Extracting Governing Laws from Sample Path Data of Non-Gaussian Stochastic Dynamical Systems

Yang Li¹ · Jinqiao Duan²

Received: 21 July 2021 / Accepted: 1 January 2022 / Published online: 21 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Advances in data science are leading to new progresses in the analysis and understanding of complex dynamics for systems with experimental and observational data. With numerous physical phenomena exhibiting bursting, flights, hopping, and intermittent features, stochastic differential equations with non-Gaussian Lévy noise are suitable to model these systems. Thus it is desirable and essential to infer such equations from available data to reasonably predict dynamical behaviors. In this work, we consider a data-driven method to extract stochastic dynamical systems with non-Gaussian asymmetric (rather than the symmetric) Lévy process, as well as Gaussian Brownian motion. We establish a theoretical framework and design a numerical algorithm to compute the asymmetric Lévy jump measure, drift and diffusion (i.e., nonlocal Kramers–Moyal formulas), hence obtaining the stochastic governing law, from noisy data. Numerical experiments on several prototypical examples confirm the efficacy and accuracy of this method. This method will become an effective tool in discovering the governing laws from available data sets and in understanding the mechanisms underlying complex random phenomena.

Keywords Nonlocal Kramers–Moyal formulas · Non-Gaussian Lévy noise · Stochastic dynamical systems · Heavy-tailed fluctuations · Rare events

Mathematics Subject Classification 60G51 · 60H10 · 65C20

1 Introduction

The extraction of physical laws from experimental or observable data is crucial to scientific and engineering applications where governing equations are unknown [1,2]. Due to...
the deficiency of scientific understanding of mechanisms underlying complex phenomena, it is sometimes not feasible to derive the explicit governing laws directly. These physical governing laws are often in the form of ordinary, partial or stochastic differential equations.

Therefore, there are recent machine learning methods devoted to discovering governing laws of nonlinear phenomena from noisy data. Some of these works are based on Kramers–Moyal formula [3–5] or Koopman generator [6–10]. Other researchers have developed a data-driven method called Sparse Identification of Nonlinear Dynamics, to learn ordinary [11], partial [12–15] and stochastic [3] differential equations from available data sets.

These techniques only focus on extracting either deterministic differential equations, or stochastic differential equations with Gaussian noise. However, there are numerous systems involving random bursting, flights, intermittent, hopping, or rare transition features in, for example, statistical physics [16], climate change [17], gene regulation [18], ecology [19], and geophysical turbulence [20]. In consideration of their jump character, these systems are suitable to be modeled as by stochastic differential equations with (non-Gaussian) Lévy processes rather than with Gaussian fluctuations alone. For instance, Böttcher [21] presented a simple construction method for a class of stochastic processes based on state space dependent mixing of Lévy processes. Ditlevsen revealed that the climate change system may be modeled as stochastic differential equations with Lévy process and Brownian motion, in the context of the Greenland ice core measurement data [17]. Based on this assertion, Zheng et al. developed a probabilistic framework to investigate the maximum likelihood climate change for an energy balance system under the combined influence of greenhouse effect and non-Gaussian $\alpha$-stable Lévy motions [22]. In the form of Lévy flights, Kharcheva et al. investigated the steady-state correlation characteristics of superdiffusion in one-dimensional confinement potential profiles [23]. Some researchers and we studied the escape phenomena [24–26] and stochastic resonance of neuron models [27] or population dynamics [28] driven by the non-Gaussian noise to detect its excitation behaviors. The non-Gaussian Lévy motions are also used to characterize random fluctuations in gene networks [29–31], current-biased long Josephson junctions [32–35], molecular motor [36] and other scientific fields [37–40].

Recently, we devised a novel data-driven approach to extract stochastic governing equations with (idealized) symmetric non-Gaussian Lévy motion from data [41], inspired by the deduction of integrodifferential Chapman–Kolmogorov equation by Gardiner [42]. In this present work, we present this technique to the stochastic dynamical systems with general asymmetric non-Gaussian Lévy noise. Numerical experiments for prototypical examples verify its effectiveness.

This work is arranged as follows. In Sect. 2, we derive formulas in the form of a theorem and a corollary to express the Lévy jump measure, drift coefficient, and diffusion coefficient, in terms of either the transition probability density or sample paths. These may be regarded as nonlocal Kramers–Moyal formulas, in contrast to the usual (local) Kramers–Moyal formulas as in [5, Ch. 3]. Then we design a numerical algorithm to compute the jump measure, drift coefficient and diffusion coefficient, hence obtaining the stochastic governing law, in Sect. 3. We test our method by numerical experiments in Sect. 4, and finally we conclude with Discussion in Sect. 5.

## 2 Method

In the previous section, we realize that random fluctuations often have both Gaussian and non-Gaussian statistical features. By Lévy–Itô decomposition theorem [43,44], a large class
of random fluctuations are indeed modeled as linear combinations of a (Gaussian) Brownian motion \( \mathbf{B}_t \) and a (non-Gaussian) Lévy process \( \mathbf{L}_t \). We thus consider an \( n \)-dimensional stochastic dynamical system in the following form

\[
\mathrm{d}\mathbf{x}(t) = \mathbf{b}(\mathbf{x}(t)) \, \mathrm{d}t + \Lambda \left( \mathbf{x}(t) \right) \, \mathrm{d}\mathbf{B}_t + \sigma \, \mathrm{d}\mathbf{L}_t, \tag{1}
\]

where \( \mathbf{B}_t = [B_{1,t}, \ldots, B_{n,t}]^T \) is an \( n \)-dimensional Brownian motion, and \( \mathbf{L}_t = [L_{1,t}, \ldots, L_{n,t}]^T \) is an \( n \)-dimensional non-Gaussian Lévy process with independent components described in Appendix A. The vector \( \mathbf{b}(\mathbf{x}) = [b_1(\mathbf{x}), \ldots, b_n(\mathbf{x})]^T \) is the drift coefficient (or vector field) in \( \mathbb{R}^n \) and \( \Lambda(\mathbf{x}) \) is an \( n \times n \) matrix. The diffusion matrix is defined by \( \sigma(\mathbf{x}) = \Lambda^T \). We take \( \sigma \) as a diagonal matrix with the positive diagonal component \( \sigma_i \), indicating the noise intensity of the corresponding component of Lévy process. Assume that the initial condition is \( \mathbf{x}(0) = \mathbf{z} \) and the jump measure of \( L_{i,t} \) is \( \nu_i(\mathrm{d}\xi) = W_i(\xi) \, \mathrm{d}\xi \), with kernel \( W_i \), for \( \xi \in \mathbb{R} \setminus \{0\} \).

For the sake of concreteness, we consider a special but significant Lévy process due to its extensive physical applications [16], the so-called \( \alpha \)-stable Lévy motion [45,46], as an example to illustrate our method. Its detailed information is present at Appendix A. In this case, the kernel function \( W^{\alpha,\beta}(\xi) \) has the following form

\[
W^{\alpha,\beta}(\xi) = \begin{cases} 
\frac{k_\alpha (1+\beta)}{2|\xi|^{1+\beta}}, & \xi > 0, \\
\frac{k_\alpha (1-\beta)}{2|\xi|^{1-\beta}}, & \xi < 0,
\end{cases}
\]

where \( \alpha \) is the stability parameter (or non-Gaussianity index) and

\[
k_\alpha = \begin{cases} 
\frac{\alpha(1-\alpha)}{T(2-\alpha) \cos(\pi \alpha/2)}, & \alpha \neq 1, \\
\frac{\alpha}{2\pi}, & \alpha = 1.
\end{cases}
\]

The skewness parameter \( \beta \) dominates the symmetry of Lévy motion, as shown in Fig. 1.

