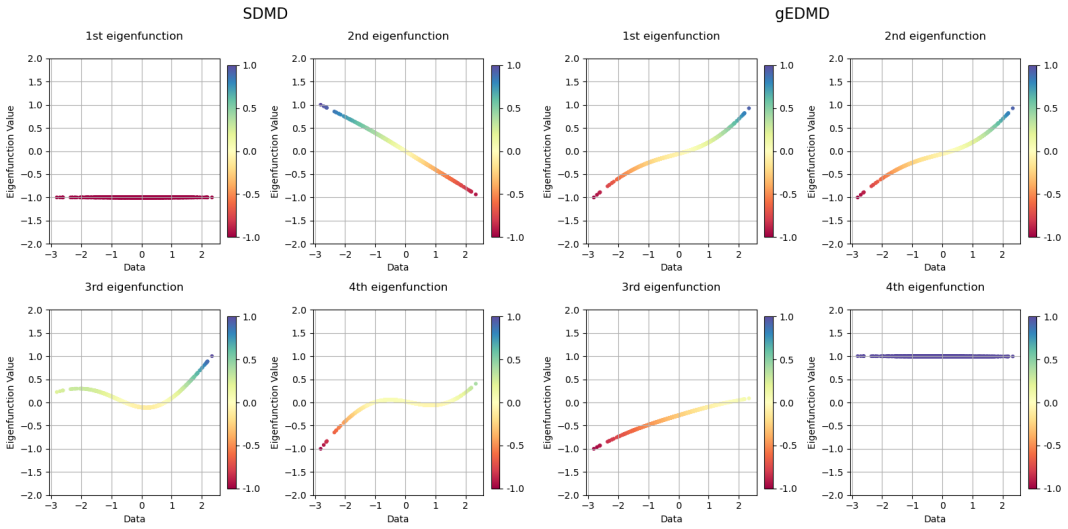
(a) Test 1: SDMD

(b) Test 1: gEDMD

(c) Test 2: SDMD

(d) Test 2: gEDMD

Figure 3: OU process: two tests on computing eigenvalues obtained from SDMD and gEDMD



(a) Test 1 (b) Test 1

(c) Test 2 (d) Test 2

Figure 4: OU process: two tests on computing eigenfunctions obtained from SDMD and gEDMD

Index	Test 1 Eigenvalues	Test 2 Eigenvalues
λ_1	$-5.22708620 \times 10^{-5}$	$-4.68367042 \times 10^{-5}$
λ_2	-1.02764488	-0.902910609
λ_3	-2.01955114	-2.04944006
λ_4	-2.94962176	-2.98903558

Table 1: First four eigenvalues of the Koopman generator estimated using SDMD.

6.3 2D Triple-Well system

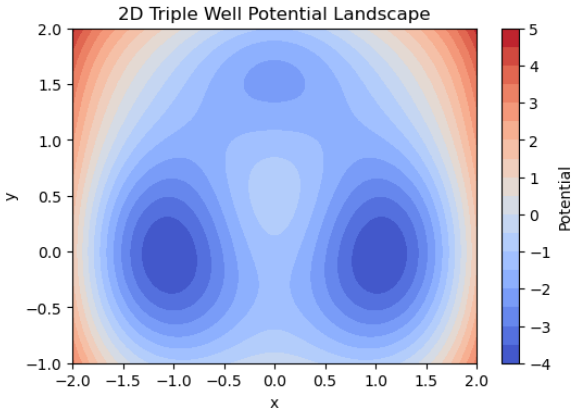
The 2D triple-well potential system [19] provides a rich setting for studying stochastic dynamics due to its intricate energy landscape, which consists of three distinct basins of attraction separated by potential barriers. The dynamics of the system are governed by the stochastic differential equation

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

where $\mathbf{X}_t = (x_t, y_t)$ represents the system's state, $V(x,y)$ is the potential function describing the energy landscape, σ is the diffusion coefficient characterizing stochastic noise, and $\mathbf{W}_t$ is a standard Wiener process.
Experiment design: In this experiment, the potential landscape is defined by

