




## Kernel-based learning framework for discovering the governing equations of stochastic jump-diffusion processes directly from data

Wenqing Sun , Jinqian Feng ,\* Jin Su , and Qin Guo

*School of Science, Xi'an Polytechnic University, Xi'an 710048, China*



(Received 5 May 2023; accepted 28 August 2023; published 15 September 2023)

Discovering the underlying mathematical-physical equations of complex systems directly from observational data has been a challenging inversion problem. We propose a data-driven framework for identifying dynamical information in stochastic diffusion or stochastic jump-diffusion systems. The probability density function is utilized to relate the Kramers-Moyal expansion to the governing equations, and the kernel density estimation method, improved by the Fourier transform idea, is used to extract the Kramers-Moyal coefficients from the time series of the state variables of the system. These coefficients provide the data expression of the governing equations of the system. Then a data-driven sparse identification algorithm is used to reconstruct the underlying dynamic equations. The proposed framework does not rely on prior assumptions, and all results are obtained directly from the data. In addition, we demonstrate its validity and accuracy using illustrative one- and two-dimensional examples.

DOI: [10.1103/PhysRevE.108.035306](https://doi.org/10.1103/PhysRevE.108.035306)

### I. INTRODUCTION

The nonlinearity, strong randomness, and multivariable characteristics of most dynamical systems in reality expose the great shortcomings and deficiencies of the traditional first-principles-based derivations and empirical analyses, which pose a great challenge for cross-integration in some critical application areas. Fortunately the continuous development of computing hardware and high-performance sensors has accumulated a huge resource of data based on actual observations or numerical simulations for the application of data-driven techniques, which provides new directions and opportunities for the inverse problem of discovering the underlying dynamics of complex dynamical systems directly from the data. Given the availability of data to facilitate and support the integration of machine learning and physical modeling, such problems have stimulated the research interests of many scholars in the engineering sciences and even in the social sciences. Examples include earth system science [1], epidemiology [2], fluid dynamics [3,4], finance [5], public transportation [6], etc.

In recent years, as the fourth paradigm of scientific discovery [7], data-based modeling has developed several mainstream approaches in the field of analysis of complex nonlinear systems, which not only improve the needs of practical applications but also provide new ideas for understanding and predicting the dynamics of complex systems. Some well-known methods include dynamic mode decomposition (DMD) methods [8–11], Koopman operator theory (which is usually combined with DMD) [12–15], physics-informed neural networks (PINNs) [16–19], and sparse identification of

nonlinear dynamical systems (SINDy) [20–24], among other techniques. However, most of these data-based modeling and analytical techniques are limited to the remarkable results achieved in solving some deterministic dynamical systems whose governing equations are mostly ordinary differential or partial differential equations. Due to the challenges of strong stochasticity, chaotic behavior, and high-dimensional properties of dynamical systems in real-world problems, it is still necessary to explore new technologies to meet constant practical needs.

A significant metric to describe the uncertainty quantification of a complex system is its probabilistic response mapping function, which contains spatial and/or temporal coordinates. It has extremely important extensions and applications in estimating the transition dynamics of complex systems, exploring the most probable transfer paths of system state variables, and even reconstructing the equations of motion of complex systems [25–27]. The stochastic evolution of the probability response function can usually be expressed by the Kramers-Moyal (KM) equation, which gives the equilibrium relationship between the first-order difference of the probability response function of the continuous process with respect to the discrete timescale in the limit sense and the infinite power series of the difference of the state variables at adjacent moments. The KM expansion has been extended from its earlier classical form based on the derivation of Markovian processes to some new variations including the KM for delayed differential equations and the nonlocal KM to analyze non-Markov processes [28,29]. The Pawula theorem proves that the well-known Fokker-Planck equation is the special form of the second-order truncated KM equation and that the drift and diffusion coefficients of the stochastic diffusion process correspond to the first two orders of KM coefficients, respectively [30]. Therefore, estimating the KM coefficients accurately is of positive significance for reconstructing the

\*School of Science, Xi'an Polytechnic University, Xi'an 710048, China; jqfeng15@126.com, wqsun98@163.com

governing equations of motion for complex systems based on the data. The Nadaraya-Watson estimator is a regression model based on the kernel density estimation method proposed successively by Nadaraya and Watson, which allows the analysis of stationary and nonstationary time series by estimating the time-dependent KM coefficients [31].

The kernel density estimation (KDE) method serves not only as a core component of the Nadaraya-Watson estimator but also as an important tool in statistical theory of data and already has some well-established theoretical studies [32–34]. In particular, in recent years, some scholars have developed many high-performance and visualized R or Python package programs by integrating the theory of KDE with machine learning methods, which have made the method become a very sought-after and effective component for studying the problems of complex systems [35–38]. The main challenges of the classical KDE methods are the computational complexity of the kernel components to estimate the probability distribution of the large sample data in the average sense, and the accurate selection of the bandwidth parameter for the true probability density function of the sample to be estimated under the asymptotic mean integrated squared error (AMISE).

The Binning technique can subdivide the sample intervals into equally spaced grids smaller than the number of samples, and reduce the computational burden of the KDE by convolving the grid weights given by the data. In the multivariate case, combining the idea of binning and the fast Fourier transform method can save a lot of computational time [39,40]. The bandwidth as a built-in parameter of the kernel function is the most important issue that needs to be considered whether in univariate or multivariate problems. There are three major types of bandwidth selectors based on the AMISE theory: The rule-of-thumb approach, the cross-validation method, and the plug-in method. However, the rule-of-thumb approach requires certain prior information (usually assuming that the true probability density function is normal), the scheme of cross-validation ideas has a high degree of variability in estimating the sample, although the estimation results of the plug-in method are more stable, and the computational complexity caused by the iteration step is higher (especially for multivariate samples). Bernacchia *et al.* proposed a self-consistent estimation method that can accurately reproduce the potential probability distribution of the sample data. This nonparametric estimation method bypasses the subjective choice of parameters, does not require any prior information, and proves the theoretical limit of its squared error proportional to the size of the data set [41]. This method has greatly inspired this paper. O'Brien *et al.* have successively expanded the application of this method in one-dimensional and multidimensional practical problems [42,43].

In this paper, the Nadaraya-Watson estimator is used to study the problem of identifying the multivariate stochastic jump-diffusion process. Random jump events in the given sample are detected and described by estimating the higher order conditional moments of the KM equation in real time. Once the drifts, diffusion coefficients, and jump parameters of the Itô stochastic differential equations for the underlying dynamics of the state variables are calculated, the corresponding abstract governing equations of motion can be given by

combining the ideas of SINDy so as to reconstruct the underlying dynamical processes of the system directly from the data. Since the estimated drift and diffusion coefficients are clearly given with the trends of the state variables, this can reduce the computational complexity of the SINDy procedure by reducing the prior information needed to construct the library of functions. The framework proposed in this paper is based on a strict theoretical basis, and there is a strong physical interpretation that “black box” models such as neural network methods do not have. The results obtained are not simple data forms but abstract mathematical equations, which can provide a new reference for obtaining potential physical laws from data.

The article is structured as follows. In Sec. II we introduce the background knowledge of the multidimensional stochastic jump-diffusion process including the derivation of its corresponding KM equation and its discriminant formula with the stochastic diffusion process. In Sec. III we complete the construction of the Nadaraya-Watson estimator using the self-consistent KDE algorithm and represent the higher order KM coefficients using the Nadaraya-Watson estimator. With the help of the trends in the data mapping relationships between the estimated KM coefficients and the state variables, the appropriate library of basis functions is constructed and the corresponding abstract mathematical expressions are given using the symbolic regression algorithm. In Sec. IV the validity and adaptiveness of the framework are tested by simulating three one-dimensional and/or two-dimensional stochastic jump-diffusion processes. Finally, Sec. V concludes the study and gives some perspectives.

## II. PROBLEM STATEMENT

Consider a class of typical multidimensional stochastic jump-diffusion dynamical systems, whose Itô interpreted governing differential equations can be given as

$$\begin{aligned} d\mathbf{x}(t) &= \mathbf{a}(\mathbf{x}, t)dt + \mathbf{b}(\mathbf{x}, t)d\mathbf{W}(t) + \mathbf{c}(\mathbf{x}, t)d\mathbf{J}(t), \\ \mathbf{x}(0) &= \mathbf{x}_0, \quad t \in [0, T], \end{aligned} \quad (1)$$

where  $\mathbf{x}_0$  is the state vector at the initial moment;  $\mathbf{x}(t) \in \mathbb{R}^d$  denotes the  $d$ -dimensional state variable;  $\mathbf{a}(\mathbf{x}, t) \in \mathbb{R}^d$  indicates the general nonlinear drift function vector, which depends on the stochastic variable vector  $\mathbf{x}(t)$ ;  $\mathbf{b}(\mathbf{x}, t) \in \mathbb{R}^{d \times s}$  represents the diffusion matrix;  $\mathbf{W}(t) \in \mathbb{R}^s$  stands for the Wiener process vector, which are independent from each other, i.e.,  $\langle dW_i dW_j \rangle = 0 (i \neq j, i, j = 1, \dots, s)$ ; the jump amplitude parameter  $\mathbf{c}(\mathbf{x}, t) \in \mathbb{R}^{d \times l}$  usually assumed as a random matrix that satisfies the normal distribution  $\mathcal{N}(0, \sigma_c^2)$ ; and  $\mathbf{J}(t) \in \mathbb{R}^l$  usually modeled as the Poisson jump processes with Borel measurable and bounded jump rate  $\lambda \in \mathbb{R}^l$  on the filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{P})$  that satisfies the condition of mutual independence, i.e.,  $\langle dJ_i dJ_j \rangle = 0 (i \neq j, i, j = 1, \dots, l)$ . In addition, the Wiener process  $\mathbf{W}(t)$  and the Poisson jump process  $\mathbf{J}(t)$  in the above equation are also independent of each other.