According to Refs. [43,47], the Fokker–Planck equation for the probability density function \( p(\mathbf{x}, t|\mathbf{z}, 0) \) for the solution of Eq. (1) is

\[
\frac{\partial p}{\partial t} = -\sum_{i=1}^n \frac{\partial}{\partial x_i} \left[ b_i p(\mathbf{x}, t|\mathbf{z}, 0) \right] + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} \left[ a_{ij} p(\mathbf{x}, t|\mathbf{z}, 0) \right] \\
- \sum_{i=1}^n \int_{\mathbb{R} \setminus \{0\}} \left[ p(\mathbf{x}, t|\mathbf{z}, 0) - p(\mathbf{x} - \sigma_i \chi_i^{\alpha}(y_i) y_i, t|\mathbf{z}, 0) \right] W_i^{\alpha,\beta}(y_i) \, \mathrm{d}y_i, \tag{2}
\]

where

\[
\chi_i^{\alpha}(y_i) = \begin{cases} 
0, & 0 < \alpha < 1, \\
\chi_{|y_i|<\sigma_i^{-1}}, & \alpha = 1, \\
1, & 1 < \alpha < 2,
\end{cases}
\]

which is the same as Eq. (22) in Ref. [48]. Recall here that \( \epsilon_i \)'s form the standard basis in \( \mathbb{R}^n \), and \( \chi_{|y_i|<\sigma_i^{-1}} \) denotes the indicator function. The initial condition is \( p(\mathbf{x}, 0|\mathbf{z}, 0) = \delta(\mathbf{x} - \mathbf{z}) \).

Here the integral in the right hand side is understood as a Cauchy principal value integral, which can be also expressed as a fractional differentiation [43,48]. Thus it is a fractional Fokker–Planck equation.
In order to discover stochastic dynamical systems with non-Gaussian Lévy noise as well as Gaussian noise from data, we derive the following theorem. It expresses jump measure, drift and diffusion in terms of the solution of Fokker–Planck equation.

In what follows, the marginal probability distribution in $i$th direction is denoted as $p_i (x_i, t | z_i, 0)$ and the set $\Gamma$ represents an $n$-dimensional cube $[-\epsilon, \epsilon]^n$.

**Theorem 1** (Relation between stochastic governing law and Fokker–Planck equation)

For every $\epsilon > 0$, the probability density function $p (x, t | z, 0)$ and the jump measure, drift and diffusion have the following relations:

1. For every $x_i$ and $z_i$ satisfying $|x_i - z_i| > \epsilon$ and $i = 1, 2, \ldots, n$,
   \[
   \lim_{t \to 0} t^{-1} p_i (x_i, t | z_i, 0) = \sigma_i^{-1} W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} (x_i - z_i) \right)
   \]
   uniformly in $x_i$ and $z_i$.

2. For $i = 1, 2, \ldots, n$,
   \[
   \lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) \ p \ (x, t | z, 0) \ dx = b_i (z) + R_{i}^{\alpha, \beta} (\epsilon),
   \]
   where $R_{i}^{\alpha, \beta} (\epsilon) =$
   \[
   \begin{cases}
   \sigma_i^{-1} \int_{-\epsilon}^{\epsilon} y_i W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i, & \alpha < 1, \\
   \sigma_i^{-1} \left[ \int_{-\epsilon}^{\epsilon} y_i W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i + \int_{1}^{\epsilon} y_i W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i \right], & \alpha = 1, \\
   -\sigma_i^{-1} \left[ \int_{-\epsilon}^{\epsilon} y_i W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i + \int_{-\epsilon}^{1} y_i W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i \right], & \alpha > 1.
   \end{cases}
   \]

3. For $i, j = 1, 2, \ldots, n$,
   \[
   \lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \ p \ (x, t | z, 0) \ dx = a_{ij} (z) + S_{ij}^{\alpha, \beta} (\epsilon),
   \]
   where $S_{ij}^{\alpha, \beta} (\epsilon) =$
   \[
   \begin{cases}
   \sigma_i^{-1} \int_{-\epsilon}^{\epsilon} \ y_i^2 W_{i}^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i, & \text{and } S_{ij}^{\alpha, \beta} (\epsilon) = 0 \text{ for } i \neq j.
   \end{cases}
   \]

Note that for the purpose of the design of subsequent numerical scheme, we reformulate these relations in Theorem 1 in the following corollary. Thus, the jump measure, drift and diffusion...
diffusion are expressed in terms of the sample paths of the stochastic differential equation (1).

**Corollary 2** (Nonlocal Kramers–Moyal formulas)

For every $\varepsilon > 0$, the sample path solution $x(t)$ of the stochastic differential equation (1) and the jump measure, drift and diffusion have the following relations:

1. For every $c_1$ and $c_2$ satisfying $c_1 < c_2 < 0$ or $0 < c_1 < c_2$, and $i = 1, 2, \ldots, n$,

   $$
   \lim_{t \to 0} t^{-1} \mathbb{P} \{ x_i(t) - z_i \in [c_1, c_2] | x(0) = z \} = \sigma_i^{-1} \int_{c_1}^{c_2} W_i^{\alpha, \beta} \left( \sigma_i^{-1} y_i \right) dy_i;
   $$

2. For $i = 1, 2, \ldots, n$,

   $$
   \lim_{t \to 0} t^{-1} \mathbb{P} \{ x(t) - z \in \Gamma | x(0) = z \} \cdot E \left[ (x_i(t) - z_i) | x(0) = z; x(t) - z \in \Gamma \right] = b_i(z) + R_i^{\alpha, \beta}(\varepsilon);
   $$

3. For $i, j = 1, 2, \ldots, n$,

   $$
   \lim_{t \to 0} t^{-1} \mathbb{P} \{ x(t) - z \in \Gamma | x(0) = z \} \cdot E \left[ (x_i(t) - z_i) (x_j(t) - z_j) | x(0) = z; x(t) - z \in \Gamma \right] = a_{ij}(z) + S_{ij}^{\alpha, \beta}(\varepsilon).
   $$

The three formulas in this corollary may be called the nonlocal Kramers–Moyal formulas, in contrast to the usual (local) Kramers–Moyal formulas as in [5, Ch. 3]. They express the jump measure, drift and diffusion in terms of the sample paths of the stochastic differential equation (1). In other words, with experimental or observational sample path data, we can extract the underlying stochastic differential equation.

The proofs of Theorem 1 and Corollary 2 are in Appendices B and C, respectively.

### 3 Numerical Algorithm

We now devise a numerical algorithm to extract a stochastic differential equation model as Eq. (1) from its sample path data. Assume that there exists a pair of data sets containing $M$ elements, respectively,

$$
Z = [z_1, z_2, \ldots, z_M],
$$

$$
X = [x_1, x_2, \ldots, x_M],
$$

where $x_j$ is the image of $z_j$ after a small evolution time $h$. In addition, we also choose a dictionary of basis functions $\Psi(x) = [\psi_1(x), \psi_2(x), \ldots, \psi_K(x)]$. With the data sets, the dictionary and the Corollary 2 in the previous section, we now propose the following algorithm to identify the kernel function and noise intensity of the Lévy motion, the drift coefficient and the diffusion matrix.

#### 3.1 Algorithm for Identification of the Lévy Motion

In order to determine the Lévy motion, we need to identify all the parameters $\alpha_i, \beta_i$ and $\sigma_i$ for $i = 1, 2, \ldots, n$. From the first assertion of Corollary 2, we see that the two sides of
the equation only depend on the difference of \( x_i \) and \( z_i \) rather than their specific positions. Thus we construct a new data set \( Y_i = [y_i^1, y_i^2, \ldots, y_i^M] \) with \( y_i = x_i - z_i \). Therefore, the probability in the left hand side of the first assertion can be approximated by the ratio of the number of the points falling into the interval \([c_1, c_2]\) to the total number \( M \) in \( Y_i \).

