Physica D 398 (2019) 69–83

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Hamiltonian systems with Lévy noise: Symplecticity, Hamilton's principle and averaging principle[☆]

Pingyuan Wei[a], Ying Chao[a], Jinqiao Duan[b,*]

[a] School of Mathematics and Statistics, & Center for Mathematical Sciences, Huazhong University of Sciences and Technology, Wuhan 430074, China
[b] Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA

ARTICLE INFO

Article history:
Received 17 January 2019
Received in revised form 23 April 2019
Accepted 28 May 2019
Available online 3 June 2019
Communicated by T. Wanner

Keywords:
Stochastic Hamiltonian systems
Lévy noise
Symplecticity
Hamilton's principle
Averaging principle

ABSTRACT

This work focuses on topics related to Hamiltonian stochastic differential equations with Lévy noise. We first show that the phase flow of the stochastic system preserves symplectic structure, and propose a stochastic version of Hamilton's principle by the corresponding formulation of the stochastic action integral and the Euler–Lagrange equation. Based on these properties, we further investigate the effective behavior of a small transversal perturbation to a completely integrable stochastic Hamiltonian system with Lévy noise. We establish an averaging principle in the sense that the action component of solution converges to the solution of a stochastic differential equation when the scale parameter goes to zero. Furthermore, we obtain the estimation for the rate of this convergence. Finally, we present an example to illustrate these results.

1. Introduction

Certain nonlinear systems have "geometric" structures, such as the Hamiltonian structure [1–3]. Hamiltonian systems of ordinary differential equations (ODEs) widely appear in celestial mechanics, statistical mechanics, geophysics, and chemical physics. They are models for the dynamics of planets, motion of particles in a fluid, and evolution of other microscopic systems [4]. Hamiltonian systems have many well-known properties. For example, it was known to Liouville that the flows of Hamiltonian systems possess the property of phase-volume preservation; Poincaré observed that the Hamiltonian flows are symplectic and geometrically preserve certain symplectic area along phase flow [5]; based on Hamilton's principle, Hamiltonian equations of motion are closely related to Euler–Lagrange differential equations [2,6]. As a matter of fact, these dynamical systems are often subject to perturbations. In the deterministic case, the perturbation theory of Hamiltonian systems have appeared long ago; see Arnold [1] and Freidlin–Wentzell [7] for details. Particularly, an averaging principle for an integrable Hamiltonian system has been studied in e.g. Arnold [1].

It is important to take randomness into account when building mathematical models for complex phenomena under uncertainty [8]. Stochastic differential equations (SDEs) with "Hamiltonian structures" are appropriate models for randomly influenced Hamiltonian systems as studied in Bismut [9], and have also drawn much attention; see, for example, Brin–Freidlin [10], MacKay [11], Misawa [12], Wu [13], Zhu–Huang [14]. In particular, Milstein et al. [15,16] proved the symplecticity for stochastic Hamiltonian systems with Brownian noise, and Wang et al. [17] proposed a version of Hamilton's principle for the same systems to construct variational integrators; Pavon [18] established variational principles in stochastic mechanics; Li [19] developed an averaging principle for a perturbed completely integrable stochastic Hamiltonian system with Brownian noise. For some specific physical Hamiltonian models, we refer to Cresson–Darses [20] and Givon et al. [21].

In view of the development on SDEs with Hamiltonian structures, the noise processes considered to date are mainly Gaussian noise in terms of Brownian motion. However, non-Gaussian random fluctuations should be introduced to capture some large moves and unpredictable events in various areas such as not only aforementioned celestial mechanics and statistical physics, but also mathematics finance and life science [8,22–24]. Lévy motions are an important and useful class of non-Gaussian processes whose sample paths are càdlàg (right-continuous with left limit at each time instant). The study on stochastic systems driven by such processes has received increasing attentions recently, especially on developing proper averaging principles for these systems. For example, Albeverio et al. [25,26] established ergodicity of Lévy-type operators and SDEs driven by jump noise with non-Lipschitz coefficients; Högele–Ruffino [27] and

☆ This work was partly supported by the NSFC grants 11531006 and 11771449.