In practice, for continuous systems, the sample data are observed at discrete moments. Therefore, a fixed very small snapshot  $\tau$  of time is allowed to obtain  $N$  observations in time period  $t \in [0, T]$ . The time-continuous Euler-Maruyama

(EM) scheme is given on the basis of a random initial state as

$$\begin{aligned} x_k^{(n+1)} &= x_k^{(n)} + a_k(\mathbf{x}^{(n)}, t_n)\tau_n \\ &+ \sum_{i=1}^s b_{k,i}(\mathbf{x}^{(n)}, t_n)\Delta W_i^{(n)} \\ &+ \sum_{j=1}^l c_{k,j}(\mathbf{x}^{(n)}, t_n)\Delta J_j^{(n)}, \end{aligned} \quad (2)$$

where  $a_k$  is the drift coefficient of the  $k$ th components of the state vector  $\mathbf{x}(t) \in \mathbb{R}^d$ ,  $b_{k,i}$  is the component of the  $k$ th row and  $i$ th column of the diffusion matrix  $\mathbf{b}$ , and  $c_{k,j}$  is the element of the  $k$ th row and  $j$ th column of the matrix of the jump amplitude  $\mathbf{c}$ , for  $k \in 1, 2, \dots, d$ ,  $i \in 1, 2, \dots, s$ , and  $j \in 1, 2, \dots, l$ .  $\Delta W_i^{(n)} = (W_i^{(n+1)} - W_i^{(n)})$ , and  $\Delta J_j^{(n)} = (J_j^{(n+1)} - J_j^{(n)})$  denote the independent increments of the Wiener process and the jump process, respectively. For the EM scheme, we need to emphasize the explicit dependence on the initial value  $x^{(0)} = x_0$  in the notation. In this case, the coefficients  $a_k$ ,  $b_{k,i}$ , and  $c_{k,j}$  are Lipschitz. In addition,  $a_k$ ,  $b_{k,i}$  and  $c_{k,j}$  all satisfy the linear growth condition, i.e., there exist  $h_1, h_2, h_3 \in (0, \infty)$  such that

$$\begin{aligned} |a_k(x)| &\leq h_1(1 + |x|), \\ |b_{k,i}(x)| &\leq h_2(1 + |x|), \\ |c_{k,j}(x)| &\leq h_3(1 + |x|). \end{aligned} \quad (3)$$

It is well known that SDE (1) admits a unique strong solution that can be approximated by the EM scheme at strong convergence order  $1/2$  [44].

The short-time propagator of the system state vector  $\mathbf{x}(t)$  can be captured by the KM expansion [45] as

$$\frac{\partial p(\mathbf{x}, t \mid \mathbf{x}', t')}{\partial t} = \mathcal{L}_{KM}[p(\mathbf{x}, t \mid \mathbf{x}', t')], \quad (4)$$

with the initial conditional probability  $p(\mathbf{x}, t \mid \mathbf{x}', t) = \delta(\mathbf{x} - \mathbf{x}')$ , where  $\delta$  is the generalized Dirac delta function.  $t$  and  $t'$  are two adjacent moments after the discretization of the continuous process, respectively, and the positive difference timescale  $\tau$  between them is a very small value that tends to 0. In this case, for a stochastic system with  $d$  components  $x_j (j = 1, \dots, d)$ , the expansion of  $\mathcal{L}_{KM}$  can be given as

$$\begin{aligned} &\mathcal{L}_{KM}[p(\mathbf{x}, t \mid \mathbf{x}', t')] \\ &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^{v_1, \dots, v_d}}{\prod_{j=1}^d \partial x_j^{v_j}} [K^{(v_1, \dots, v_d)}(\mathbf{x}, t) p(\mathbf{x}, t \mid \mathbf{x}', t')], \end{aligned} \quad (5)$$

where  $v_j$  denotes the differential order for the  $j$ th component of the state vector  $\mathbf{x}$ , which takes on the entire set of natural numbers  $\mathbb{N}$ ; the subscript of the summation symbol  $n = \sum_{j=1}^d v_j$ .

The conditional moments of the system (1) with finite time span  $\tau$  are usually defined as

$$\begin{aligned} K^{(v_1, v_2, \dots, v_d)}(\mathbf{x}, t) &= \left\langle \prod_{j=1}^d dx_j(t)^{v_j} \right\rangle_{\mathbf{x}(t)=\mathbf{x}} \\ &= \left\langle \prod_{j=1}^d (x_j(t + \tau) - x_j(t))^{v_j} \right\rangle_{\mathbf{x}(t)=\mathbf{x}} \\ &= \int_{\mathbb{R}^d} \left[ \prod_{j=1}^d (x_j(t + \tau) - x_j(t))^{v_j} \right] \\ &\quad \times p(\mathbf{x}, t + \tau \mid \mathbf{x}', t) d\mathbf{x}. \end{aligned} \quad (6)$$

The conditional moments  $K(\mathbf{x}, t)$  and the KM coefficients  $D(\mathbf{x}, t)$  of the system satisfy the following relationship:

$$D^{(v_1, v_2, \dots, v_d)}(\mathbf{x}, t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} K^{(v_1, v_2, \dots, v_d)}(\mathbf{x}, t). \quad (7)$$

According to the Pawula theorem, the KM expansion of a general stochastic diffusion process is truncated at  $n = 2$ . However, there exist nonvanishing KM coefficients greater than 2 for stochastic processes with random jump event effects [46].

### III. SYSTEM IDENTIFICATION FRAMEWORK

In general, the observed data are either stationary or non-stationary. Most of the measured time series are nonstationary, which means that the KM coefficients may be explicitly time-dependent. The kernel function estimation method proposed by Nadaraya and Watson allows us to estimate the time-dependent local KM coefficients. This method can be used to analyze both stationary and nonstationary stochastic processes. The statistical features of the noise can be analyzed based on the limiting vanishing behavior of the KM condition moments with timescale  $\tau$ , and then the governing equations of the dynamics can be reconstructed.

To estimate the conditional moments in Eq. (6) from the time-series sample, the conditional probability density function of the state transfer process should be considered. According to Bayesian theory, the conditional probability density function of the system at the  $r$ th moment shows the dependence on its state at the previous  $r - 1$  moments, and for general processes its conditional probability can be defined as

$$\begin{aligned} &p(\mathbf{x}_r, t_r \mid \mathbf{x}_{r-1}, t_{r-1}; \dots; \mathbf{x}_1, t_1) \\ &= \frac{p_r(\mathbf{x}_r, t_r; \dots; \mathbf{x}_1, t_1)}{p_{r-1}(\mathbf{x}_{r-1}, t_{r-1}; \dots; \mathbf{x}_1, t_1)} \\ &= \frac{p_r(\mathbf{x}_r, t_r; \dots; \mathbf{x}_1, t_1)}{\int p_r(\mathbf{x}_r, t_r; \dots; \mathbf{x}_1, t_1) d\mathbf{x}_r}. \end{aligned} \quad (8)$$

Through the probabilistic transformation of the state variables of the original stochastic system in Eq. (8), Eq. (7) can

be rewritten as

$$\begin{aligned} D^{(v_1, v_2, \dots, v_d)}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{\int_{\mathbb{R}^d} \left[ \frac{1}{\tau} \prod_{j=1}^d [x_j(t + \tau) - x_j(t)]^{v_j} \right] p(\mathbf{x}, t) d\mathbf{x}_t}{\int_{\mathbb{R}^d} p(\mathbf{x}, t) d\mathbf{x}_t}. \end{aligned} \quad (9)$$

In applied science, the problem of point estimation of continuous density from a discrete set of data is very common. As a very theoretically mature and widely used data smoothing method, the KDE method can estimate the potential distribution of the data based only on the data itself by choosing a variety of kernel functions. For the multivariate problem studied in this paper,  $\hat{p}(\mathbf{x}, t)$  in Eq. (9) is an approximate estimate of the joint probability density function  $p(\mathbf{x}, t)$ , which can be classically written by KDE theory [33] as

$$\begin{aligned} \hat{p}(\mathbf{x}, t) &= \frac{1}{N} \sum_{i=1}^N |\mathbf{H}|^{-1/2} K[\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{x}_i)] \\ &= \frac{1}{N} \sum_{i=1}^N K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i), \end{aligned} \quad (10)$$

where  $\mathbf{K}$  is the kernel function operator,  $K_{\mathbf{H}}(\cdot) = |\mathbf{H}|^{-1/2} K[\mathbf{H}^{-1/2}(\cdot)]$ ,  $\mathbf{H}$  is a bandwidth matrix of  $d \times d$  dimensions, which is symmetric and positive definite, and  $d$  is determined by the state variable  $\mathbf{x}$  of the problem under study.  $\mathbf{x}$  is the center of the kernel component, and the traditional KDE is estimated with each data point as the center.

The traditional form of KDE like Eq. (10) requires a choice of kernel bandwidth parameters. However, this selection is not straightforward; it may require making assumptions about the potential distribution of the sample data and then selecting the optimal bandwidth by constructing an AMISE model based on the differentiability of the kernel function. With the principle of not assuming additional prior information and saving computational resources, we prefer to focus on finding the optimal shape of the kernel rather than finding the optimal bandwidth for the kernel function. Inspired by the idea of an enhanced self-consistent KDE of arbitrary dimensionality proposed by Bernacchia [41], Eq. (10) can be rewritten in the form of a sum of convolutions between data-centric kernel functions and delta functions:

$$\begin{aligned} \hat{p}(\mathbf{x}, t) &= \frac{1}{N} \sum_{i=1}^N K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i) \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} K(\mathbf{z}) \cdot \delta(\mathbf{x} - \mathbf{x}_i - \mathbf{z}) d\mathbf{z} \\ &= \frac{1}{N} \sum_{i=1}^N K(\mathbf{x}) \times \delta(\mathbf{x} - \mathbf{x}_i), \end{aligned} \quad (11)$$

where  $\times$  is the convolution symbol and  $\delta$  is the generalized Dirac delta function.