Specifically, we consider the integral on \( 2N + 2 \) intervals \([\varepsilon, m\varepsilon], [m\varepsilon, m^2\varepsilon], \ldots, [m^N\varepsilon, m^{N+1}\varepsilon] \) and \([-m\varepsilon, -\varepsilon], [-m^2\varepsilon, -m\varepsilon], \ldots, [-m^{N+1}\varepsilon, -m^N\varepsilon]\) with the positive integer \( N \), the positive real number \( \varepsilon \) and the real number \( m > 1 \). This is illustrated in Fig. 2. For the convenience of representations, we neglect the subscript “i” of \( \alpha_i, \beta_i \) and \( \sigma_i \) in the following equations. Assume that there are \( n_0^+, n_1^+, \ldots, n_N^+ \) and \( n_0^-, n_1^-, \ldots, n_N^- \) points from the data set \( Y_i \) which fall into these intervals respectively. Therefore,

\[
h^{-1} \mathbb{P} \left\{ x_i (h) - z_i \in [m^k \varepsilon, m^{k+1} \varepsilon] \right\} \approx h^{-1} M^{-1} n_k^+.\]

On the other hand, the integration from the right hand side yields

\[
\sigma^{-1} \int_{m^k \varepsilon}^{m^{k+1} \varepsilon} W_i^a \beta (\sigma^{-1} y_i) \, dy_i
\]

\[
= \sigma^a k_\alpha \frac{(1 + \beta)}{2} \int_{m^k \varepsilon}^{m^{k+1} \varepsilon} |y_i|^{-1+\alpha} \, dy_i
\]

\[
= \sigma^a k_\alpha (1 + \beta) \alpha^{-1} \varepsilon^{-\alpha} m^{-\alpha} (1 - m^{-\alpha})/2.
\]

Combining the two equations, we have \( N + 1 \) equalities

\[
\sigma^a k_\alpha (1 + \beta) \alpha^{-1} \varepsilon^{-\alpha} m^{-\alpha} (1 - m^{-\alpha})/2 = h^{-1} M^{-1} n_k^+, \quad k = 0, 1, \ldots, N. \tag{4}
\]

On the other \( N + 1 \) intervals, we also have

\[
\sigma^a k_\alpha (1 - \beta) \alpha^{-1} \varepsilon^{-\alpha} m^{-\alpha} (1 - m^{-\alpha})/2 = h^{-1} M^{-1} n_k^-, \quad k = 0, 1, \ldots, N. \tag{5}
\]

The summations of Eqs. (4) and (5) yield

\[
\sigma^a k_\alpha \alpha^{-1} \varepsilon^{-\alpha} m^{-\alpha} (1 - m^{-\alpha}) = h^{-1} M^{-1} (n_k^+ + n_k^-), \quad k = 0, 1, \ldots, N.
\]

The ratios of the first equation (\( k = 0 \)) to the other \( N \) equations lead to the solutions

\[
\alpha = (k \ln m)^{-1} \ln \frac{n_0^+ + n_0^-}{n_k^+ + n_k^-}, \quad k = 1, 2, \ldots, N. \tag{6}
\]

If \( N = 1 \), it is the optimal solution \( \tilde{\alpha} \) of the parameter \( \alpha \). In order to make full use of the data information, a bigger \( N \) can be selected to identify the optimal solution \( \tilde{\alpha} \) as the mean value of Eq. (6). Let \( \rho = \sum k n_k^- / \sum k n_k^+ \). Then the ratio of the summation of Eqs. (4) over the summation of Eqs. (5) yields

\[
\beta = \frac{1 - \rho}{1 + \rho}. \tag{7}
\]

Therefore, the noise intensity \( \sigma \) is computed as

\[
\sigma = \left[ \frac{\tilde{\alpha} \varepsilon^{\tilde{\alpha} m^{\tilde{\alpha}} (n_k^+ + n_k^-)} }{k_\alpha h M (1 - m^{-\tilde{\alpha}})} \right]^{1/\tilde{\alpha}}, \quad k = 0, 1, \ldots, N. \tag{8}
\]

Hence, the optimal noise intensity \( \tilde{\sigma} \) can be identified as the mean value of Eq. (8). By going through all \( n \) dimensions for \( i \), we obtain \( 3n \) parameters \( \alpha_i, \beta_i \) and \( \sigma_i \) to determine the Lévy jump measure as well as its noise intensity.
Fig. 2 Integration intervals for $\varepsilon = 1, m = 5$ and $N = 1$

Note that if we consider the symmetric Lévy motion, then $\rho = 1$ and $\beta = 0$ in Eq. (7). Let $n_k = n_k^+ + n_k^-$, thus Eqs. (6) and (8) recover the computations of the symmetric case. It is also important to emphasize that this method avoids the problem of curse of dimensionality since we can deal with every component independently. In other words, the increase of dimensionality does not necessarily require more data.

3.2 Algorithm for Identification of the Drift Term

In order to identify the drift coefficient $b(x)$, we approximate its every component in terms of the dictionary of basis functions $\Psi(x)$ as $b_i(x) \approx \sum_{k=1}^{K} c_{i,k} \psi_k(x)$, $i = 1, 2, \ldots, n$. According to the second assertion of Corollary 2, the computation of the drift term only requires the data within the cube $x-z \in \Gamma$. Therefore, after deleting the data outside the cube $\Gamma$ in the data sets $Z$ and $M$, we obtain the new data sets with $\hat{M}$ elements respectively

$$\hat{Z} = [\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_{\hat{M}}].$$

$$\hat{X} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{\hat{M}}].$$

Based on the second assertion of Corollary 2, we derive the following group of equations

$$A \tilde{c}_i = B_i,$$

$$A = \begin{bmatrix}
\psi_1(\hat{z}_1) & \cdots & \psi_K(\hat{z}_1) \\
\vdots & \ddots & \vdots \\
\psi_1(\hat{z}_{\hat{M}}) & \cdots & \psi_K(\hat{z}_{\hat{M}})
\end{bmatrix},$$

$$B_i = \hat{M} M^{-1} h^{-1} \left[ \hat{x}_{i,1} - \hat{z}_{i,1}, \ldots, \hat{x}_{i,\hat{M}} - \hat{z}_{i,\hat{M}} \right]^T - R_{i}^{\alpha,\beta}(\varepsilon).$$

Here, the computation of $R_{i}^{\alpha,\beta}(\varepsilon)$ requires the estimated parameters $\alpha_i, \beta_i$ and $\sigma_i$ in Sect. 3.1. This group of equations may be solved via the least squares approximation $\min_{\tilde{c}_i} \| A \tilde{c}_i - B_i \|_2^2$. It is well known that its optimal solution is provided by

$$\tilde{c}_i = \left( A^T A \right)^{-1} \left( A^T B_i \right).$$
3.3 Algorithm for Identification of the Diffusion Term

Finally, the diffusion matrix \( a(x) \) remains to be found. Based on the dictionary of basis functions \( \Psi(x) \), its component can be estimated as \( a_{ij}(x) \approx \sum_{k=1}^{K} d_{ij,k} \psi_k(x) \), \( i, j = 1, 2, \ldots, n \).

According to the third assertion of Corollary 2, we derive the following group of equations

\[
\begin{align*}
A d_{ij} &= B_{ij}, \\
A &= \begin{bmatrix}
\psi_1(\hat{z}_1) & \cdots & \psi_K(\hat{z}_1) \\
\vdots & \ddots & \vdots \\
\psi_1(\hat{z}_M) & \cdots & \psi_K(\hat{z}_M)
\end{bmatrix}, \\
B_{ij} &= \hat{M} M^{-1} h^{-1} \left[ \left( \hat{x}_{i,1} - \hat{z}_{i,1} \right) \left( \hat{x}_{j,1} - \hat{z}_{j,1} \right), \\
&\quad \ldots, \\
&\quad \left( \hat{x}_{i,M} - \hat{z}_{i,M} \right) \left( \hat{x}_{j,M} - \hat{z}_{j,M} \right) \right]^T - S_{ij}^{\alpha,\beta}(\varepsilon).
\end{align*}
\] (12)

Subsequently, we also consider the least squares problem \( \min_{d_{ij}} \| A d_{ij} - B_{ij} \|_2^2 \) to find the optimal solution. It is well known that this leads to the solution

\[
\hat{d}_{ij} = \left( A^T A \right)^{-1} \left( A^T B_{ij} \right).
\] (13)

Since the diffusion matrix is symmetric, we just need to compute the coefficients for \( i = 1, 2, \ldots, n, j = i, i+1, \ldots, n \).

Remark that it is not unique to infer the coefficient \( A(x) \) in terms of the diffusion matrix \( a(x) \). Via the symmetry of \( a(x) \), the orthogonal matrix \( Q \) that is composed of the eigenvectors of \( a(x) \) can diagonalize the diffusion matrix, i.e., \( J = Q^T a Q \) with the component in the diagonal \( J_{ii} \geq 0 \). Then \( J = \left( Q^T A \right) \left( Q^T A \right)^T \). Denoting \( \sqrt{J} \) as the diagonal matrix with the diagonal component \( \sqrt{J_{ii}} \), the matrix \( \Lambda = Q \sqrt{J} \) is a solution of this equation. Note that for arbitrary orthogonal matrix \( T \), the matrix \( Q \sqrt{JT} \) is also a solution. Even so, they are statistically equivalent as they have the same Fokker–Planck equation for a given diffusion matrix.