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

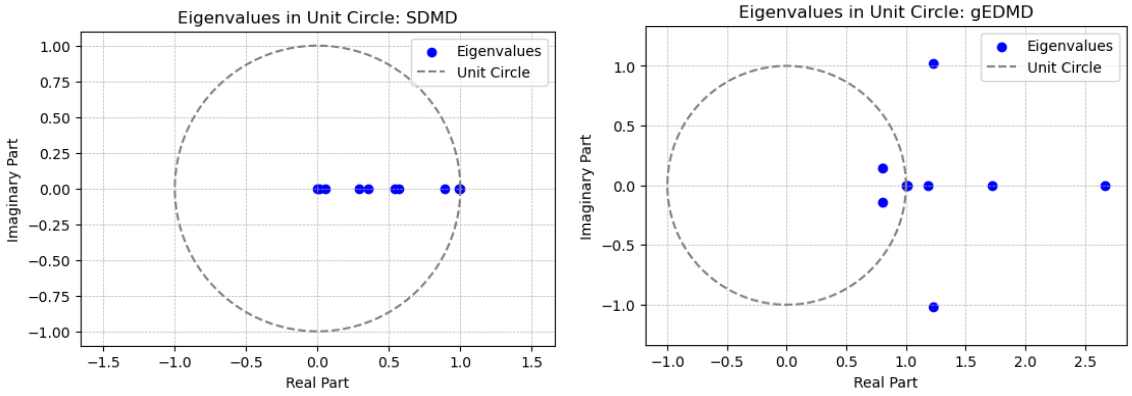
The spatial domain is defined by $x \in [-2,2]$ and $y \in [-1,2]$, with $m = 35$ points uniformly sampled along each dimension to create a grid of initial conditions. The trajectories are generated using the Euler-Maruyama (EM) method for numerical integration, with an integration step size of $h = 1 \times 10^{-3}$ and $n_{\text{steps}} = 100$ steps simulated for each trajectory. Here is the landscape of this potential well system:



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

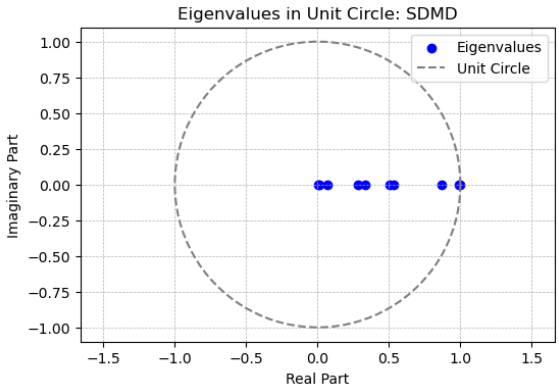
Table 2: First three eigenvalues of the Koopman generator estimated using SDMD.

Experiment Result: Notice that, same as the last example on OU process, our SDMD method computes the approximated eigenvalues of the Koopman semigroup $\mathscr{K}^{\Delta t}$ in Figure 5, **NOT** of the generator directly. The