By using the idea of the Fourier transform, we can easily obtain the equivalent form of KDE  $\hat{p}(\mathbf{x}, t)$  in frequency

space

$$\begin{aligned} \phi(\omega) &= \mathcal{F}_{\omega}[\hat{p}(\mathbf{x}, t)] = \mathcal{F}_{\omega} \left[ \frac{1}{N} \sum_{j=1}^N K(\mathbf{x}) \times \delta(\mathbf{x} - \mathbf{x}_j) \right] \\ &= \kappa(\omega) \cdot \frac{1}{N} \sum_{j=1}^N e^{i\mathbf{x}_j \omega} \\ &= \kappa(\omega) \cdot \mathcal{C}(\omega), \end{aligned} \quad (12)$$

where  $\mathcal{F}_{\omega}$  denotes the Fourier transform operator for mapping from data space  $\mathbf{x}$  to frequency space  $\omega$ ;  $\kappa$  represents the Fourier transform of kernel  $K$ ; and  $\mathcal{C}(\omega)$  is denoted as the empirical characteristic function of the data. In fact, there is agreed to exist an optimal transformation kernel  $\hat{\kappa}(\omega) = N \cdot (N - 1 + |\phi|^{-2})^{-1}$  which minimizes the mean square error between the true probability density function  $p(\mathbf{x}, t)$  of the sample data and the optimal KDE result  $\hat{p}(\mathbf{x}, t)$  [42].

The relationship between the optimal transform kernel and the empirical characteristic function is defined as

$$\hat{\kappa}(\omega) \equiv \frac{N}{2(N-1)} \left[ 1 + \sqrt{1 - \frac{4(N-1)}{N^2 |\mathcal{C}(\omega)|^2} I_{\bar{A}}(\omega)} \right], \quad (13)$$

where  $I_{\bar{A}}$  denotes the frequency filter, which is 1 for the set  $\bar{A}$  of frequencies accepted in the KDE and 0 otherwise. For the stability of the KDE, the set of acceptable frequencies  $\{\omega | \omega \in \bar{A}\}$  needs to be specified to satisfy  $|\mathcal{C}(\omega)|^2 \geq C_{\min}^2 = 4(N-1)N^{-2}$ . The bounded frequency set  $\bar{A}$  not only can exclude arbitrary additional subsets of other admissible frequencies, but also reflects the arbitrariness of the initial guess  $\phi_0$  of the iterative solution. Furthermore, the stability condition on  $\bar{A}$  forces  $\kappa(\omega)$  to be real-valued, which ensures that the kernel  $K(\mathbf{x})$  assumed in its data space is symmetric. If the empirical characteristic function is also integrable,  $\phi$  can converge to the true potential distribution as  $N$  increases. Finally, once the optimal  $\hat{\phi}$  is obtained, the estimation of the probability density function  $\hat{p}(\mathbf{x}, t)$  of the sample space can be reverted using the inverse Fourier transform.

Thus based on the idea of KDE, Eq. (9) can be further transformed into the following Nadaraya-Watson estimator:

$$\begin{aligned} D^{(v_1, v_2, \dots, v_d)}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{\sum_{i=1}^N \left[ \frac{1}{\tau} \prod_{j=1}^d [x_j(t + \tau) - x_j(t)]^{v_j} \right] \times \mathcal{F}_t^{-1}(\hat{\phi}(\omega))}{\sum_{i=1}^N \mathcal{F}_t^{-1}[\hat{\phi}(\omega)]}. \end{aligned} \quad (14)$$

Clearly, the data requirements of the proposed computational framework depend largely on the estimation of the underlying distribution of the time series. To reduce the cost of the computation of this framework, an effective approach is to use the nonuniform fast Fourier transform method [47] for acceleration. This fast self-consistent KDE algorithmic framework converges as the number of sample data  $N$  increases, and not only has a computational accuracy that is not inferior to the state-of-the-art classical KDE estimation, but also overcomes the problems of dimensional catastrophe and computational complexity to some extent, and even the

convergence is further accelerated when the sample data exceeds  $10^4$ ; these findings are illustrated by examples by O'Brien [43].

### A. Case I: Stochastic diffusion process

When the optimal probability density function  $\hat{p}(\mathbf{x}, t)$  is estimated precisely, by means of the Nadaraya-Watson estimator in Eq. (14), we can obtain the KM coefficients of arbitrary order for the potential dynamical system (1). The KM coefficients are important to characterize the governing equations of motion for potential dynamical models, especially for general stochastic diffusion processes, i.e., processes in which the system (1) without admixture of random jump events.

$$\begin{aligned} d\mathbf{x}(t) &= \mathbf{a}(\mathbf{x}, t)dt + \mathbf{b}(\mathbf{x}, t)d\mathbf{W}(t), \\ \mathbf{x}(0) &= \mathbf{x}_0, \quad t \in [0, T]. \end{aligned} \quad (15)$$

According to the Pawula theorem, the first two orders of KM coefficients of the diffusion process correspond exactly to the drift and quadratic diffusion coefficients of its governing equations. For only the differential order of the  $k$ th component  $x_k$  ( $k = 1, \dots, d$ ) of the state vector  $\mathbf{x}$  can be taken to the set of positive integers, and the differential orders of all other components are 0. Thus, there are

$$\begin{aligned} D^{(v_k=1, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle dx_k(t) \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle [x_k(t + \tau) - x_k(t)]^1 \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left[ a_k \tau + \sum_{i=1}^s b_{k,i} \langle dW_i \rangle \right] \\ &= a_k, \end{aligned} \quad (16)$$

$$\begin{aligned} D^{(v_k=2, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle dx_k^2(t) \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle [x_k(t + \tau) - x_k(t)]^2 \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left\langle \left[ a_k \tau + \sum_{i=1}^s b_{k,i} dW_i \right]^2 \right\rangle \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left\langle a_k^2 \tau^2 + \left( \sum_{i=1}^s b_{k,i} dW_i \right)^2 \right\rangle \\ &\quad + 2a_k \sum_{i=1}^s b_{k,i} \langle dW_i \rangle = \sum_{i=1}^s b_{k,i}^2, \end{aligned} \quad (17)$$

$$D^{(v_k \geq 3, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) = 0. \quad (18)$$

### B. Case II: Stochastic jump-diffusion process

Similarly, for the multidimensional stochastic jump-diffusion system (1), the relationship between the corresponding KM coefficients and the main parameter terms of the

equation can be summarized as follows:

$$\begin{aligned} D^{(v_k=1, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle dx_k(t) \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle [x_k(t + \tau) - x_k(t)]^1 \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left[ a_k \tau + \sum_{i=1}^s b_{k,i} \langle dW_i \rangle + \sum_{j=1}^l \langle c_{k,j} \rangle \langle dJ_j \rangle \right] \\ &= a_k, \end{aligned} \quad (19)$$

$$\begin{aligned} D^{(v_k=2, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle dx_k^2(t) \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle [x_k(t + \tau) - x_k(t)]^2 \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left\langle a_k^2 \tau^2 + \left( \sum_{i=1}^s b_{k,i} dW_i \right)^2 + \left( \sum_{j=1}^l c_{k,j} dJ_j \right)^2 \right\rangle \\ &\quad + 2a_k \sum_{i=1}^s b_{k,i} \langle dW_i \rangle + 2a_k \left\langle \sum_{j=1}^l c_{k,j} dJ_j \right\rangle \\ &\quad + 2 \left\langle \left( \sum_{i=1}^s b_{k,i} dW_i \right) \left( \sum_{j=1}^l c_{k,j} dJ_j \right) \right\rangle \\ &= \sum_{i=1}^s b_{k,i}^2 + \sum_{j=1}^l \sigma_{c_{k,j}}^2 \lambda_j, \end{aligned} \quad (20)$$

$$\begin{aligned} D^{(v_k=2n, n>1, v_{k'}=0(k' \neq k))}(\mathbf{x}, t) &= \langle [dx_k^{2n}] \rangle |_{\mathbf{x}(t)=\mathbf{x}} \\ &= \sum_{j=1}^l \frac{2n!}{2^n (n!)} (\sigma_{c_{k,j}}^2)^n \lambda_j. \end{aligned} \quad (21)$$

According to the derivation of Eq. (21), for the special hypothesis  $l = 1$ , i.e., the case where the  $j$ th component  $x_j$  of the state vector  $\mathbf{x}$  is excited by only one corresponding jump term  $dJ_j$ . The variance  $\sigma_{c_{k,j}}^2$  of the jump amplitude parameter obeying the Gaussian distribution and the jump rate  $\lambda_j$  of the Poisson jump term can be given by

$$\begin{aligned} \sigma_{c_{k,j}}^2 &= \frac{D^{(v_k=6, v_{k'}=0(k' \neq k))}(\mathbf{x}, t)}{5D^{(v_k=4, v_{k'}=0(k' \neq k))}(\mathbf{x}, t)}, \\ \lambda_j &= \frac{D^{(v_k=4, v_{k'}=0(k' \neq k))}(\mathbf{x}, t)}{3\sigma_{c_{k,j}}^4}. \end{aligned} \quad (22)$$

Furthermore, for a very small timescale  $\tau$ , it is allowed to construct a function  $Q(\mathbf{x})$  for distinguishing the jumpy and purely diffusive behavior of the stochastic process through the higher order conditional moments obtained in the sampled time series  $\mathbf{x}$ :

$$Q(\mathbf{x}) = \frac{K^{(v_k=6, v_{k'}=0(k' \neq k))}(\mathbf{x}, \tau)}{5K^{(v_k=4, v_{k'}=0(k' \neq k))}(\mathbf{x}, \tau)} \approx \begin{cases} b_{k,i}^2 \tau, & \text{diffusive,} \\ \sigma_{c_{k,j}}^2, & \text{jumpy,} \end{cases} \quad (23)$$

here  $b_{k,i}$  is also considered as the diffusion coefficient in the case when the  $k$ th state component is also perturbed by only one Gaussian white noise process.