4 Examples

We now present two prototypical examples to demonstrate our method for discovering stochastic dynamical systems from sample path data sets.

Example 1 Consider a stochastic Lorenz system

\[
\begin{align*}
\mathrm{d}x_1 &= 10 (-x_1 + x_2) \, \mathrm{d}t + (1 + x_3) \, \mathrm{d}B_{1,t} + \mathrm{d}B_{2,t} + 2 \, \mathrm{d}L_{1,t}, \\
\mathrm{d}x_2 &= (4x_1 - x_2 - x_1 x_3) \, \mathrm{d}t + x_2 \, \mathrm{d}B_{2,t} + \mathrm{d}L_{2,t}, \\
\mathrm{d}x_3 &= (-8/3 \, x_3 + x_1 x_2) \, \mathrm{d}t + x_1 \, \mathrm{d}B_{3,t} + 0.5 \, \mathrm{d}L_{3,t}.
\end{align*}
\]
Table 1  Identified Lévy motion for the three-dimensional Lorenz system

| Parameter | L₁ (L₂) | L₁ (L₂) | L₁ (L₂) |
|-----------|---------|---------|---------|
| True      | Learned | True    | Learned |
| α         | 0.5     | 0.498   | 1       | 1.0038  | 1.5 | 1.4832 |
| β         | 0.5     | 0.5066  | 0       | 0.0003  | −0.5 | −0.5118 |
| σ²        | 2       | 2.01    | 1       | 0.9956  | 0.5  | 0.4907 |

Table 2  Identified drift term for the three-dimensional Lorenz system

| Basis | b₁ (x) | b₂ (x) | b₃ (x) |
|-------|--------|--------|--------|
|       | True   | Learned | True   | Learned | True   | Learned |
| 1     | 0      | 0      | 0      | 0       | 0      | 0       |
| x₁    | −10    | −9.9801| 4      | 3.9941  | 0      | 0       |
| x₂    | 10     | 9.9666 | −1     | −1.0036 | 0      | 0       |
| x₃    | 0      | 0      | 0      | 0       | −2.6667| −2.6502 |
| x₁x₂  | 0      | 0      | 0      | 0       | 0      | 0       |
| x₁x₃  | 0      | 0      | −1     | −0.9917 | 0      | 0       |
| x₂x₃  | 0      | 0      | 0      | 0       | 0      | 0       |
| x₃²   | 0      | 0      | 0      | 0       | 0      | 0       |

The Lévy noise intensity $\sigma_1 = 2$, $\sigma_2 = 1$, $\sigma_3 = 0.5$ and the drift and the diffusion coefficients

$$b(x) = \begin{bmatrix} 10(-x_1 + x_2), 4x_1 - x_2 - x_1x_3, -\frac{8}{3}x_3 + x_1x_2 \end{bmatrix}^T,$$

$$a(x) = \begin{bmatrix} 2 + 2x_3 + x_3^2 & x_2 & 0 \\ x_2 & x_2 & 0 \\ 0 & 0 & x_1^2 \end{bmatrix}.$$

The time step is fixed as $h = 0.001$ and the chosen initial points $Z = [z_1, z_2, \ldots, z_M]$ are distributed uniformly within the cube $[-2, 2] \times [-2, 2] \times [-2, 2]$ with a mesh $400 \times 400 \times 400$. The data set $X$ is generated by Euler scheme of numerical scheme starting from $Z$ after the time $h$ with $\alpha_1 = 0.5$, $\beta_1 = 0.5$, $\alpha_2 = 1$, $\beta_2 = 0$ and $\alpha_3 = 1.5$, $\beta_3 = −0.5$. The dictionary is selected as polynomials to estimate the drift and diffusion terms.

With the parameters fixed as $N = 2$, $\epsilon = 1$ and $m = 5$, we compute the three groups of the parameters $\alpha$, $\beta$ and $\sigma$, and the drift and diffusion terms via our proposed algorithms and list the results in Tables 1, 2 and 3. As we see, the results agree well with the true parameters.

**Example 2**  Consider a gene regulation system driven by both Gaussian Brownian noise and non-Gaussian Lévy noise with a rational drift coefficient [49]

$$dx(t) = \left[ \frac{k f x^2(t)}{x^2(t) + K_d} - k_d x(t) + R_{bas} \right] dt + \frac{x(t)}{\sqrt{x^2(t) + 0.5}} dB_t + 0.5dL_t,$$
Table 3  Identified diffusion term for the three-dimensional Lorenz system

| Basis | $a_{11}(x)$ | $a_{12}(x)$ | $a_{13}(x)$ | $a_{22}(x)$ | $a_{23}(x)$ | $a_{33}(x)$ |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
|       | True        | Learned     | True        | Learned     | True        | Learned     | True        | Learned     | True        | Learned     |
| 1     | 2           | 2.2748      | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_1$ | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_2$ | 0           | 0           | 1           | 0.9968      | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_3$ | 2           | 1.993       | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_1^2$ | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_1x_2$ | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_1x_3$ | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_2^2$ | 0           | 0           | 0           | 0           | 1           | 1.0139      | 0           | 0           | 0           | 0           |
| $x_2x_3$ | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
| $x_3^2$ | 1           | 0.9982      | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           |
Parameter True Learned
\[ \alpha \] 1.5 1.5406
\[ \beta \] −0.5 −0.5237
\[ \sigma \] 0.5 0.538

where the system parameters are \( k_f = 6 \, \text{min}^{-1} \), \( K_d = 10 \), \( k_d = 1 \, \text{min}^{-1} \), and \( R_{bas} = 0.4 \, \text{min}^{-1} \). Then we have the drift coefficient \( b (x) = k_f x^2 / (x^2 + K_d) - k_d x + R_{bas} \), the diffusion coefficient \( a (x) = x^2 / (x^2 + 0.5) \) and the Lévy noise intensity \( \sigma = 0.5 \). The time step is fixed as \( h = 0.001 \) and the chosen \( 10^7 \) initial points \( Z = [z_1, z_2, \ldots, z_M] \) are distributed uniformly within the interval \([0, 5]\). The simulated data set \( X \) is generated by Euler scheme of numerical scheme starting from \( Z \) after the time \( h \). In order to check out the validation of our method to non-polynomial basis functions, we choose the following dictionary as in Ref. [3]

\[
\Psi (x) = [1, x, x^2, x^3, \sin x, \cos 11x, \sin 11x, -10 \tanh^2 (10x) + 10, -10 \tanh^2 (10x - 10) + 10, \exp \{-50x^2\}, \exp \{-50(x - 3)^2\}, \exp \{-0.3x^2\}, \exp \{-0.3(x - 3)^2\}, \exp \{-2(x - 2)^2\}, \exp \{-50(x - 4)^2\}, \exp \{-0.6(x - 4)^2\}, \exp \{-0.6(x - 3)^2\}, -2 \tanh^2 (2x - 4) + 2, \tanh^2 (x - 4) + 1].
\]

Here we only show the case with \( \alpha = 1.5 \) and \( \beta = -0.5 \). We choose the parameters as \( N = 2, \varepsilon = 1 \) and \( m = 5 \). According to our proposed method in Sect. 3, the learned values of the stability parameter \( \alpha \), the skewness parameter \( \beta \) and the Lévy noise intensity \( \sigma \) are listed in Table 4. It is seen that they are consistent with the true values. Employing Eqs. (11) and (13), we compute the least square solutions \( \tilde{c} \) and \( \tilde{d} \) as

\[
\tilde{c} = [-118.8168, -28.0260, 18.4281, -2.0028, 0.5302, 0, 0, 0, 0, 0, 0, 0, 0, 62.0744, 50.5969, 0, 6.2844, 0, 20.0348, -20.7253, 0, 26.9544]^T,
\]

\[
\tilde{d} = [-121.6943, -30.3672, 20.5902, -2.2510, 24.4084, 0, 0, 0, 0, 0, 68.1945, 49.7766, 0, 5.9156, 0, 19.6789, -19.9246, 0, 25.1716]^T.
\]

According to these learned coefficients, the estimated drift and diffusion terms are plotted in Fig. 3, indicating that the learned results provide a reasonable approximation in parameter range of interest. This implies that our method is valid even for stochastic dynamical systems with non-polynomial drift and diffusion terms and non-polynomial basis functions.