* Corresponding author.
E-mail addresses: weipingyuan@hust.edu.cn (P. Wei), yingchao1993@hust.edu.cn (Y. Chao), duan@iit.edu (J. Duan).

https://doi.org/10.1016/j.physd.2019.05.013
0167-2789/© 2019 Published by Elsevier B.V.

Gargate–Ruffino [28] focused on averaging along foliated Brownian and Lévy diffusions, respectively, which generalized the approach by Li [19], and Högele–da Costa [29] further studied strong averaging along foliated Lévy diffusions with heavy tails on compact leaves. For more information on averaging principle for stochastic systems driven by Lévy noise, we refer to Xu et al. [30] and Bao et al. [31]. ODEs and SDEs with "Hamiltonian structures" usually exhibit some extraordinary properties. Nevertheless, averaging principles for SDEs driven by Lévy noise with "Hamiltonian structures", and even some basic dynamics such as symplecticity (invariance under a transformation) and Lévy-type stochastic Hamilton's principle of these systems, have not yet been considered to date to the best of our knowledge.

In this present paper, we consider stochastic Hamiltonian systems with Lévy noise on symplectic manifolds. They are defined as Marcus SDEs whose drift vector fields and diffusion vector fields are Hamiltonian vector fields. Note that the Marcus integral [32–34] in Lévy case has the advantage of leading to ordinary chain rule of the Newton–Leibniz type under a transformation. This property makes the Marcus integral natural to use especially in connection with SDEs on manifolds [35].

We first demonstrate that the phase flow of a stochastic Hamiltonian system with Lévy noise preserves symplectic structure, and then propose the formulation of Lévy-type stochastic action integral and Euler–Lagrange equation of motions, as well as the stochastic Hamilton's principle. These properties are derived by using the calculus of variations, and the demand of the systems being in Marcus sense will simplify the stochastic differential calculations in the proofs. It is important to note that the stochastic Hamiltonian systems with Lévy noise should be understood as special nonconservative systems, for which the Lévy noise is a nonconservative 'force'. The symplecticity here is presented for the whole stochastic system instead of the original deterministic Hamiltonian system without the nonconservative force. The stochastic Hamilton's principle is also proposed on the basis of nonconservative mechanical systems.

Based on these foundational work, we further investigate the effective behavior of a small transversal perturbation to a (completely) integrable stochastic Hamiltonian system with Lévy noise. As this integrable stochastic system is perturbed by a transversal smooth vector field of order ε (ε is a small parameter), the solution to the perturbed equation will not preserve the properties mentioned above. The main idea we will use is to consider the solution along the rescaled time t/ε. The motion splits into two parts with fast rotation along the unperturbed trajectories and slow motion across them. Indeed, by an action–angle coordinate, the fast rotation is a diffusion on the invariant torus and the slow motion is governed by the transversal component. When averaged by ergodic invariant measure on torus, the evolution of action component of the motion does not depend on the angular variable when ε tends to zero. The essential transversal behavior is captured by a system of ODEs for the transversal component and this result is referred as an averaging principle. The estimation for rate of convergence for this averaging principle is also established. Some inspiration for this part came from Li [19], as well as Högele–de Costa [29]. The main novelty of our work is that the model we consider here combines features of a Hamiltonian structure with stochastic non-Gaussian Lévy noise.

The rest of this paper is organized as follows. In Section 2, we recall basic concepts about Hamiltonian vector fields and Lévy motions, and then present the definition of stochastic Hamiltonian system with Lévy noise, together with the existence and uniqueness of the solution. In Section 3, we show that the phase flow of this stochastic system preserves the symplectic structure. By considering a stochastic Hamiltonian system with Lévy noise as a special nonconservative system, we propose a stochastic version of Hamilton's principle. The goal of this section is to better understand such a system and to establish foundation for the following sections of this paper. In Section 4, we investigate an integrable stochastic Hamiltonian system, with Lévy noise, perturbed by a transversal smooth vector field. After discussing the ergodic behavior and some technical issues, we establish an averaging principle, together with a specific illustrative example.