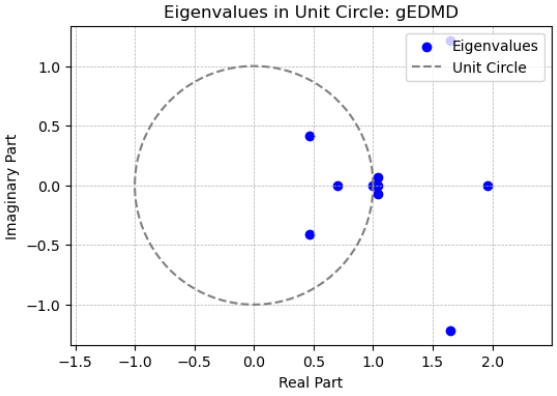


(a) Test 1: SDMD

(b) Test 1: gEDMD

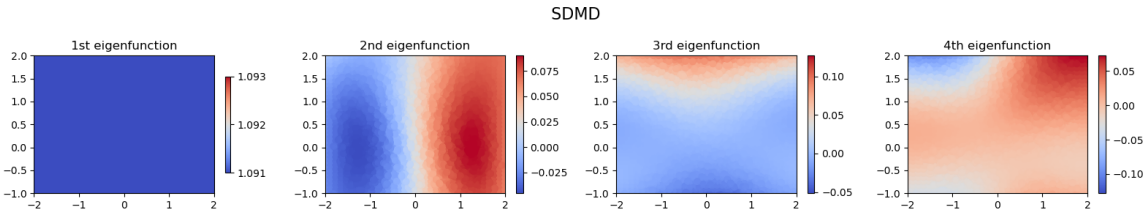


(c) Test 2: SDMD

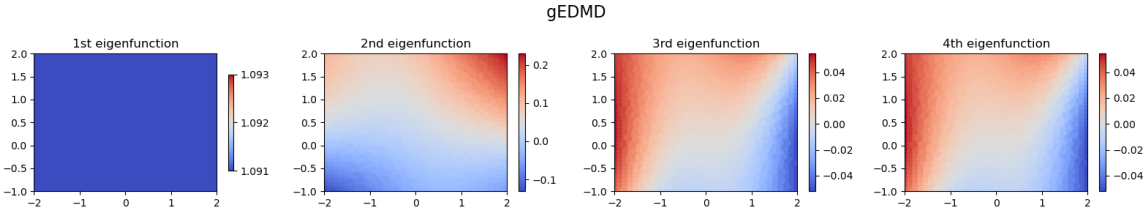


(d) Test 2: gEDMD

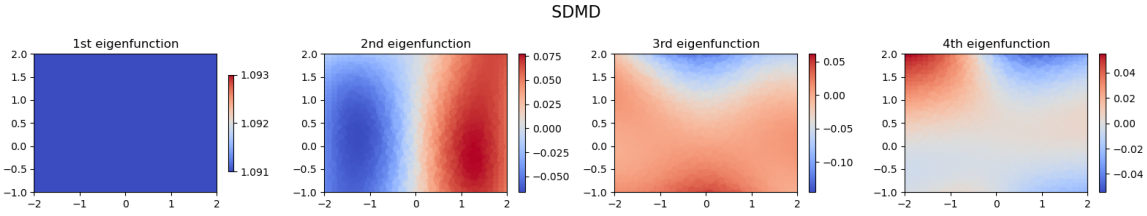
Figure 5: Triple-well: two tests on computing eigenvalues obtained from SDMD and gEDMD



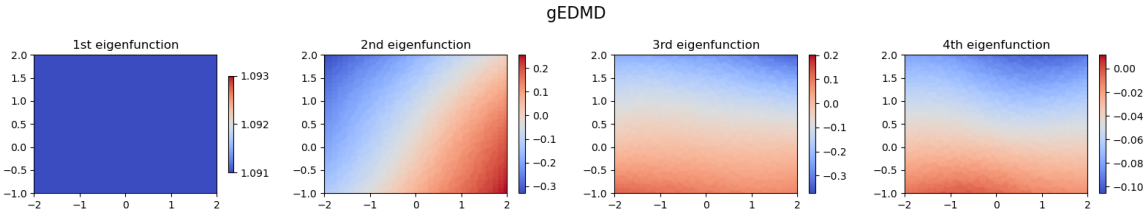
(a) Test 1: SDMD



(b) Test 1: gEDMD



(c) Test 2: SDMD



(d) Test 2: gEDMD

Figure 6: Triple-well: two tests on computing eigenfunctions obtained from SDMD and gEDMD

eigenvalues of the Koopman generator quantify the rates of transitions between basins, with smaller eigenvalues corresponding to slower transitions that reflect the metastable behavior of the system. The eigenvalues for metastable systems like the triple-well potential provides insights into the number and stability of wells. Here's a more detailed analysis for Figure 5, Figure 6 and Table 2: The eigenvalue $\lambda = 1$ corresponds to the system's steady state. The associated eigenfunction of the Koopman operator represents an invariant observable of the system, which reflects quantities that remain unchanged under the system's dynamics. 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 intra-basin dynamics and other transient behaviors that decay quickly. 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 both the number of basins and the hierarchy of transition rates among them.