When dealing with actual collected data, we cannot obtain infinite time resolution, which means that the above limit  $\tau \rightarrow 0$  is not possible. The best case is to analyze the smallest possible time difference. For stochastic diffusion processes, the KM coefficients have a theoretical value of 0 with differential order  $n \geq 3$ , but it is often found that these terms do not vanish due to finite time effects. In order to effectively reduce the error in the estimation of conditional moments, an arbitrary-order finite-time correction method is introduced, which is effective for both arbitrary Markovian diffusion processes and jump diffusion processes [48].

$$\begin{aligned}
D^{(1)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} M_1, \\
D^{(2)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} [M_2 - M_1^2], \\
D^{(3)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} [M_3 - 3M_1M_2 + 2M_1^3], \\
D^{(4)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} [M_4 - 4M_1M_3 - 3M_2^2 + 12M_1^2M_2 - 6M_1^4], \\
D^{(5)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} [M_5 - 5M_1M_4 - 10M_2M_3 + 30M_1M_2^2 \\
&\quad + 20M_1^2M_3 - 60M_1^3M_2 + 24M_1^5], \\
D^{(6)} &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} [M_6 - 6M_1M_5 - 10M_3^2 - 15M_2M_4 \\
&\quad + 30M_2^3 + 120M_1M_2M_3 + 30M_1^2M_4 - 270M_1^2M_2^2 \\
&\quad - 120M_1^3M_3 + 360M_1^4M_2 - 120M_1^6], \quad (24)
\end{aligned}$$

where  $M$  denotes the conditional moment and its subscript denotes the differential order.

However, the abstract mathematical symbolic relationship between the KM coefficients and the state variables is the core of the inverse problem of identifying the underlying dynamics from the data, rather than the simple data results of the statistical measures. The development and evolution of the SINDY idea cannot be ignored as a vital help in achieving this goal, so a similar structure is introduced in the proposed computational framework. Therefore, the optimization problem can be expressed as

$$\tilde{\Xi} = \arg \min_{\Xi \in \mathbb{R}^m} \frac{1}{2N_s} \|D^{(v)} - \Theta(\mathbf{x})\Xi\|_2^2 + R(\Xi), \quad (25)$$

where  $m$  indicates the number of function elements in the assumed basis function library;  $N_s$  denotes the number of input data;  $\Theta(\mathbf{x})$  is the library of candidate basis functions that may contain polynomials, exponential functions, and trigonometric functions;  $\Xi$  is the sparse coefficient matrix composed of sparse vectors corresponding to each function element in the library of basis functions; and  $R(\Xi)$  is a regularization factor used to improve the sparsity of the solution of the model. In this paper, it is appropriately considered as regularized in the  $l_1$ -norm sense, and then the optimization model becomes a Lasso regression.

Based on the relationship between the KM coefficients  $D^{(v)}(\mathbf{x}, t)$  ( $\mathbf{v} = \{v_1, \dots, v_d\}$ ) and the state vector  $\mathbf{x}$  calculated by Eq. (14), the trend of the evolution of the KM coefficients with the state variables can be initially determined, which can

provide useful prior information for the selection of the library of basis functions in the framework of the symbolic regression algorithm and can reduce the computational complexity in the iterative process. The above symbolic regression idea can give an explicit mathematical expression of the KM coefficients with the state vector  $\mathbf{x}$ . Therefore, the underlying dynamics of the stochastic system can be reconstructed by relating the KM coefficients to the drift, diffusion coefficients and the jump parameters of the system in the framework of the Itô stochastic differential equation.

We give the exact calculation steps in the pseudocode Algorithm 1, and the concise flowchart is in Fig. 1.

---

#### Algorithm 1 The Kernel-based system identification algorithm.

---

- Input:** The data  $\mathbf{x}(t)$  observed for a multivariate stochastic dynamical system with a similar structure to Eqs. (1) or (15) under the Itô interpretation, and a certain timescale  $\tau$ .
- Output:** The drift function  $a(\mathbf{x}, t)$ , diffusion term  $b(\mathbf{x}, t)$ , variance  $\sigma_c^2$  of the jump amplitude parameter, and jump rate  $\lambda$  of the jump term for the stochastic dynamical system under the Itô interpretation.
- 1: Use the self-consistent KDE (11)-(13) to estimate the joint probability distribution of the historical time of the observed data of the system.
  - 2: Calculate the higher-order KM coefficients  $D^{(v)}(\mathbf{x}, t)$  of the system according to the Nadaraya-Watson estimation model (14).
  - 3: Divide the conditional moments  $K^{(v)}$  by the very small timescale  $\tau$  to obtain the KM coefficients  $D^{(v)}$  and correct them appropriately using (24).
  - 4: Use Eq. (23) to determine the type of potential stochastic system for the time-series, while the variance  $\sigma_c^2$  of the distribution of the jump magnitude coefficients and the jump rate  $\lambda$  of the jump excitation are given by Eq. (22).
  - 5: Select the appropriate candidate function library  $\Theta(\mathbf{x})$  based on the known data mapping relationship  $(\mathbf{x}, D^{(v)}(\mathbf{x}, t))$ .
  - 6: **for** epoch:  $i = 1 \rightarrow m$  **do**
  - 7:      $\tilde{\Xi} = \arg \min_{\Xi \in \mathbb{R}^m} \frac{1}{2N_s} \|D^{(v)} - \Theta(\mathbf{x})\Xi\|_2^2 + R(\Xi)$ ,
  - 8:     Set a small threshold  $\epsilon$
  - 9:     If  $|\tilde{\Xi}_j| \leq \epsilon$ , set  $\tilde{\Xi}_j = 0$ , ( $j = 1, \dots, i$ ).
  - 10: **return** the optimal sparse matrix  $\tilde{\Xi}$ .
- 

## IV. NUMERICAL EXPERIMENTS

The time series sampled in the real world show different types of stationary or nonstationary dynamic characteristics due to local randomness. Nonstationary dynamic problems caused by random jump events have been widely involved in option pricing [49], fluid mechanics [50], energy conversion [51], and other fields. In this paper, a nonparametric approach is applied to separate the deterministic drift and diffusion coefficients from the data of the stochastic jump process, and the parameters of the jump intensity and the distribution of the jump components are effectively identified. In order to test the superiority of the proposed algorithm, we compared it with *kramersmoyal* [52], an open-source package written in Python, which is derived from calculating KM coefficients using the traditional KDE method [53]. The experimental data

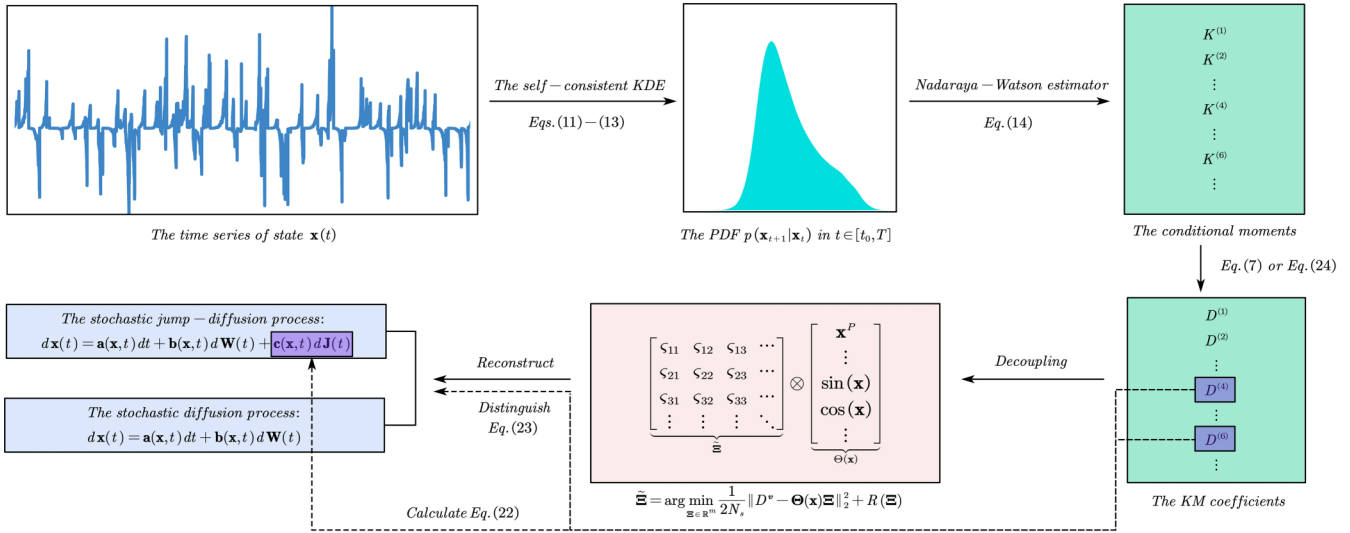


FIG. 1. The flow diagram of the proposed algorithmic framework.

for all numerical examples in this paper are obtained with the Euler-Mayurama integration scheme.

**A. One-dimensional example**

Consider a simple one-dimensional example for a jump-diffusion process with a nonlinear drift term, whose governing equations of motion can be defined as

$$dx_t = N(x, t)dt + \sqrt{G(x, t)}dW(t) + \xi dJ(t), \tag{26}$$

where the drift function  $N(x, t) = -0.5x^3 - 3$ ; the quadratic diffusion term, in the context of making the equation satisfy the Lipschitz condition and the linear growth condition, considers a representative nonlinear structure for harmonic excitation as  $G(x, t) = [\sin(x) + 5]^2$ ;  $W(t)$  is the Wiener process; the jump amplitude parameter  $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$ , where  $\sigma_\xi^2 = 0.8$ ; the jump rate of the Poisson jump term is  $\lambda = 1$ ; and the initial value  $x_0$  is a randomly specified real number from the standard normal distribution. The trajectory of the system with random jump term collected by using the Euler-Mayurama method in the time interval  $t \in [0, 4000]$  is on the timescale  $\tau = 0.001$ . For simplicity of illustration, the time series for  $t \in [0, 100]$  is given in Fig. 2. It can be seen that the fluctuation of the trajectory of the system with additional random jump events is more inhomogeneous and unstable than the general purely stochastic diffusion process.