2. Preliminaries

2.1. Stochastic Hamiltonian systems with Lévy noise

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geqslant 0}, P)$ be a filtered probability space endowed with a Poisson random measure N on $(\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}_+$ with jump intensity measure $\nu = \mathbb{E}N(1,\cdot)$. Denote by $\widetilde{N}$ the associated compensated Poisson random measure, that is, $\widetilde{N}(dt,dz) = N(dt,dz) - \nu(dz)dt$. We assume that the filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$ satisfies the usual conditions [36]. Let $L_t = L(t)$ be a d-dimensional Lévy process with the generating triplet (γ, A, ν). By Lévy–Itô decomposition [22,23,32],

$$L_t = \gamma t + B_A(t) + \int_{|z|<1} z\widetilde{N}(t,dz) + \int_{|z|\geq 1} zN(t,dz),$$

where $\gamma \in \mathbb{R}^d$ is a drift vector, $B_A(t)$ is an independent d-dimensional Brownian motion with covariance matrix A, and the last two terms describe the 'small jumps' and 'big jumps' of Lévy process, respectively. In the following, we denote $L_c(t) = \gamma t + B_A(t)$ as the continuous part of L_t and $L_d(t) = L_t - L_c(t)$ as the discontinuous part.

Given a smooth Hamiltonian H_0 and a family of n smooth Hamiltonians $\{H_k\}_{k=1}^n$ on a smooth $2n$-dimensional manifold M [1,37]. We denote by V_0 and V_k $(k = 1, 2, \ldots, d)$ the corresponding Hamiltonian vector fields, that is,

$$dH_0(\cdot) = \omega^2(\cdot, V_0),\ \ dH_k(\cdot) = \omega^2(\cdot, V_k),$$

where ω^2 is the symplectic form. Note that we use the symbol with superscript 2 for the symplectic form to avoid confusion with the customary symbol for chance variable on sample space Ω.

We shall consider stochastic Hamiltonian systems driven by non-Gaussian Lévy noise, which are described by the following SDEs in the Marcus form on M:

$$dX = V_0(X)dt + \sum_{k=1}^{d} V_k(X) \diamond dL^k(t),\ \ X_0 := X(t_0) = x \in M, \tag{2.1}$$

or equivalently,

$$X_t = x + \int_0^t V_0(X_s)ds + \sum_{k=1}^{d} \int_0^t V_k(X_{s-}) \diamond dL^k(s), \tag{2.2}$$

where "$\diamond$" stands for Marcus integral [32–34] defined by

$$\int_0^t V_k(X_{s-}) \diamond dL_k(s) = \int_0^t V_k(X_{s-}) \circ dL_c^k(s) + \int_0^t V_k(X_{s-})dL_d^k(s) + \sum_{0\leqslant s\leqslant t} \left[\phi(\Delta L^k(s), V_k(X_{s-}), X_{s-}) - X_{s-} - V_k(X_{s-})\Delta L^k(s)\right] \tag{2.3}$$

with $\int \circ dL_c^k(s)$ denoting the Stratonovich integral, $\int dL_d^k(s)$ denoting the Itô integral and $\phi(l, v(x), x)$ being the value at $t = 1$ of the solution of the following ODE:

$$\frac{d}{dt}\xi(t) = v(\xi(t))l,\ \ \xi(0) = x. \tag{2.4}$$

Note that Marcus SDEs (2.1) satisfy chain rule under a transformation (change of variable) and $\mathbb{P}(X_0 \in M) = 1$ implies that $\mathbb{P}(X_t \in M, t \geqslant 0) = 1$, for details see Kurtz et al. [33].