6.4 Neuroscience example

7 Conclusion

This paper presents SDMD, a novel computational framework for estimating Koopman operators in stochastic dynamical systems. Our approach addresses several challenges in the field and provides a robust foundation for analyzing stochastic dynamics. The proposed method integrates perturbation theory to explicitly account for sampling time, offering stability guarantees even when dealing with unbounded Koopman generators. 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, make the framework particularly suitable for neural network implementations. Rigorous convergence analysis further supports the method's reliability, providing probabilistic error bounds and finite-dimensional approximations. Numerical experiments on examples such as the Stuart-Landau equation, Ornstein-Uhlenbeck process, and triple-well system validate the framework's ability to accurately approximate eigenvalues and eigenfunctions of the Koopman operator.

Future directions include extending the framework of perturbation theory to high-dimensional systems, developing efficient implementations for large-scale applications, and integrating advanced machine learning techniques for automated basis selection. Applying SDMD to more real-world problems in molecular dynamics, climate science, and finance would demonstrate its practical value. Further theoretical advancements, such as handling general noise types and optimizing sampling strategies, could broaden the applicability of the method.

Acknowledgement

References

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

[2] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 12 2012.

[3] Matthew J. Colbrook, Qin Li, Ryan V. Raut, and Alex Townsend. Beyond expectations: Residual dynamic mode decomposition and variance for stochastic dynamical systems. *ArXiv*, abs/2308.10697, 2023.

[4] Matthew J Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of koopman operators for dynamical systems. *Communications on Pure and Applied Mathematics*, 77(1):221–283, 2024.

[5] K.J. 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.

[6] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[7] Dimitrios Giannakis, Gary Froyland, Benjamin Lintner, Max Pike, and Joanna Slawinska. Spectral analysis of climate dynamics with operator-theoretic approaches. In *AGU Fall Meeting Abstracts*, volume 2021, pages A15E–1677, December 2021.

[8] Vladimir Gol'dshtein and Alexander Ukhlov. Weighted sobolev spaces and embedding theorems. *Transactions of the American Mathematical Society*, 361:3829–3850, 2007.

[9] Isao Ishikawa, Yuka Hashimoto, Masahiro Ikeda, and Yoshinobu Kawahara. Koopman operators with intrinsic observables in rigged reproducing kernel hilbert spaces. *arXiv preprint arXiv:2403.02524*, 2024.

[10] Stefan Klus, Feliks Nüske, Sebastian Peitz, Jan-Hendrik Niemann, Cecilia Clementi, and Christof Schütte. Data-driven approximation of the koopman generator: Model reduction, system identification, and control. *Physica D: Nonlinear Phenomena*, 406:132416, 2020.

[11] Vladimir Kostic, Pietro Novelli, Andreas Maurer, Carlo Ciliberto, Lorenzo Rosasco, and Massimiliano Pontil. Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces, 2022.

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

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

[14] Liam Llamazares-Elias, Samir Llamazares-Elias, Jonas Latz, and Stefan Klus. Data-driven approximation of koopman operators and generators: Convergence rates and error bounds, 2024.

[15] Jordan Mann and J. Nathan Kutz. Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, 16(11):1643–1655, 2016.

[16] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. Vampnets for deep learning of molecular kinetics. *Nature Communications*, 9, 2017.

[17] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41:309–325, 2005.

[18] Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual review of fluid mechanics*, 45(1):357–378, 2013.