(1) For the case of the stochastic diffusion process, Eq. (26) does not contain the jump term. Comparative results between the drift and diffusion coefficients calculated with the proposed method and those calculated with *kramersmoyal* are given in Fig. 3, respectively. It is clear that the *kramersmoyal* package developed based on the idea of traditional KDE has great volatility in the estimation of KM coefficients in the edge part, which is well handled by our method. The specific reason may mainly stems from the difficulty of traditional KDE to effectively capture the heavy-tailed probability distribution of the time series of the stochastic dynamical system.

(2) For the stochastic jump-diffusion process in the form as Eq. (26), obviously, the stochastic dynamical system

becomes more complex as the random excitation term increases. The variance  $\hat{\sigma}_\xi^2 = 0.7930$  of the distribution of the random jump amplitude parameter  $\xi$  can be estimated by Eq. (22), and the jump rate  $\lambda = 0.9683$  of the jump term can also be obtained. The coupled Poisson jump term is distinguished by Eq. (23), i.e., the system is not a purely stochastic diffusion process. Furthermore, based on Eqs. (19) and (20), we can derive the drift term  $\hat{N}(x, t) = D^{(1)}(x, t)$  and the diffusion term  $\sqrt{\hat{G}}(x, t) = \sqrt{D^{(2)}(x, t) - \hat{\sigma}_\xi^2 \hat{\lambda}}$ . Thus, the nonlinear drift and diffusion terms for system (26) estimated directly by Eq. (14) are shown in Fig. 4. As the complexity of the random perturbations increases, the shortcomings of *kramersmoyal* in clarifying the jump noise and accurately estimating the diffusion term become more apparent.

According to the data mapping relationships given by the drift and diffusion terms respectively in Fig. 4, we can appropriately select the candidate function library as  $\Theta(x) = \{1, x, x^2, x^3, \sin(x), \cos(x)\}$ . For demonstration purposes, the

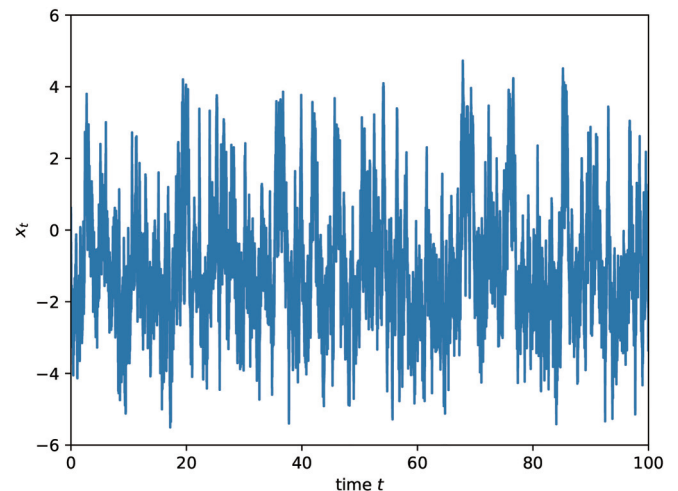


FIG. 2. The trajectory of the state variable of the stochastic jump-diffusion process (26) in the time interval  $t \in [0, 100]$ .

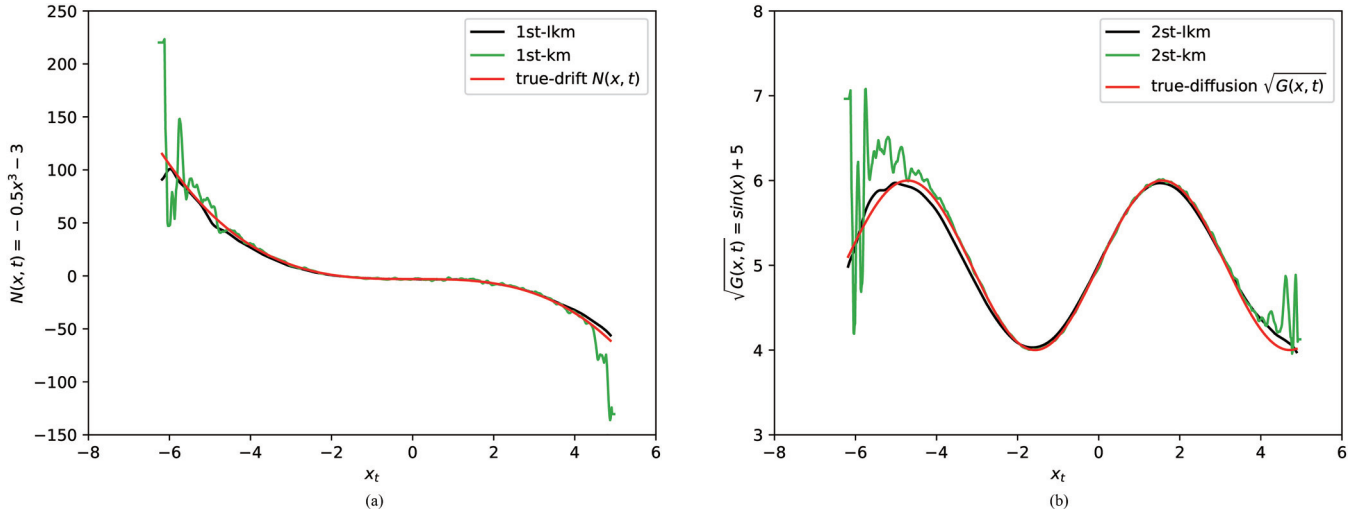


FIG. 3. For the case of the stochastic diffusion process, (a) true and estimated drift terms and (b) true and estimated diffusion terms. The green solid lines labels “1st-km” or “2st-km” are the results calculated with *kramersmoyal*, and the black solid lines named “1st-lkm” or “2st-lkm” are the results estimated with our improved method.

library here is relatively simple, and in fact, as long as the KM coefficients are estimated accurately, it is easy to give a suitable expression through a class of symbolic regression algorithms. In Table I we give mathematical explicit expressions for the drift and diffusion terms, respectively. In fact, the random jump events exacerbate the unsteadiness and randomness of the data samples, and for the stochastic diffusion model, the identification results of the proposed algorithm will only be better than those given in Table I.

### B. Two-dimensional example I

Simple systems in one dimension are not representative in real situations, although they are common. To further illustrate the validity and generalizability of the proposed model for multidimensional systems in practical applications, this subsection will consider a dynamical problem evolving from the background of the sound absorption process of the membrane-type acoustic metamaterial. Membrane-type acoustic metamaterials have received widespread attention from scholars in the fields of physical and acoustic engineering because of their good sound insulation, low mass, and low construction cost [54,55], and the coupled structural-acoustic response of the system can be explained by a combination

of structural and acoustic mode shapes [56]. Therefore, this example considers a new membrane-type acoustic metamaterial, and its specific structure is shown in Fig. 5. It is a coupled resonant system formed by a membrane-type acoustic metamaterial in a closed cavity under the action of magnetic poles and external plane waves. The model consists of two external magnets placed symmetrically and coaxially, a membrane with a centered magnetic-proof mass. The interaction between the magnets on both sides of the magnetic-proof mass and the external magnets creates an attractive force on both sides of the mass, and the mass is also subjected to air loading due to incident acoustic waves, which in turn resonates.

The law of motion of the membrane vibrating and bending with time under the combined action of the attractive force of the magnets and the incident acoustic waves can be described by the following differential equation [57]:

$$M\ddot{x}(t) + (2\zeta_i\omega_i M + \rho c\beta^2 S)\dot{x}(t) + \left(K + \frac{\rho c^2\beta^2 S}{D} + k_1\right)x(t) + k_3x^3(t) = 2\beta S p_{in}(t), \quad (27)$$

TABLE I. Identified drift and diffusion terms for one-dimensional system (26).

Basis $\Theta(x)$	$N(x, t)$		$\sqrt{G(x, t)}$	
	True	Learned	True	Learned
1	-3	-3.0049	5	4.9509
$x$	0	0	0	0
$x^2$	0	0	0	0
$x^3$	-0.5	-0.4592	0	0
$\sin(x)$	0	0	1	1.0151
$\cos(x)$	0	0	0	0

where  $\ddot{x}$ ,  $\dot{x}$ , and  $x$  denote the acceleration, velocity, and displacement variables of the membrane during vibration, respectively.  $M = M_i/\phi_{i,0}^2$ ,  $K = M\omega_i^2$ ,  $\beta = \langle\phi_i\rangle/\phi_{i,0}$ , where  $\phi_{i,0}$  and  $\langle\phi_i\rangle$  are the maximum and the surface averaged values of the  $i$ th modal shape, respectively.  $M_i$ ,  $\zeta_i$ , and  $\omega_i$  are the model mass, the damping ratio, and the intrinsic frequency of the  $i$ th modal shape, respectively.  $\rho$  and  $c$  are the density of air and the speed of sound, respectively.  $S$  is the surface area of the membrane, and  $D$  is the effective depth of the cavity.  $k_1$  and  $k_3$  are the linear and nonlinear stiffness coefficients derived from the magnetic force, respectively.  $p_{in}$  is regarded as the power spectrum of the acoustic wave incident on the absorber, which is used to describe the intensity of the acoustic wave.