We remark that, by Lévy–Itô decomposition, the systems (2.1) with Lévy triplet being $(0, I, 0)$ are stochastic Hamiltonian systems with Brownian noise [15–17], and the systems (2.1) without Lévy term are deterministic Hamiltonian systems.

2.2. Existence and uniqueness

In order to ensure the existence and uniqueness for the stochastic dynamical systems with Hamiltonian structure, we will need to make some assumptions. First we rewrite the Marcus equations (2.1) and (2.2) in the Itô form [22,32]. This can be carried out by employing the Lévy–Itô decomposition. Note that it is convenient to write the d-dimensional Brownian term in the form: $B_A(t) = \sigma B(t)$ [8,22], where $B(t)$ is a d'-dimensional standard Brownian motion and σ is a $d \times d'$ nonzero matrix for which $A = \sigma\sigma^T$. For simplicity, we consider the Brownian term as a standard Brownian motion here, i.e., we set $A = I$. Then we obtain, for $1 \leqslant i \leqslant 2n$, $t \geqslant 0$,

$$\begin{aligned} dX_t^i = {} & V_0^i(X_t)dt + \sum_{k=1}^{d} \gamma^k V_k^i(X_t)dt + \sum_{k=1}^{d} V_k^i(X_t)dB^k(t) \\ & + \frac{1}{2}\sum_{k=1}^{d} V_k \cdot \nabla V_k^i(X_t)dt \\ & + \int_{|z|<1} [\phi^i(z)(X_{t-}) - X_{t-}^i]\tilde{N}(dt, dz) \\ & + \int_{|z|\geqslant 1} [\phi^i(z)(X_{t-}) - X_{t-}^i]N(dt, dz) \\ & + \int_{|z|<1} [\phi^i(z)(X_{t-}) - X_t^i - \sum_{k=1}^{d} z^k V_k^i(X_{t-})]\nu(dz)dt. \end{aligned} \tag{2.5}$$

Denote by $\hat{D}V(x)$ the vector in M whose ith component is $\max_{1\leqslant k\leqslant d} |V_k \cdot \nabla V_k^i(x)|$ for $1 \leqslant i \leqslant 2n, x \in M$. We make the following assumptions.

A1. *The vector field V_0 is locally Lipschitz and the vector fields V_k $(k = 1, 2, \ldots, d)$ and $\hat{D}V(x)$ are globally Lipschitz in the following sense:*
(i) For any $x \in M$, there exists a neighborhood M_0 of x such that $V_0|_{M_0}$ is Lipschitz continuous, i.e. there is a constant $N_1(M_0) > 0$ such that,

$$|V_0(x_1) - V_0(x_2)| \leqslant N_1|x_1 - x_2|, \ x_1, x_2 \in M_0.$$

(ii) There is a constant $N_2 > 0$ such that,

$$\max_{1\leqslant k\leqslant d} |V_k(x_1) - V_k(x_2)|^2 + |\hat{D}V(x_1) - \hat{D}V(x_2)|^2 \le N_1|x_1 - x_2|^2, \\ x_1, x_2 \in M.$$

A2. *One sided linear growth condition: There exists a constant $N_3 > 0$ such that*

$$\sum_{k=1}^{d} V_k^2(x) + 2x \cdot V_0(x) \le N_3(1 + |x|^2), \ x \in M.$$

Theorem 2.1. *Under assumptions* **A1** *and* **A2**, *there exists a unique global solution to (2.5), and the solution process is adapted and càdlàg.*

Proof. This follows immediately from [25, Theorem 3.1] and [22, Lemma 6.10.3]; see also [38]. □

3. Symplecticity and stochastic Hamilton's principle

In this section we present several facts about the stochastic Hamiltonian system with Lévy noise, such as the property of preserving symplectic structure and stochastic Hamilton's principle, which will help us to better understand such systems from the viewpoint of geometry and physics and further allow us in the next sections to confine our studies to its special structure.