[19] Paul D. Monsour, Neelang Parghi, Christof Schütte, and Marco Sarich. Metastability and markov state models in molecular dynamics modeling, analysis. 2016.

[20] Frank Noé and Feliks Nuske. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Modeling & Simulation*, 11(2):635–655, 2013.

[21] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin Heidelberg, 2010.

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

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

[24] Friedrich M. Philipp, Manuel Schaller, Septimus Boshoff, Sebastian Peitz, Feliks Nüske, and Karl Worthmann. Variance representations and convergence rates for data-driven approximations of koopman operators, 2024.

[25] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.

[26] Claude Sammut and Geoffrey I. Webb, editors. *McDiarmid's Inequality*, pages 651–652. Springer US, Boston, MA, 2010.

[27] Christof Schütte, Stefan Klus, and Carsten Hartmann. Overcoming the timescale barrier in molecular dynamics: Transfer operators, variational principles and machine learning. *Acta Numerica*, 32:517–673, 2023.

[28] Alexis Tantet, Mickaël D. Chekroun, Henk A. Dijkstra, and J. David Neelin. Ruelle-pollicott resonances of stochastic systems in reduced state space. part ii: Stochastic hopf bifurcation, 2020.

[29] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications, 2014.

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

[31] Mathias Wanner and Igor Mezic. Robust approximation of the stochastic koopman operator. *SIAM Journal on Applied Dynamical Systems*, 21(3):1930–1951, 2022.

[32] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, June 2015.

A

A.1 Derivation for Eq. (2.4)

Given $f \in C_b^2(M)$. From Itô's formula [22], we know that:

$$df(\mathbf{X}_t) = \left(b(\mathbf{X}_t) \cdot \nabla f(\mathbf{X}_t) + \frac{1}{2}\sigma^2(\mathbf{X}_t) : \nabla^2 f(\mathbf{X}_t) \right) dt + \sigma(\mathbf{X}_t) \cdot \nabla f(\mathbf{X}_t) d\boldsymbol{W}_t, \mathbf{X}_0 = \mathbf{x}$$

Taking expectation on its integral form, we have:

$$\mathbb{E}[f(\mathbf{X}_t)|\mathbf{X}_0 = \mathbf{x}] - f(\mathbf{x}) = \int_0^t \left(b(\mathbf{X}_s) \cdot \nabla f(\mathbf{X}_s) + \frac{1}{2}\sigma^2(\mathbf{X}_s) : \nabla^2 f(\mathbf{X}_s) \right) ds$$

Let $t \to 0$, we have:

$$\begin{aligned} \mathscr{A} f(\mathbf{x}) &:= \lim_{t \to 0} \frac{\mathbb{E}[f(\mathbf{X}_t) - f(\mathbf{x})|\mathbf{X}_0 = \mathbf{x}]}{t} \\ &= b(\mathbf{x}) \cdot \nabla f(\mathbf{x}) + \frac{1}{2}\sigma^2(\mathbf{x}) : \nabla^2 f(\mathbf{x}) \end{aligned}$$

Remark A.1. A good candidate of $\mathscr{D}(\mathscr{A})$ can be a weighted Sobolev space $H_\rho^2(M)$ [8], containing all measurable functions $\psi : M \to \mathbb{R}$ with finite norm:

$$\|\psi\|_{H_\rho^2(M)} = \left(\sum_{|\alpha| \le 2} \int_M |D^\alpha \psi|^2 d\rho \right)^{1/2},$$

where $D^\alpha \psi$ represents the weak derivative [6] of order α.