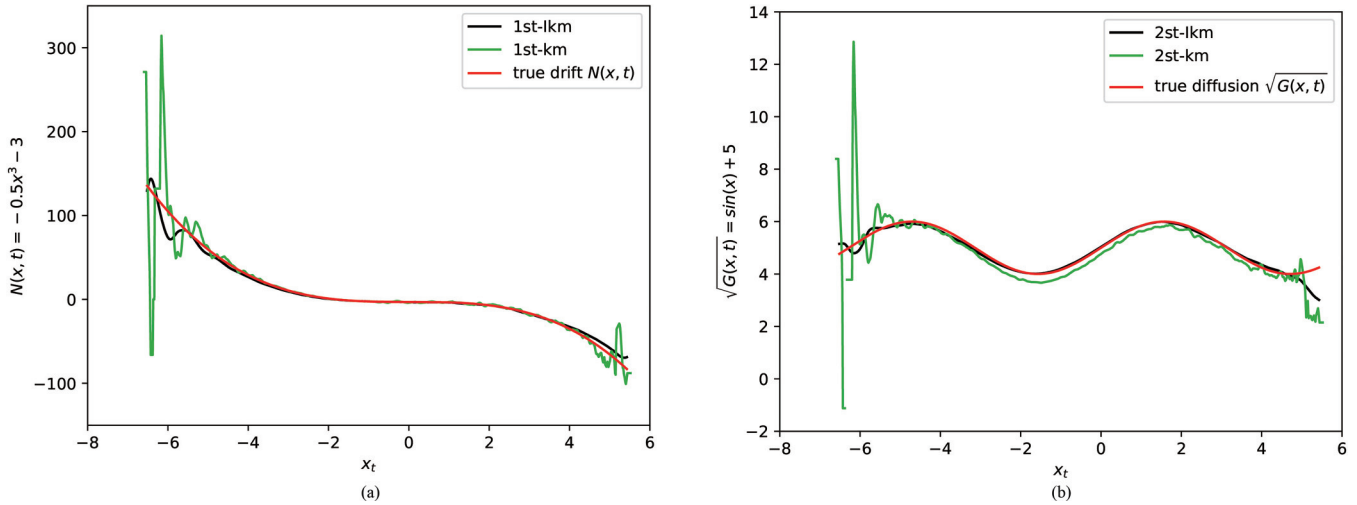


FIG. 4. For the case of the stochastic jump-diffusion process, (a) true and estimated drift terms and (b) true and estimated diffusion terms. The green solid lines named “1st-km” or “2st-km” are the results calculated by *kramersmoyal*, and the black solid lines named “1st-lkm” or “2st-lkm” are the results estimated by using our improved method.

To facilitate the numerical validation of the proposed model, consider its dimensionless equation as

$$\ddot{x}(t) + \alpha_1 x(t) + \alpha_2 \dot{x}(t) + \alpha_3 x^3(t) = \Gamma(t), \quad (28)$$

where  $\alpha_1$  and  $\alpha_2$  denote the stiffness and damping coefficients, respectively, and  $\alpha_3$  denotes the nonlinear stiffness coefficient. Assume that the coupling term  $\Gamma(t)$  between the power spectrum for the incident acoustic wave and the timescale in Eq. (28) is a hybrid noise structure, which consists of the standard Gaussian white noises  $W_t$  and Poisson jump processes  $J_t$ . In this example, it is practical to assume that the incident acoustic wave is simulated by Gaussian white noise and Poisson jump noise. The random jump process can enrich the acoustic wave cacophony making the analysis of the problem more realistic.

Defining  $y \triangleq dx/dt$  and organizing the terms, it can be further transformed into the following two-dimensional first-

order stochastic differential equation:

$$\begin{aligned} dx &= y dt, \\ dy &= (-\alpha_1 x - \alpha_3 x^3 - \alpha_2 y) dt + \eta dW_t + \xi dJ_t, \end{aligned} \quad (29)$$

where the drift coefficients are  $a_1 = y$  and  $a_2 = -\alpha_1 x - \alpha_3 x^3 - \alpha_2 y$ , respectively. The diffusion coefficients are  $b_1 = 0$  and  $b_2 = \eta$ , respectively.  $dW_t$  and  $dJ_t$  are independent increments of the standard Gaussian white noise and Poisson jump process that are independent of each other, respectively. Given the parameter values  $\alpha_1 = -1, \alpha_2 = 10, \alpha_3 = 3, \eta = 0.8, \xi \sim \mathcal{N}(0, \sigma_\xi^2 = 0.6)$ , and  $\lambda = 0.9$ . The time series collected using the Euler-Mayurama integration method in the time interval of  $[0, 9000]$  at a timescale of 0.001, with sample data size of  $9 \times 10^6$ . For  $t \in [0, 1000]$ , the phase diagram of system (29) is shown in Fig. 6(a), and the trajectories of the state displacement component  $x$  and the state velocity component  $y$  with time  $t$  are shown in Figs. 6(b) and 6(c), respectively.

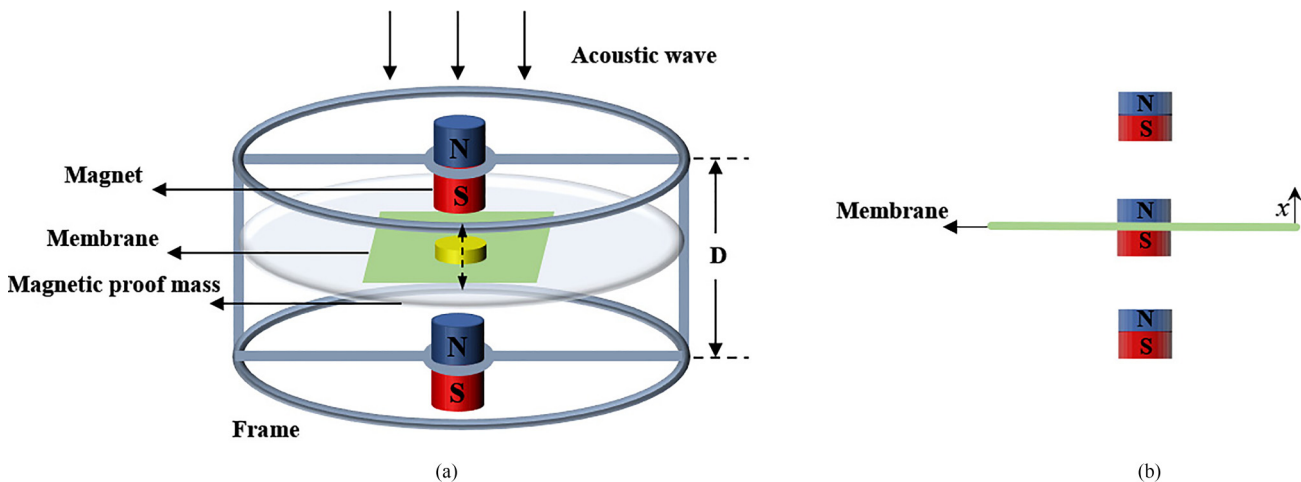


FIG. 5. Diagrams of the membrane-type acoustic metamaterial: (a) three-dimensional and (b) side plane.

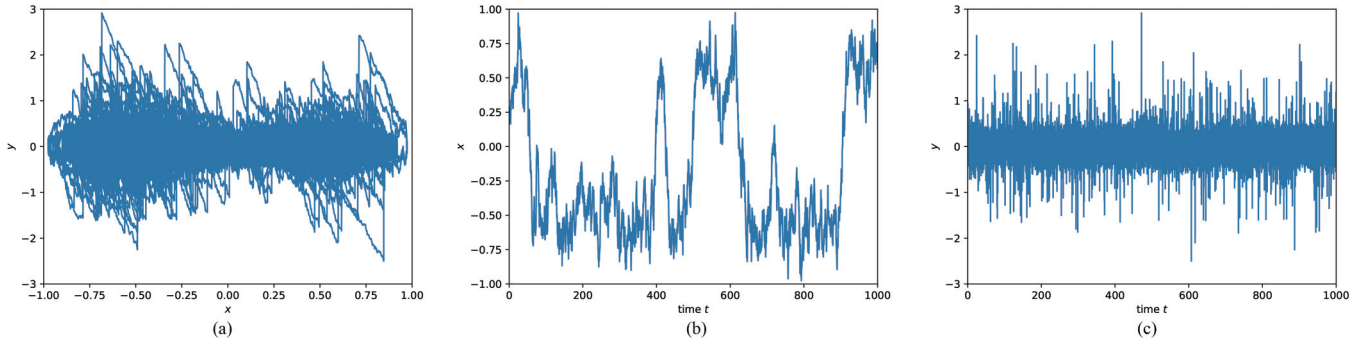


FIG. 6. (a) Phase diagram of the state variables of system (29). Trajectories of (b) the displacement component  $x$  and (c) the velocity component  $y$  over time.

According to Eq. (22), we can calculate the variance  $\hat{\sigma}_{\xi}^2 = 0.5269$  of the distribution of the jump magnitude coefficient  $\xi$  and the jump rate  $\hat{\lambda} = 0.9576$  with estimated higher order conditional moments  $\hat{D}^{(0,6)}$  and  $\hat{D}^{(0,4)}$ . Once the jump terms are clarified, it is easy to estimate the remaining drift and diffusion terms of the system based on Eqs. (19) and (20). The error thermograms of the drift and diffusion coefficients computed by our proposed framework and the *kramersmoyal* method with respect to the true values are given in Fig. 7, respectively. It is clear that the results estimated by the *kramersmoyal* method are not as good as those computed by our proposed method. Although the trajectories of the states in these edge regions appear rarely, they may contain important information about the jump phenomenon, which is important for the efficient identification of the stochastic jump-diffusion process. For the state vector  $[x, y]$  of system (29), it is appropriate to consider polynomials within three degrees of freedom (including their cross terms) as the basis func-

tion library  $\Theta(x) = \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$ . The results of the drift and diffusion terms estimated by the associated KM coefficients are shown in Table II. The two-dimensional strong nonlinear system is more complex, and the proposed system identification algorithm shows acceptable results.