3.1. Preservation of symplectic structure

Phase flows of both deterministic Hamiltonian systems and stochastic Hamiltonian systems with Brownian noise are known to preserve symplectic structure [1,5,9]. We next show that stochastic Hamiltonian systems with Lévy noise in the Marcus sense also have this intrinsic property.

Keeping in mind that Marcus integral satisfies the change of variable formula [33, Section 4], for simplicity, we rewrite systems (2.1) in their canonical coordinates. That is, with $X = (Q, P)$, $X_0 = (q, p)$, $V = (\frac{\partial H}{\partial P}, -\frac{\partial H}{\partial Q})$ and $V_k = (\frac{\partial H_k}{\partial P}, -\frac{\partial H_k}{\partial Q})$, $k = 1, \ldots, d$, canonical stochastic Hamiltonian systems with Lévy noise are

$$dQ = \frac{\partial H}{\partial P}(Q, P)dt + \sum_{k=1}^{d} \frac{\partial H_k}{\partial P}(Q, P) \diamond dL^k(t), \ \ Q(t_0) = q, \tag{3.1}$$

$$dP = -\frac{\partial H}{\partial Q}(Q, P)dt - \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q}(Q, P) \diamond dL^k(t), \ \ P(t_0) = p. \tag{3.2}$$

Note that $dp \wedge dq = \sum_{i=1}^{n} dp_i \wedge dq_i$ determines a differential two-form. We are interested in systems (3.1)–(3.2) such that the transformation $(p, q) \to (P, Q)$ preserves symplectic structure as follows:

$$dP \wedge dQ = dp \wedge dq, \\ \text{i.e.,} \quad \sum_{i=1}^{n} dP_i \wedge dQ_i = \sum_{i=1}^{n} dp_i \wedge dq_i. \tag{3.3}$$

To avoid confusion, we should note that the differentials in (3.1)–(3.2) and (3.3) have different meanings: In (3.1)–(3.2), P, Q are treated as functions of time and p, q are fixed parameters, while, in (3.3), the differentiation is made with respect to the initial data p, q.

Geometrically, (3.3) means that the sum of the oriented areas of projections is an integral invariant [1,16]. Consequently, for such systems, all exterior powers of the two-form are also invariant, and the case of nth exterior power gives the preservation of phase volume.

Theorem 3.1 (*Symplecticity*)**.** *The stochastic Hamiltonian system (3.1)–(3.2) preserves symplectic structure.*

The proof of this theorem is based on the differential transformation in the sense of Marcus. It is given in the Appendix.

3.2. Stochastic Hamilton's principle with Lévy noise

For conservative mechanical systems, the classical Hamilton's principle asserts that the dynamics of systems are determined by a variational problem for Lagrangian, and it gives a relationship between the Euler–Lagrange equation and the action integral of the motion [1]. For the situation of nonconservative mechanical systems, the form of the action integral and that of the Euler–Lagrange equation must be changed [6,17]. In this subsection, we would like to propose a stochastic version of Hamilton's principle for a stochastic Hamiltonian system with Lévy noise by viewing it as a special nonconservative system.

We recall some results of nonconservative mechanical systems at first. Let $\mathbf{F}$ be a nonconservative generalized force. The work done by this nonconservative generalized force is defined as

$$W = -\mathbf{F}\cdot\mathbf{r}, \tag{3.4}$$

where $\mathbf{r} = \mathbf{r}(q,t)$ being a position vector. As a nonconservative generalized force is independent of generalized configuration q, the variation of W satisfies

$$\delta W = \mathbf{F}\cdot\delta\mathbf{r} = \mathbf{F}\cdot\frac{\partial \mathbf{r}}{\partial q}\delta q.$$