5 Discussion

In this paper, we have designed a machine learning method to discover stochastic dynamical systems with both non-Gaussian Lévy noise and Gaussian noise (Brownian motion), from
observational, experimental or simulated data sets. We have generalized our recently proposed method from the special stochastic dynamical systems with symmetric Lévy motion to the general systems with asymmetric Lévy motion. Based on the expressions (i.e., nonlocal Kramers–Moyal formulas) of the jump measure, drift coefficient and diffusion coefficient in terms of sample paths, we have further devised a numerical algorithm for our machine learning method. The verification of this method on prototypical systems shows its efficacy and accuracy. Our novel approach provides a data-driven tool to extract stochastic governing laws for complex phenomena, under general non-Gaussian fluctuations.

Comparing with our previously published method, there are three differences in this present work. First, it is generalized from stochastic systems with symmetric Lévy motion to systems with asymmetric Lévy motion. The symmetric case is only an idealized situation. Thus what we should identify are not only the stability parameter and the Lévy noise intensity but also the skewness parameter. The application range of our numerical algorithm is thus much wider. Second, through removing the dependence of the different Lévy components, the first assertion of Theorem 1 is transformed into one-dimensional computation via the marginal probability distribution. Finally, the integration domain in the second and third assertions of Theorem 1 is altered from a ball to a cube due to the independence of the different Lévy components.

Remark that it is still a challenge to extend our approach to identify the stochastic governing equations with multiplicative Lévy noise. In this case, the multiplicative Lévy noise destroys the “space homogeneity” in the first assertion of Theorem 1, leading to the situation that the function $W$ depends on both $x$ and $z$ individually (not just depends only on the spatial dislocation $x - z$).

Finally, note that our data-driven method can be also used to extract dynamical behaviors from sample paths. If we get large amounts of observational or experimental data for complex systems such as ecosystems [50], molecular dynamics [51], population dynamics [52], tensile behaviour of stainless steel [53], living neural networks [54], etc, we can use the method proposed to model it as stochastic differential equations and further investigate the escape events or other noise-induced phenomena. For example, the maximum likelihood transition path can be quantified by numerically computing the conditional probability density related to the estimated stochastic systems [4,22], and the mean exit time can be calculated by Dynkin formula [55], with the coefficients approximated from data.

**Acknowledgements** This research was supported by the National Natural Science Foundation of China (Grants 11771449).
Appendix A: \( \alpha \)-Stable Lévy Processes

A scalar \( \alpha \)-stable Lévy process \( L_t \) is a stochastic process with the following conditions:

(i) \( L_0 = 0 \), a.s.;
(ii) Independent increments: for any choice of \( n \geq 1 \) and \( t_0 < t_1 < \ldots < t_n-1 < t_n \), the random variables \( L_{t_0}, L_{t_1} - L_{t_0}, L_{t_2} - L_{t_1}, \ldots, L_{t_n} - L_{t_{n-1}} \) are independent;
(iii) Stationary increments: \( L_t - L_s \sim S_\alpha ((t - s)^{1/\alpha}, \beta, 0) \);
(iv) Stochastically continuous sample paths: for every \( s > 0 \), \( L_t \to L_s \) in probability, as \( t \to s \).

The \( \alpha \)-stable Lévy motion is a special but most popular type of the Lévy process defined by the stable random variable with the distribution \( S_\alpha (\delta, \beta, \lambda) \) [57–59]. Usually, \( \alpha \in (0, 2] \) is called the stability parameter, \( \delta \in (0, \infty) \) is the skewness parameter and \( \lambda \in (-\infty, \infty) \) is the shift parameter.

A stable random variable \( X \) with \( 0 < \alpha < 2 \) has the following “heavy tail” estimate:

\[
\lim_{x \to \infty} y^{\alpha} P (X > y) = C_\alpha \frac{1 + \beta}{2} \delta^\alpha,
\]

where \( C_\alpha \) is a positive constant depending on \( \alpha \). In other words, the tail estimate decays polynomially. The \( \alpha \)-stable Lévy motion has larger jumps with lower jump frequencies for smaller \( \alpha \) (\( 0 < \alpha < 1 \)), while it has smaller jump sizes with higher jump frequencies for larger \( \alpha \) (\( 1 < \alpha < 2 \)). The special case \( \alpha = 2 \) corresponds to (Gaussian) Brownian motion. For more information about Lévy process, refer to Refs. [43,44].

Appendix B: Proof of Theorem 1

(i) Since \( p (x, 0|z, 0) = \delta (x - z) \), then \( p_i (x_i, t|z_i, 0) = \delta (x_i - z_i) \) and \( p_i (x_i, 0|z_i, 0) = 0 \) for arbitrary \( x_i \) and \( z_i \) satisfying \( |x_i - z_i| > \epsilon \). For the convenience of representations, denote \( x^i \) as the rest of the vector \( x \) with \( x_i \) being removed. Then

\[
\lim_{t \to 0} t^{-1} p_i (x_i, t|z_i, 0) = \lim_{t \to 0} t^{-1} \left[ p_i (x_i, t|z_i, 0) - p_i (x_i, 0|z_i, 0) \right] \\
= \int_{\mathbb{R}^{n-1}} \left. \frac{\partial p (x, t|z, 0)}{\partial t} \right|_{t=0} dx^i \\
= - \int_{\mathbb{R}^{n-1}} \nabla \cdot [b p (x, 0|z, 0)] dx^i + \frac{1}{2} \int_{\mathbb{R}^{n-1}} \text{Tr} [H (a p (x, 0|z, 0))] dx^i \\
- \sum_{j=1}^{n} \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}\setminus\{0\}} \left[ p (x, 0|z, 0) - p (x - \sigma_j y_j e_j, 0|z, 0) \right] W_{\alpha, \beta} (y_j) dy_j dx^i.
\]
Since \( |x_i - z_i| > \varepsilon, p(x, 0|z, 0) = 0 \). Thus the above equation is reduced as
\[
\sum_{j=1}^{n} \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}\setminus\{0\}} p(x - \sigma_j y_j e_j, 0|z, 0) W_j^{\alpha,\beta}(y_j) \, dy_j \, dx^i.
\]
Since
\[
p(x - \sigma_j y_j e_j, 0|z, 0) = \sigma_j^{-1} \delta(x_1 - z_1) \ldots \delta(x_{j-1} - z_{j-1}) \delta(x_j - \sigma_j y_j - z_j)
\]
its integral is equal to zero for \( j \neq i \) due to \( \delta(x_i - z_i) = 0 \). Hence, we have
\[
\sum_{j=1}^{n} \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}\setminus\{0\}} p(x - \sigma_j y_j e_j, 0|z, 0) W_j^{\alpha,\beta}(y_j) \, dy_j \, dx^i = \sigma_i^{-1} W_i^{\alpha,\beta}(\sigma_i^{-1}(x_i - z_i)).
\]
(ii) According to the Fokker–Planck equation (2), we have
\[
\lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) p(x, t|z, 0) \, dx
\]
\[
= \lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) [p(x, t|z, 0) - p(x, 0|z, 0) + p(x, 0|z, 0)] \, dx
\]
\[
= \int_{x - z \in \Gamma} (x_i - z_i) \left. \frac{\partial p(x, t|z, 0)}{\partial t} \right|_{t=0} \, dx + \lim_{t \to 0} \left[ t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) \delta(x - z) \, dx \right]
\]
\[
= - \sum_{j=1}^{n} \int_{x - z \in \Gamma} (x_i - z_i) \frac{\partial}{\partial x_j} \left[ b_j(x) p(x, 0|z, 0) \right] \, dx
\]
\[
+ \frac{1}{2} \sum_{k,l=1}^{n} \int_{x - z \in \Gamma} (x_i - z_i) \frac{\partial^2}{\partial x_k \partial x_l} [a_{kl}(x) p(x, 0|z, 0)] \, dx
\]
\[
- \sum_{j=1}^{n} \int_{x - z \in \Gamma} \int_{\mathbb{R}\setminus\{0\}} (x_i - z_i) \left[ p(x, 0|z, 0) - p(x - \sigma_j y_j e_j, 0|z, 0) \right] \, dx^i W_j^{\alpha,\beta}(y_j) \, dy_j \, dx.
\]
The application of integration by parts into the first term leads to
\[
\sum_{j=1}^{n} \int_{x - z \in \Gamma} (x_i - z_i) \frac{\partial}{\partial x_j} \left[ b_j(x) p(x, 0|z, 0) \right] \, dx
\]
\[
= - \sum_{j=1}^{n} \int_{x - z \in \Gamma} b_j(x) p(x, 0|z, 0) \frac{\partial}{\partial x_j} (x_i - z_i) \, dx
\]
\[
= - \sum_{j=1}^{n} \int_{x - z \in \Gamma} b_j(x) \delta(x - z) \delta_{ij} \, dx
\]
\[
= - b_i(z).
\]
Therein, the boundary condition vanishes since $p(x,0|z,0) = 0$ as $|x_j - z_j| = \varepsilon$. For the second integration, we use integration by parts twice