A.2 EDMD

For any observable $f \in \mathscr{F}$, using the matrix representation $K_{\Delta t}$ of the projected Koopman operator $\mathscr{P}_N \mathscr{K}^{\Delta t} \mathscr{P}_N$ with respect to the basis $\{\psi_1, \ldots, \psi_N\}$, we have:

$$\begin{aligned} \mathscr{P}_N \mathscr{K}^{\Delta t} \mathscr{P}_N f &= \mathscr{P}_N \mathscr{K}^{\Delta t} \left(\sum_{j=1}^N c_j \psi_j \right) \\ &= \sum_{j=1}^N c_j \left(\mathscr{P}_N \mathscr{K}^{\Delta t} \psi_j \right) \\ &= \sum_{j=1}^N c_j \left(\sum_{i=1}^N \psi_i [K_{\Delta t}]_{ij} \right) \\ &= \sum_{i,j=1}^N \psi_i [K_{\Delta t}]_{ij} c_j \end{aligned} \tag{A.1}$$

where $c_j = \langle f, \psi_j \rangle_\rho$ for some measure ρ, $[K_{\Delta t}]_{ij}$ denotes the ij-th element of matrix $K_{\Delta t}$ and $\mathscr{P}_N : \mathscr{F} \to \mathscr{F}_N$ is the orthogonal projection operator.

A.3 McDiarmid's Inequality

Definition A.2 (Bounded Differences Property)**.** A function $f : \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_n \to \mathbb{R}$ 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 $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2, \ldots, x_n \in \mathscr{X}_n$,

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

Lemma A.3 (McDiarmid's Inequality)**.** Let $f : \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_n \to \mathbb{R}$ satisfy the bounded differences property with bounds $c_1, c_2, \ldots, c_n$ as in A.2. Consider independent random variables $X_1, X_2, \ldots, X_n$ where $X_i \in \mathscr{X}_i$ for all i. Then, for any $\varepsilon > 0$,

$$\mathbb{P}(|f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)]| \geq \varepsilon) \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

Proof. See [26]. □

A.4 First Trotter–Kato Approximation Theorem

Theorem A.4 (First Trotter–Kato Approximation Theorem)**.** Let $(\mathscr{K}^t)_{t\geq 0}$ and $(\mathscr{K}_N^t)_{t\geq 0}$, $N \in \mathbb{N}$, be strongly continuous semigroups on M with generators $\mathscr{A}$ and $\mathscr{A}_N$, respectively, and assume that they satisfy the estimate

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

and some constants $D \geq 1$, $\omega \in \mathbb{R}$. Take $\mathscr{D}$ to be a core for $\mathscr{A}$ and consider the following assertions.

(a) $\mathscr{D} \subset \mathscr{D}(\mathscr{A}_N)$ for all $N \in \mathbb{N}$ and $\mathscr{A}_N f \to \mathscr{A} f$ for all $f \in \mathscr{D}$.

(b) $\mathscr{K}_N^t f \to \mathscr{K}^t f$ for all $f \in \mathscr{F}$, uniformly for t in compact interval.

Then the implication (a) $\implies$ (b) holds, while (b) does not imply (a).

Proof. See [5, Section 4.8] □

B

B.1 2D Stuart-Landau equation: Phase Diffusion Equation

The SL equation not only has the polar coordiantes form, but also has phase coordinates form [28, Section 4] given in the following:

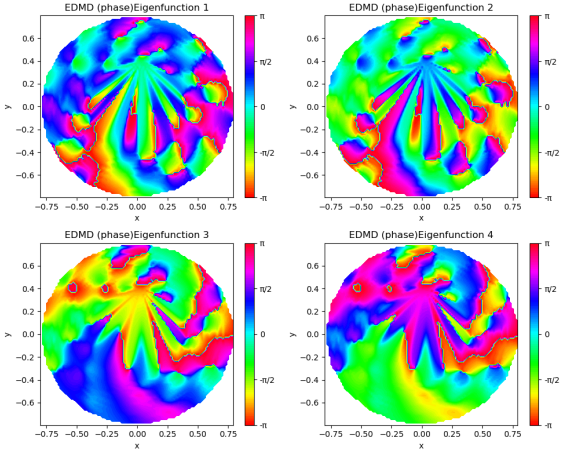
$$dr = (\delta r - \kappa r^3 + \frac{\varepsilon^2}{2r})\,dt + \varepsilon\,dW_r,$$
$$d\phi = \omega_f\,dt + \frac{\varepsilon}{r}\,dW_\theta - \tilde{\beta}\frac{\varepsilon}{r}\,dW_r,$$