### C. Two-dimensional example II

More practically, the system is coupled with multiple random noises of different intensities, including Gaussian white noise and Poisson jump noise. To illustrate the level of accuracy of the estimation and identification of the proposed method, on the basis of system (29), we proceed to consider a two-dimensional stochastic system excited by a mixture of three Gaussian white noises of different intensities and four

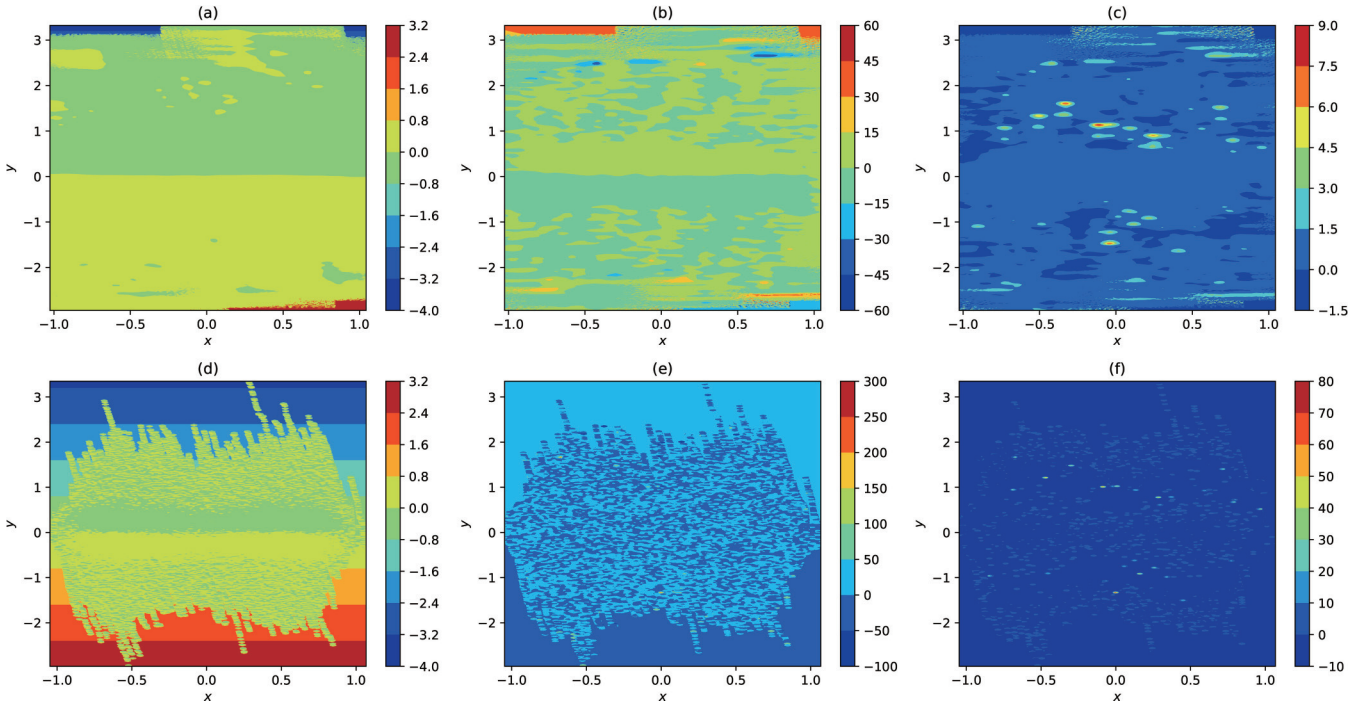


FIG. 7. Error between the solutions computed by our proposed framework and the true drift coefficients (a)  $a_1$ , (b)  $a_2$  and the true diffusion coefficient (c)  $b_2$ , respectively. The error between the solution calculated with *kramersmoyal* and the true drift coefficients (d)  $a_1$ , (e)  $a_2$ , and the true diffusion coefficient (f)  $b_2$ , respectively.

TABLE II. Identified drift and diffusion terms for two-dimensional system (29).

Basis $\Theta(x)$	$a_1(x, t)$		$a_2(x, t)$		$b_1(x, t)$		$b_2(x, t)$	
	True	Learned	True	Learned	True	Learned	True	Learned
1	0	0	0	0	0	0	0.8	0.7513
$x$	0	0	1	1.6261	0	0	0	0
$y$	1	0.9710	-10	-10.0640	0	0	0	0
$x^2$	0	0	0	0	0	0	0	0
$xy$	0	0	0	0	0	0	0	0
$y^2$	0	0	0	0	0	0	0	0
$x^3$	0	0	-3	-2.5097	0	0	0	0
$x^2y$	0	0	0	0	0	0	0	0
$xy^2$	0	0	0	0	0	0	0	0
$y^3$	0	0	0	0	0	0	0	0

jump processes with different jump rates:

$$d \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} * \begin{bmatrix} dW_1 \\ dW_2 \\ dW_3 \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix} * \begin{bmatrix} dJ_1 \\ dJ_2 \\ dJ_3 \\ dJ_4 \end{bmatrix}, \quad (30)$$

where  $*$  denotes the multiplication symbol of the matrix; the drift coefficients are  $a_1 = y$ ,  $a_2 = x - 3x^3 - 10y$ . The diffusion coefficient component vectors are  $\mathbf{b}_1 = [b_{11}, b_{12}, b_{13}] = [0.2, 0.4, 0.6]$  and  $\mathbf{b}_2 = [b_{21}, b_{22}, b_{23}] = [0.3, 0.5, 0.7]$ , respectively.  $dW_1, dW_2$ , and  $dW_3$  are Gaussian white noises with mean value of 0 and variances of 2, 4, and 6, respectively. The variances of the Gaussian distributions obeyed by the jump amplitude coefficients are  $\sigma_c^2 = [0.2, 0.4, 0.6, 0.8; 0.3, 0.5, 0.7, 0.9]$ . The jump rates  $\lambda = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]^T$  of the jump terms  $d\mathbf{J} = [dJ_1, dJ_2, dJ_3, dJ_4]^T$  are  $[0.2, 0.4, 0.6, 0.8]^T$ . For the time interval  $t \in [0, 9000]$ , the time series generated by the Euler-Mayurama method for system (30) on the timescale  $\tau = 0.001$  is shown in Fig. 8.

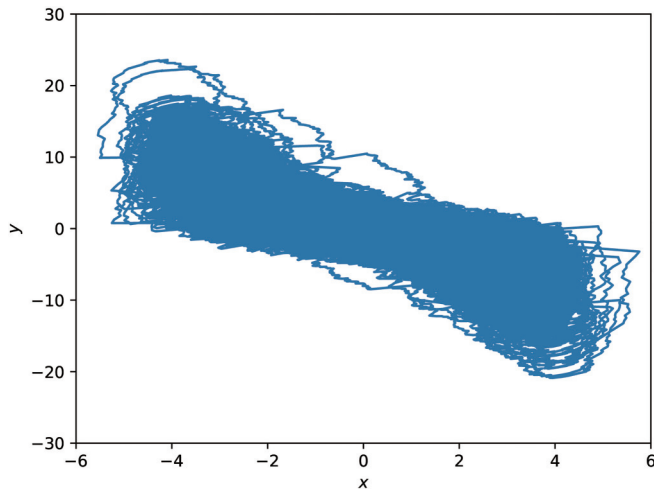


FIG. 8. Phase diagram of system (30) over time.

Although for the case of multiple noise mixing, we may not be able to effectively obtain the diffusion coefficients of the corresponding components and the statistical properties of each noise according to Eqs. (20) and (21), the corresponding drift coefficients can still be easily calculated with Eq. (19). The candidate function library is still selected as  $\Theta(x) = \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$ . Table III shows that the proposed algorithmic framework can achieve good identification of deterministic parts even for the multinoise system (30).

### V. CONCLUSION

Complex systems involving a large number of degrees of freedom in nature generally exhibit nonstationary dynamics. Such instabilities are usually accompanied by random jump processes, and separating the class of effects caused by random jump events from diffusion processes is an important issue for a detailed understanding of the stochastic dynamics of complex systems. In this paper we address this general problem by introducing a nonparametric method for KDE. By incorporating the KDE technique, which is improved by the idea of the Fourier transform, into the general KM coefficients of the stochastic jump-diffusion model, the drift and diffusion terms under the Itô interpretation are thus identified from observations whose physical laws are unknown. It provides the possibility for the symbolic regression algorithm to efficiently

TABLE III. Identified drift terms for two-dimensional system (30).

Basis $\Theta(x)$	$a_1(x, t)$		$a_2(x, t)$	
	True	Learned	True	Learned
1	0	0	0	0
$x$	0	0	1	1.2291
$y$	1	0.9859	-10	-9.9207
$x^2$	0	0	0	0
$xy$	0	0	0	0
$y^2$	0	0	0	0
$x^3$	0	0	-3	-2.9929
$x^2y$	0	0	0	0
$xy^2$	0	0	0	0
$y^3$	0	0	0	0

identify the stochastic dynamical systems by establishing the exact relationship between the higher-order KM coefficients and the main structural terms of the equations in Itô's framework. The only information relied upon in this paper is a time series of the observed system without any other prior assumptions. The proposed computational framework can effectively identify the potential governing equations not only for the stochastic jump-diffusion systems with multiple degrees of freedom in the form of Eq. (1), but also for arbitrary Markovian stochastic diffusion processes in the form of Eq. (15).

Comparisons with the method mentioned in Ref. [53] reflect the higher accuracy of the proposed framework. Moreover, the proposed computational framework has better accuracy of global estimation for computing KM coefficients. It can capture the influence exerted by sparse and random jump events more precisely and use it to improve the accu-

racy of the recognition of the system. In addition, It should be noted that the systems mentioned in this paper are all autonomous systems, i.e., the equations of the systems do not explicitly contain the time  $t$ . Similar inverse problems for nonautonomous systems that explicitly contain time  $t$  will be considered in the future.

The relevant code can be obtained by contacting the authors.

#### ACKNOWLEDGMENTS

This work was supported by the Natural Science Basic Research Program of Shaanxi Province (2023-JC-YB-063) and the Postgraduate Innovation Fund Project of Xi'an Polytechnic University (No. chx2022024).