$$\sum_{k,l=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} (x_i - z_i) \frac{\partial^2}{\partial x_k \partial x_l} [a_{kl}(\mathbf{x}) p(x,0|z,0)] d\mathbf{x}$$

$$= - \sum_{k,l=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \frac{\partial}{\partial x_k} (x_i - z_i) \frac{\partial}{\partial x_l} [a_{kl}(\mathbf{x}) p(x,0|z,0)] d\mathbf{x}$$

$$= \sum_{k,l=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} (x_i - z_i) a_{kl}(\mathbf{x}) p(x,0|z,0) \frac{\partial \delta_{ik}}{\partial x_l} d\mathbf{x}$$

$$= 0.$$  

For the third integration, we derive it separately. First, according to Tonelli’s theorem [60], we obtain

$$\sum_{j=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \int_{\mathbb{R}\setminus\{0\}} (x_i - z_i) p(x,0|z,0) W^{|\alpha,\beta|}_j (y_j) dy_j d\mathbf{x}$$

$$= \sum_{j=1}^{n} \int_{\mathbb{R}\setminus\{0\}} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} (x_i - z_i) \delta(\mathbf{x} - \mathbf{z}) d\mathbf{x} W^{|\alpha,\beta|}_j (y_j) dy_j$$

$$= 0.$$  

Second,

$$\sum_{j=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \int_{\mathbb{R}\setminus\{0\}} (x_i - z_i) p(x - \sigma_y y_j e_j,0|z,0) W^{|\alpha,\beta|}_j (y_j) dy_j d\mathbf{x}$$

$$= \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \int_{\mathbb{R}\setminus\{0\}} (x_i - z_i) p(x - \sigma_i y_i e_i,0|z,0) W^{|\alpha,\beta|}_i (y_i) dy_i d\mathbf{x}$$

$$= \sigma^{-1}_i \int_{-\varepsilon}^{\varepsilon} y_i W^{|\alpha,\beta|}_i \left(\sigma^{-1}_i y_i\right) dy_i.$$  

Finally, according to Tonelli’s theorem and integration by parts, we have

$$\sum_{j=1}^{n} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \int_{\mathbb{R}\setminus\{0\}} (x_i - z_i) \sigma_j^\alpha \chi^\alpha_j (y_j) y_j \frac{\partial}{\partial x_j} p(x,0|z,0) W^{|\alpha,\beta|}_j (y_j) dy_j d\mathbf{x}$$

$$= \sum_{j=1}^{n} \int_{\mathbb{R}\setminus\{0\}} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} \sigma_j^\alpha \chi^\alpha_j (y_j) y_j \frac{\partial}{\partial x_j} p(x,0|z,0) d\mathbf{x} W^{|\alpha,\beta|}_j (y_j) dy_j$$

$$= - \sum_{j=1}^{n} \int_{\mathbb{R}\setminus\{0\}} \int_{\mathbf{x} - \mathbf{z} \in \Gamma} p(x,0|z,0) \frac{\partial}{\partial x_j} (x_i - z_i) d\mathbf{x} W^{|\alpha,\beta|}_j (y_j) dy_j$$

$$= \begin{cases} 0, & \alpha < 1, \\ -\sigma^{-1}_i \int_{-1}^{1} y_i W^{|\alpha,\beta|}_i \left(\sigma^{-1}_i y_i\right) dy_i, & \alpha = 1, \\ -\sigma^{-1}_i \int_{-\infty}^{\infty} y_i W^{|\alpha,\beta|}_i \left(\sigma^{-1}_i y_i\right) dy_i, & \alpha > 1. \end{cases}$$
Hence,

\[
\lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) p(x, t|z, 0) \, dx = b_i(z) + R_i^{\alpha, \beta}(\varepsilon).
\]

(iii) According to the Fokker–Planck equation (2), we have

\[
\lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) p(x, t|z, 0) \, dx
\]

\[
= \lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) [p(x, t|z, 0) - p(x, 0|z, 0) + p(x, 0|z, 0)] \, dx
\]

\[
= \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \left. \frac{\partial p(x, t|z, 0)}{\partial t} \right|_{t=0} \, dx
\]

\[
+ \lim_{t \to 0} t^{-1} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \, dx
\]

\[
= -n \sum_{k=1}^{\alpha} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \frac{\partial}{\partial x_k} [b_k(x) p(x, 0|z, 0)] \, dx
\]

\[
+ \frac{1}{2} \sum_{k, l=1}^{\alpha} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \frac{\partial^2}{\partial x_k \partial x_l} [a_{kl}(x) p(x, 0|z, 0)] \, dx
\]

\[
- \sum_{k=1}^{\alpha} \int_{x - z \in \Gamma} \int_{\mathbb{R} \setminus \{0\}} (x_i - z_i) (x_j - z_j) [p(x, 0|z, 0) - p(x - \sigma_k y_k e_k, 0|z, 0)
\]

\[
- \sigma_k \chi_k^\alpha(y_k) y_k \frac{\partial}{\partial x_k} p(x, 0|z, 0)] \, dy_k \, dx
\]

\[
\int_{\mathbb{R} \setminus \{0\}} W_k^{\alpha, \beta}(y_k) \, dy_k.
\]

The application of integration by parts into the first term yields

\[
\sum_{k=1}^{\alpha} \int_{x - z \in \Gamma} (x_i - z_i) (x_j - z_j) \frac{\partial}{\partial x_k} [b_k(x) p(x, 0|z, 0)] \, dx
\]

\[
= -\sum_{k=1}^{\alpha} \int_{x - z \in \Gamma} b_k(x) p(x, 0|z, 0) \frac{\partial}{\partial x_k} [(x_i - z_i) (x_j - z_j)] \, dx
\]

\[
= -\sum_{k=1}^{\alpha} \int_{x - z \in \Gamma} b_k(x) \delta(x - z) [\delta_{ik} (x_j - z_j) + \delta_{jk} (x_i - z_i)] \, dx
\]

\[
= 0.
\]
Therein, the boundary condition vanishes since \( p(x, 0|z, 0) = 0 \) as \(|x_k - z_k| = \varepsilon\). For the second integration, we use integration by parts again.

\[
\frac{1}{2} \sum_{k,l=1}^{n} \int_{x-z \in \Gamma} (x_i - z_i) (x_j - z_j) \frac{\partial^2}{\partial x_k \partial x_l} [a_{kl}(x) p(x, 0|z, 0)] \, dx
\]

\[
= -\frac{1}{2} \sum_{k,l=1}^{n} \int_{x-z \in \Gamma} \frac{\partial}{\partial x_k} [(x_i - z_i) (x_j - z_j)] \frac{\partial}{\partial x_l} [a_{kl}(x) p(x, 0|z, 0)] \, dx
\]

\[
= \frac{1}{2} \sum_{k,l=1}^{n} \int_{x-z \in \Gamma} a_{kl}(x) p(x, 0|z, 0) \left[ \frac{\partial}{\partial x_l} [\delta_{jk} (x_j - z_j)] + \delta_{jk} (x_i - z_i) \right] \, dx
\]

\[
= \frac{1}{2} \sum_{k,l=1}^{n} \int_{x-z \in \Gamma} a_{kl}(x) [\delta (x - z) (\delta_{lk} \delta_{jl} + \delta_{il} \delta_{jk})] \, dx
\]

\[
= \frac{1}{2} \left[ a_{ij}(z) + a_{ji}(z) \right]
\]

\[
= a_{ij}(z).
\]