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

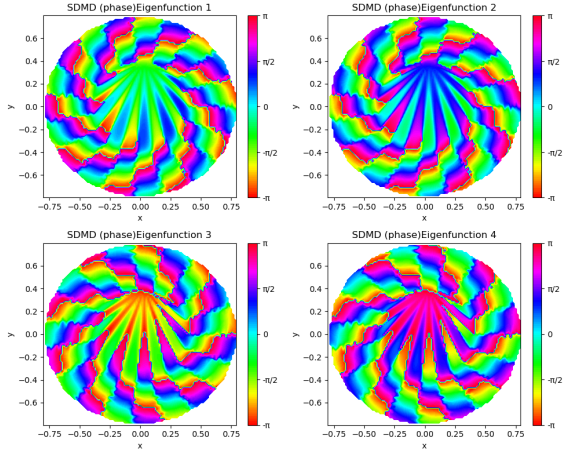
For $\delta > 0$ and $\varepsilon\sqrt{\kappa}/\delta \ll 1$, the eigenfunctions are given by:

- $l = 0$:
$$\psi_{0n} = \exp(i(n(\theta - \tilde{\beta}\log\frac{r}{R})))$$
- $l > 0$:
$$\psi_{ln} \propto H_l(\sqrt{2\delta}\frac{r-R}{\varepsilon})\exp(i(n(\theta - \tilde{\beta}\log\frac{r}{R})))$$
where H_l is the l-th order Hermite polynomial.

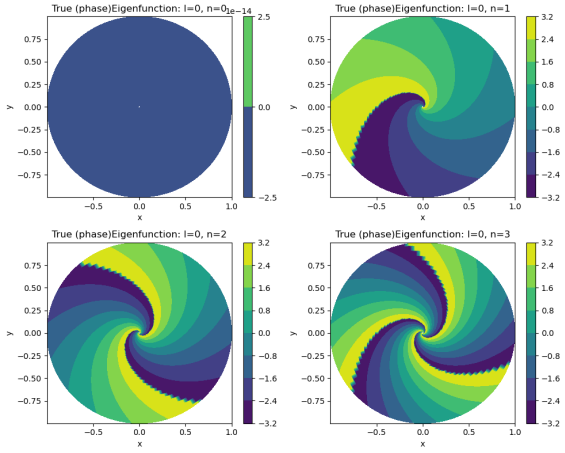
Figure 7 compares the analytical, EDMD-approximated, and SDMD-approximated eigenfunctions for different modes. The analytical eigenfunctions are normalized by a factor of $1/\sqrt{2^{|n|}|n|!}$, combining the phase dynamics through $\exp(in(\theta - \tilde{\beta}\log(r/R)))$ and radial structure through Hermite polynomials $H_l(\sqrt{2\delta}(r-R)/\varepsilon)$. Both EDMD and SDMD approximate these eigenfunctions using Fourier basis functions, with SDMD showing better accuracy in capturing the phase structure. Since our chosen data is for $r \in [0.4, 0.8]$, we do not have information for the disk area with $r < 0.4$. For $r \in [0.4, 0.8]$, we can see that the eigenfunctions by SDMD also have the rotational behaviour as in the true eigenfunctions of Figure 7.



(a) Eigenfunctions by SDMD



(b) Eigenfunctions by SDMD



(c) True eigenfunctions

Figure 7: Comparison of eigenvalues estimated using EDMD and SDMD for the stochastic Stuart-Landau system with Fourier basis. (7a) shows the eigenvalues obtained from EDMD, while (7b) shows those obtained from SDMD.