- 
- [1] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhath, Deep learning and process understanding for data-driven Earth system science, *Nature (London)* **566**, 195 (2019).
  - [2] E. Kuhl, Data-driven modeling of COVID-19—Lessons learned, *Extreme Mech. Lett.* **40**, 100921 (2020).
  - [3] J. Zhang and W. Ma, Data-driven discovery of governing equations for fluid dynamics based on molecular simulation, *J. Fluid Mech.* **892**, A5 (2020).
  - [4] K. Wang, Y. Chen, M. Mehana, N. Lubbers, K. C. Bennett, Q. Kang, H. S. Viswanathan, and T. C. Germann, A physics-informed and hierarchically regularized data-driven model for predicting fluid flow through porous media, *J. Comput. Phys.* **443**, 110526 (2021).
  - [5] M. Rishhechi Fayyaz, M. R. Rasouli, and B. Amiri, A data-driven and network-aware approach for credit risk prediction in supply chain finance, *Ind. Manage. Data Syst.* **121**, 785 (2020).
  - [6] A. M. Avila and I. Mezić, Data-driven analysis and forecasting of highway traffic dynamics, *Nat. Commun.* **11**, 2090 (2020).
  - [7] K. M. Tolle, D. S. W. Tansley, and A. J. G. Hey, The fourth paradigm: Data-intensive scientific discovery [Point of View], *Proc. IEEE* **99**, 1334 (2011).
  - [8] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* **656**, 5 (2010).
  - [9] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, Sparsity-promoting dynamic mode decomposition, *Phys. Fluids* **26**, 024103 (2014).
  - [10] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (Society for Industrial and Applied Mathematics, Philadelphia, 2016).
  - [11] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition, *Exp. Fluids* **57**, 42 (2016).
  - [12] N. Takeishi, Y. Kawahara, and T. Yairi, Subspace dynamic mode decomposition for stochastic Koopman analysis, *Phys. Rev. E* **96**, 033310 (2017).
  - [13] W. T. Redman, Renormalization group as a Koopman operator, *Phys. Rev. E* **101**, 060104(R) (2020).
  - [14] S. Wang and Y. Lan, Probing the phase space of coupled oscillators with Koopman analysis, *Phys. Rev. E* **104**, 034211 (2021).
  - [15] Y. Zhen, B. Chapron, E. Memin, and L. Peng, Eigenvalues of autocovariance matrix: A practical method to identify the Koopman eigenfrequencies, *Phys. Rev. E* **105**, 034205 (2022).
  - [16] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.* **3**, 422 (2021).
  - [17] Z. Chen, Y. Liu, and H. Sun, Physics-informed learning of governing equations from scarce data, *Nat. Commun.* **12**, 6136 (2021).
  - [18] A. D. Jagtap, E. Kharazmi, and G. E. Karniadakis, Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems, *Comput. Methods Appl. Mech. Eng.* **365**, 113028 (2020).
  - [19] P. Ren, C. Rao, Y. Liu, J.-X. Wang, and H. Sun, PhyCRNet: Physics-informed convolutional-recurrent network for solving spatiotemporal PDEs, *Comput. Methods Appl. Mech. Eng.* **389**, 114399 (2022).
  - [20] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
  - [21] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci. USA* **116**, 22445 (2019).
  - [22] E. Kaiser, J. N. Kutz, and S. L. Brunton, Sparse identification of nonlinear dynamics for model predictive control in the low-data limit, *Proc. R. Soc. A* **474**, 20180335 (2018).
  - [23] K. Champion, S. L. Brunton, and J. N. Kutz, Discovery of nonlinear multiscale systems: Sampling strategies and embeddings, *SIAM J. Appl. Dyn. Syst.* **18**, 312 (2019).
  - [24] D. A. Messenger and D. M. Bortz, Weak SINDy for partial differential equations, *J. Comput. Phys.* **443**, 110525 (2021).
  - [25] M. Dai, T. Gao, Y. Lu, Y. Zheng, and J. Duan, Detecting the maximum likelihood transition path from data of stochastic dynamical systems, *Chaos* **30**, 113124 (2020).
  - [26] W. Wei, J. Hu, J. Chen, and J. Duan, Most probable transitions from metastable to oscillatory regimes in a carbon cycle system, *Chaos* **31**, 121102 (2021).

- [27] P. Zhou, S. Wang, T. Li, and Q. Nie, Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics, *Nat. Commun.* **12**, 5609 (2021).
- [28] T. D. Frank, Kramers–Moyal expansion for stochastic differential equations with single and multiple delays: Applications to financial physics and neurophysics, *Phys. Lett. A* **360**, 552 (2007).
- [29] Y. Lu, Y. Li, and J. Duan, Extracting stochastic governing laws by nonlocal Kramers–Moyal formulas, *Philos. Trans. R. Soc. A* **380**, 20210195 (2022).
- [30] M. R. R. Tabar, Kramers–Moyal expansion and Fokker–Planck equation, in *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems* (Springer International Publishing, Cham, 2019), pp. 19–29.
- [31] M. R. R. Tabar, The Kramers–Moyal coefficients of non-stationary time series and in the presence of microstructure (measurement) noise, in *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems* (Springer International Publishing, Cham, 2019), pp. 181–189.
- [32] M. P. Wand and M. C. Jones, *Kernel Smoothing* (Springer US, Boston, 1995).
- [33] A. Gramacki, *Nonparametric Kernel Density Estimation and Its Computational Aspects*, Studies in Big Data, Vol. 37 (Springer International Publishing, Cham, 2018).
- [34] J. E. Chacón and T. Duong, *Multivariate Kernel Smoothing and Its Applications* (Chapman and Hall/CRC, New York, 2018).
- [35] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, Kernel density estimation via diffusion, *Ann. Statist.* **38**, 2916 (2010).
- [36] J.-C. Hua, F. Noorian, D. Moss, P. H. W. Leong, and G. H. Gunaratne, High-dimensional time series prediction using kernel-based Koopman mode regression, *Nonlinear Dyn.* **90**, 1785 (2017).
- [37] O. Lindenbaum, A. Sagiv, G. Mishne, and R. Talmon, Kernel-based parameter estimation of dynamical systems with unknown observation functions, *Chaos* **31**, 043118 (2021).
- [38] B. Hamzi and H. Owhadi, Learning dynamical systems from data: A simple cross-validation perspective, Part I: Parametric kernel flows, *Physica D* **421**, 132817 (2021).
- [39] M. P. Wand, Fast computation of multivariate kernel estimators, *J. Comput. Graphical Stat.* **3**, 433 (1994).
- [40] A. Gramacki and J. Gramacki, FFT-based fast bandwidth selector for multivariate kernel density estimation, *Comput. Stat. Data Anal.* **106**, 27 (2017).
- [41] A. Bernacchia and S. Pigolotti, Self-consistent method for density estimation: Density estimation, *J. R. Stat. Soc. B* **73**, 407 (2011).
- [42] T. A. O’Brien, W. D. Collins, S. A. Rauscher, and T. D. Ringler, Reducing the computational cost of the ECF using a nuFFT: A fast and objective probability density estimation method, *Comput. Stat. Data Anal.* **79**, 222 (2014).
- [43] T. A. O’Brien, K. Kashinath, N. R. Cavanaugh, W. D. Collins, and J. P. O’Brien, A fast and objective multidimensional Kernel density estimation method: FastKDE, *Comput. Stat. Data Anal.* **101**, 148 (2016).
- [44] P. Przybyłowicz and M. Szölgyenyi, Existence, uniqueness, and approximation of solutions of jump-diffusion SDEs with discontinuous drift, *Appl. Math. Comput.* **403**, 126191 (2021).
- [45] M. R. R. Tabar, Influence of finite time step in estimating of the Kramers–Moyal coefficients, in *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems* (Springer International Publishing, Cham, 2019), pp. 191–205.
- [46] M. Anvari, M. R. R. Tabar, J. Peinke, and K. Lehnertz, Disentangling the stochastic behavior of complex time series, *Sci. Rep.* **6**, 35435 (2016).
- [47] L. Greengard and J.-Y. Lee, Accelerating the nonuniform fast Fourier transform, *SIAM Rev.* **46**, 443 (2004).
- [48] L. R. Gorjão, D. Witthaut, K. Lehnertz, and P. G. Lind, Arbitrary-order finite-time corrections for the Kramers–Moyal operator, *Entropy* **23**, 517 (2021).
- [49] M. Shirzadi, M. Rostami, M. Dehghan, and X. Li, American options pricing under regime-switching jump-diffusion models with meshfree finite point method, *Chaos Solitons Fractals* **166**, 112919 (2023).
- [50] Y.-H. Chen, C.-X. Xiao, H. Li, E. Fratini, P. Baglioni, and S.-H. Chen, Water dynamics in C–S–H and M–S–H cement pastes: A revised jump-diffusion and rotation-diffusion model, *Phys. B: Condens. Matter* **627**, 413542 (2022).
- [51] X. Song, P. Johnson, and P. Duck, A novel combination of Mycielski–Markov, regime switching and jump diffusion models for solar energy, *Appl. Energy* **301**, 117457 (2021).
- [52] Rydin, KramersMoyal, <https://github.com/LRydin/KramersMoyal> (2023).
- [53] L. R. Gorjão and F. Meirinhos, kramersmoyal: Kramers–Moyal coefficients for stochastic processes, *J. Open Source Soft.* **4**, 1693 (2019).
- [54] Z. Yang, J. Mei, M. Yang, N. H. Chan, and P. Sheng, Membrane-Type Acoustic Metamaterial with Negative Dynamic Mass, *Phys. Rev. Lett.* **101**, 204301 (2008).
- [55] X. Li, J. Zhao, W. Wang, T. Xing, L. Zhu, Y. Liu, and X. Li, Tunable acoustic insulation characteristics of membrane-type acoustic metamaterials array with compact magnets, *Appl. Acoust.* **187**, 108514 (2022).
- [56] F. Bucciarelli and M. Meo, Broadening sound absorption coefficient with hybrid resonances, *Appl. Acoust.* **160**, 107136 (2020).
- [57] X. Li, T. Xing, J. Zhao, and X. Gai, Broadband low frequency sound absorption using a monostable acoustic metamaterial, *J. Acoust. Soc. Am.* **147**, EL113 (2020).