We still derive the third integration separately. First, according to Tonelli’s theorem, we obtain

\[
\sum_{j=1}^{n} \int_{x-z \in \Gamma} \int_{\mathbb{R} \setminus \{0\}} (x_i - z_i) (x_j - z_j) p(x, 0|z, 0) W_j^{\alpha,\beta} (y_j) \, dy_j \, dx
\]

\[
= \sum_{j=1}^{n} \int_{\mathbb{R} \setminus \{0\}} \int_{x-z \in \Gamma} (x_j - z_j) \delta (x - z) \, dx W_j^{\alpha,\beta} (y_j) \, dy_j
\]

\[
= 0.
\]

Second, for \( i \neq j \),

\[
\sum_{k=1}^{n} \int_{x-z \in \Gamma} \int_{\mathbb{R} \setminus \{0\}} (x_i - z_i) (x_j - z_j) p(x - \sigma_k y_k e_k, 0|z, 0) W_k^{\alpha,\beta} (y_k) \, dy_k \, dx = 0
\]

and for \( i = j \),

\[
\sum_{k=1}^{n} \int_{x-z \in \Gamma} \int_{\mathbb{R} \setminus \{0\}} (x_i - z_i)^2 p(x - \sigma_k y_k e_k, 0|z, 0) W_k^{\alpha,\beta} (y_k) \, dy_k \, dx
\]

\[
= \int_{x-z \in \Gamma} \int_{\mathbb{R} \setminus \{0\}} (x_i - z_i)^2 p(x - \sigma_i y_i e_i, 0|z, 0) W_i^{\alpha,\beta} (y_i) \, dy_i \, dx
\]

\[
= \sigma_i^{-1} \int_{-\varepsilon}^{\varepsilon} y_i^2 W_i^{\alpha,\beta} (\sigma_i^{-1} y_i) \, dy_i.
\]
This integration is bounded according to the definition of the jump measure. Finally, according to Tonelli’s theorem and integration by parts, we have
\[
\sum_{k=1}^{n} \int_{x-z \in \Gamma} \int_{\mathbb{R}\setminus[0]} (x_i - z_i) (x_j - z_j) \sigma_k \chi_k^y \frac{\partial p (x, 0|z, 0)}{\partial x_k} W_k^{\alpha, \beta} (y_k) \, dy_k \, dx
\]
\[
= \sum_{k=1}^{n} \int_{\mathbb{R}\setminus[0]} \int_{x-z \in \Gamma} (x_i - z_i) (x_j - z_j) \frac{\partial}{\partial x_k} p (x, 0|z, 0) \, dx \sigma_k \chi_k^y \gamma_k W_k^{\alpha, \beta} (y_k) \, dy_k
\]
\[
= - \sum_{k=1}^{n} \int_{\mathbb{R}\setminus[0]} \int_{x-z \in \Gamma} \delta (x - z) \left[ \delta_{jk} (x_j - z_j) + \delta_{ik} (x_i - z_i) \right] \, dx \sigma_k \chi_k^y \gamma_k W_k^{\alpha, \beta} (y_k) \, dy_k
\]
\[
= 0.
\]
Hence,
\[
\lim_{t \to 0} t^{-1} \int_{x-z \in \Gamma} (x_i - z_i) (x_j - z_j) p (x, t|z, 0) \, dx = a_{ij} (z) + b_{ij} (\varepsilon).
\]
The proof is complete.

Appendix C: Proof of Corollary 2

(i) This is derived directly by integrating the equation in the first assertion of Theorem 1 about \(x_i\) on the interval \([c_1, c_2]\).

(ii) Let the set \(dU = [u_1, u_1 + du_1) \times [u_2, u_2 + du_2) \times \cdots \times [u_n, u_n + du_n]\). Then we have
\[
\mathbb{P} \{ x (t) \in dU; \ x (t) - z \in \Gamma | x (0) = z \}
\]
\[
= \mathbb{P} \{ x (t) \in dU | x (0) = z \} \cdot \mathbb{P} \{ x (t) - z \in \Gamma | x (0) = z \}.
\]
Thus
\[
\int_{x-z \in \Gamma} (x_i - z_i) p (x, t|z, 0) \, dx
\]
\[
= \int_{u-z \in \Gamma} (u_i - z_i) \mathbb{P} \{ x (t) \in dU | x (0) = z \}
\]
\[
= \int_{u-z \in \Gamma} (u_i - z_i) \mathbb{P} \{ x (t) \in dU; \ x (t) - z \in \Gamma | x (0) = z \}
\]
\[
= \mathbb{P} \{ x (t) - z \in \Gamma | x (0) = z \} \cdot \int_{|u-z| < \varepsilon} (u_i - z_i) \, \mathbb{P} \{ x (t) \in dU | x (0) = z \}
\]
\[
= \mathbb{P} \{ x (t) - z \in \Gamma | x (0) = z \} \cdot \mathbb{E} [ (x_i (t) - z_i) | x (0) = z; \ x (t) - z \in \Gamma].
\]
Hence, the conclusion is immediately deduced
\[
\lim_{t \to 0} t^{-1} \mathbb{P} \{ x (t) - z \in \Gamma | x (0) = z \} \cdot \mathbb{E} [ (x_i (t) - z_i) | x (0) = z; \ x (t) - z \in \Gamma]
\]
\[
= b_i (z) + R^{\alpha, \beta}_i (\varepsilon).
\]
(iii) This proof is similar to the second conclusion. The proof is complete.
References

1. Hosni, H., Vulpiani, A.: Forecasting in light of big data. Philos. Technol. 31, 557–569 (2018)
2. Baldovin, M., Cecconi, F., Cencini, M., Puglisi, A., Vulpiani, A.: The role of data in model building and prediction: a survey through examples. Entropy 20, 807 (2018)
3. Boninsegni, L., Nüske, F., Clementi, C.: Sparse learning of stochastic dynamical equations. J. Chem. Phys. 148, 241723 (2018)
4. Dai, M., Gao, T., Lu, Y., Zheng, Y., Duan, J.: Detecting the maximum likelihood transition path from data of stochastic dynamical systems. Chaos 30, 113124 (2020)
5. Reza Rahimi Tabar, M.: Analysis and data-based reconstruction of complex nonlinear dynamical systems. Springer, New York (2019)
6. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. 25, 1307–1346 (2015)
7. Klus, S., Nüske, F., Peitz, S., Niemann, J.H., Clementi, C., Schütte, C.: Data-driven approximation of the Koopman generator: model reduction, system identification, and control. Phys. D Nonlinear Phenom. 406, 132416 (2020)
8. Wu, D., Fu, M., Duan, J.: Discovering mean residence time and escape probability from data of stochastic dynamical systems. Chaos 29, 093122 (2019)
9. Lu, Y., Duan, J.: Discovering transition phenomena from data of stochastic dynamical systems with Lévy noise. Chaos 30, 093110 (2020)
10. Zhang, Y., Duan, J., Jin, Y., Li, Y.: Extracting non-Gaussian governing laws from data on mean exit time. Chaos 30, 113112 (2020)
11. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. 113, 3932–3937 (2016)
12. González-García, R., Rico-Martínez, R., Kevrekidis, I.G.: Identification of distributed parameter systems: a neural net based approach. Comput. Chem. Eng. 22, S965–S968 (1998)
13. Schaeffer, H., Caltisch, R., Hauck, C.D., Osher, S.: Sparse dynamics for partial differential equations. Proc. Natl. Acad. Sci. 110, 6634–6639 (2013)
14. Rudy, S., Alla, A., Brunton, S.L., Kutz, J.N.: Data-driven identification of parametric partial differential Equations. SIAM 18, 643–660 (2019)
15. Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Data-driven discovery of partial differential equations. Sci. Adv. 3, e1602614 (2017)
16. Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I. (eds.): Lévy processes: theory and applications. Springer, New York (2001)
17. Ditlvesen, P.D.: Observation of α-stable noise induced millennial climate changes from an ice-core record. Geophys. Res. Lett. 26, 1441–1444 (1999)
18. Raser, J.M., Shea, E.K.O.: Noise in gene expression: origins, consequences, and control. Science 309(5738), 2010–2013 (2005)
19. Jourdain, B., Méleard, S., Woyczynski, W.A.: Lévy flights in evolutionary ecology. J. Math. Biol. 65, 677–707 (2012)
20. Weeks, E.R., Solomon, T.H., Urbach, J.S., Swinney, H.L.: Observation of anomalous diffusion and Lévy flights BT-Lévy Flights and Related Topics in Physics. In: M. F. Shlesinger, G. M. Zaslavsky, U. Frisch (Eds.). Springer, Berlin, Heidelberg (1995)
21. Böttcher, B.: Feller processes: the next generation in modeling Brownian motion. Lévy processes and beyond. PLoS ONE 5, e15102 (2010)
22. Zheng, Y., Yang, F., Duan, J., Sun, X., Fu, L., Kurths, J.: The maximum likelihood climate change for global warming under the influence of greenhouse effect and Lévy noise. Chaos 30, 013132 (2020)
23. Kharcheva, A.A., Dubkov, A.A., Dybiec, B., Spagnolo, B., Valenti, D.: Spectral characteristics of steady-state Lévy flights in confinement potential profiles. J. Stat. Mech. Theory E 2016, 054039 (2016)
24. Cai, R., Chen, X., Duan, J., Kurths, J., Li, X.: Lévy noise-induced escape in an excitable system. J. Stat. Mech. Theory Exp. 6, 063503 (2017)
25. Cai, R., He, Z., Liu, Y., Duan, J., Kurths, J., Li, X.: Effects of Lévy noise on the FitzHugh–Nagumo model: a perspective on the maximal likely trajectories. J. Theor. Biol. 480, 166–174 (2019)
26. Dubkov, A.A., LaCognata, A., Spagnolo, B.: The problem of analytical calculation of barrier crossing characteristics for Lévy flights. J. Stat. Mech. Theory E 2009, P01002 (2009)
27. Patel, A., Kosko, B.: Stochastic resonance in continuous and spiking neuron models with Lévy noise. IEEE Trans. Neural Netw. 19, 1993–2008 (2008)
28. La Cognata, A., Valenti, D., Dubkov, A.A., Spagnolo, B.: Dynamics of two competing species in the presence of Lévy noise sources. Phys. Rev. E 82, 11121 (2010)
29. Cheng, X., Wang, H., Wang, X., Duan, J., Li, X.: Most probable transition pathways and maximal likely trajectories in a genetic regulatory system. Phys. A Stat. Mech. Appl. 531, 121779 (2019)
30. Serdukova, L., Zheng, Y., Duan, J., Kurths, J.: Stochastic basins of attraction for metastable states. Chaos 26, 073117 (2016)
31. Wu, F., Chen, X., Zheng, Y., Duan, J., Li, X.: Most probable transition pathways and maximal likely trajectories in a genetic regulatory system. Phys. A Stat. Mech. Appl. 531, 121779 (2019)
32. Serdukova, L., Zheng, Y., Duan, J., Kurths, J.: Stochastic basins of attraction for metastable states. Chaos 26, 073117 (2016)
33. Wu, F., Chen, X., Zheng, Y., Duan, J., Kurths, J., Li, X.: Lévy noise induced transition and enhanced stability in a gene regulatory network. Chaos 28, 075510 (2018)
34. Guarcello, C., Valenti, D., Spagnolo, B., Pierro, V., Valenti, D.: Statistics of residence time for Lévy flights in unstable parabolic potentials. Phys. Rev. E 102, 042142 (2020)
35. Guarcello, C., Filatrella, G., Spagnolo, B., Pierro, V., Filatrella, G.: Anomalous transport effects on switching currents of graphene-based Josephson junctions. Nanotechnology 28, 134001 (2017)
36. Lisowski, B., Valenti, D., Spagnolo, B., Bier, M., Gudowska-Nowak, E.: Stepping molecular motor amid Lévy white noise. Phys. Rev. E 91, 042713 (2015)
37. Liu, Q., Jiang, D., Hayat, T., Ahmad, B.: Analysis of a delayed vaccinated SIR epidemic model with temporary immunity and Lévy jumps. Nonlinear Anal. Hybrid Syst. 27, 29–43 (2018)
38. Li, Y., Duan, J., Liu, X., Zhang, Y.: Most probable dynamics of stochastic dynamical systems with exponentially light jump fluctuations. Chaos 30, 063142 (2020)
39. Dubkov, A.A., Dybiec, B., Spagnolo, B., Kharcheva, A., Guarcello, C., Valenti, D.: Statistics of residence time for Lévy flights in unstable parabolic potentials. Phys. Rev. E 102, 042142 (2020)
40. Carollo, A., Spagnolo, B., Valenti, D.: Uhlmann curvature in dissipative phase transitions. Sci. Rep.-Nat. 8, 9852 (2018)
41. Li, Y., Duan, J.: A data-driven approach for discovering stochastic dynamical systems with non-Gaussian Levy noise. Phys. D Nonlinear Phenom. 417, 132830 (2021)
42. Gardiner, C.W.: Handbook of Stochastic Methods. Springer, Berlin (1985)
43. Duan, J.: An Introduction to Stochastic Dynamics. Cambridge University Press, New York (2015)
44. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge University Press, New York (2009)
45. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. CRC Press, London (2003)
46. Janicki, A., Weron, A.: Simulation and Chaotic Behavior of α-Stable Stochastic Processes. Wroclaw University of Technology, Hugo Steinhaus Center, Wroclaw (1994)
47. Sun, X., Li, X., Zheng, Y.: Governing equations for probability densities of Marcus stochastic differential equations with Lévy noise. Stoch. Dynam. 17, 1750033 (2016)
48. Dubkov, A., Spagnolo, B.: Generalized Wiener process and Kolmogorov’s equation for diffusion induced by non-Gaussian noise source. Fluct. Noise Lett. 5, L267–L274 (2005)
49. Liu, Q., Jia, Y.: Fluctuations-induced switch in the gene transcriptional regulatory system. Phys. Rev. E 70, 041907 (2004)
50. Ciuchi, S., De Pasquale, F., Spagnolo, B.: Self-regulation mechanism of an ecosystem in a non-Gaussian fluctuation regime. Phys. Rev. E 54, 706 (1996)
51. Pizzolato, N., Fiasconaro, A., Adorno, D.P., Spagnolo, B.: Resonant activation in polymer translocation: new insights into the escape dynamics of molecules driven by an oscillating field. Phys. Biol. 7, 034001 (2010)
52. Denaro, G., Valenti, D., Spagnolo, B., Basiline, G., Mazzola, S., Zgozi, S.W., Aronica, S., Bonanno, A.: Dynamics of two picophytoplankton groups in mediterranean sea: Analysis of the deep chlorophyll maximum by a stochastic advection-reaction-diffusion model. PLoS ONE 8, e66765 (2013)
53. Qiu, C., Al Kindi, M., Aladawi, A.S., Al Hatmi, I.: A comprehensive study on microstructure and tensile behaviour of a selectively laser melted stainless steel. Sci. Rep. 8, 1–16 (2018)
54. Mikhaylov, A., Pimashkin, A., Figareva, Y., Gerasimova, S., Gryaznov, E., Shchankov, S., Zuev, A., Talanov, M., Lavrov, I., Demin, V., Erokhin, V., Lobov, S., Mukhina, I., Kazantzev, V., Wu, H., Spagnolo, B.: Neurohybrid memristive CMOS-integrated systems for biosensors and neuroprosthetics. Front. Neurosci. 14, 358 (2020)
55. Gao, T., Duan, J., Li, X., Song, R.: Mean exit time and escape probability for dynamical systems driven by Lévy noises. SIAM J. Sci. Comput. 36, A887–A906 (2014)
56. Li, Y.: https://github.com/liyangnuaa/Machine-learning-for-asymmetric-Levy-motion. GitHub (2020)
57. Adler, R., Feldman, R., Taqu, M.: A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhäuser, Boston (1998)
58. Kyprianou, A.E.: Fluctuations of Lévy Processes with Applications: Introductory Lectures. Springer, Berlin, Heidelberg (2014)
59. Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I.: Lévy Processes: Theory and Applications. Springer/Birkhäuser, Boston (2001)
60. Krantz, S.G., Parks, H.R.: Geometric Integration Theory. Springer/Birkhäuser, Boston (2008)

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