Let $L(q,\dot{q},t)$ be a Lagrangian with respect to original conservative Hamiltonian system, and it is connected with Hamiltonian H through the equation

$$L = p\cdot\dot{q} - H, \tag{3.5}$$

where $p = \frac{\partial L}{\partial \dot{q}}$ is the Legendre transform. Consider $\gamma = \{q(t) : t_0 \leqslant t \leqslant t_1\}$ as a temporally parameterized curve in the configuration space. Under the influence of $\mathbf{F}$, the action integral of this curve is defined by

$$\mathcal{S}[\gamma] = \int_{t_0}^{t_1} (L(\gamma(t),\dot{\gamma}(t),t) - W(\gamma(t)))dt. \tag{3.6}$$

Hamilton's principle of nonconservative mechanical systems asserts that $\delta\mathcal{S} = 0$ is equal to the following Euler–Lagrange equation holds:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \mathbf{F}\cdot\frac{\partial \mathbf{r}}{\partial q}. \tag{3.7}$$

Here the Lagrangian L is considered as a function with independent variables q, $\dot{q}$ and t.

It is known to [6] that the Euler–Lagrange equations of motion have the property of redundancy. As the value of Lagrangian is invariant to variable transformations, Lagrangian L can be transformed from the variable set $\{q\}$ to a redundant variable set $\{Q^*, P^*\}$ by

$$\begin{aligned} L(q,\dot{q},t) &= L(q(Q^*,P^*,t),\dot{q}(Q^*,P^*,\dot{Q^*},\dot{P^*},t),t) \\ &= L(Q^*,P^*,\dot{Q^*},\dot{P^*},t). \end{aligned}$$

With generalized independent variables Q^*, P^*, $\dot{Q^*}$, $\dot{P^*}$ and t, the generalized Euler–Lagrange equations of motion can be represented as,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{P^*}} - \frac{\partial L}{\partial P^*} = \mathbf{F}\cdot\frac{\partial \mathbf{r}}{\partial P^*}, \tag{3.8}$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{Q^*}} - \frac{\partial L}{\partial Q^*} = \mathbf{F}\cdot\frac{\partial \mathbf{r}}{\partial Q^*} \tag{3.9}$$

with the position vector $\mathbf{r} = \mathbf{r}(Q^*,P^*,t)$. Based on (3.8)–(3.9), for a nonconservative system with nonconservative force $\mathbf{F}$, the corresponding generalized Hamiltonian equations take the following form [6]

$$\dot{Q^*} = \frac{\partial H}{\partial P^*} - \frac{\partial \mathbf{r}}{\partial P^*}\cdot\mathbf{F}, \tag{3.10}$$

$$\dot{P^*} = -\frac{\partial H}{\partial Q^*} + \frac{\partial \mathbf{r}}{\partial Q^*}\cdot\mathbf{F}. \tag{3.11}$$

Lévy noise as a kind of random fluctuating force, can be treated as a special nonconservative force [14,22]. We rewrite a stochastic Hamiltonian system with Lévy noise (3.1)–(3.2) in the following form

$$\dot{Q} = \frac{\partial H}{\partial P} + \frac{\partial \bar{H}}{\partial P}\diamond \dot{L}(t), \tag{3.12}$$

$$\dot{P} = -\frac{\partial H}{\partial Q} - \frac{\partial \bar{H}}{\partial Q}\diamond \dot{L}(t). \tag{3.13}$$

where $\bar{H} = (H_1, H_2, \ldots, H_d)$. It is natural to compare (3.10)–(3.11) with (3.12)–(3.13). Formally, the associations between $\mathbf{F}$ and $\dot{L}(t)$, as well as $\mathbf{r}$ and $-\bar{H}$ are reasonable. Under this consideration, we can thus view stochastic Hamiltonian systems with Lévy noise as a special class of nonconservative system. In other words, stochastic Hamiltonian systems with Lévy noise are Hamiltonian systems in certain generalized sense, which are disturbed by certain nonconservative force (i.e., Lévy noise).

It should be noted that the random fluctuating force here, i.e. Lévy noise, is different from usual nonconservative forces which dissipate energy of the system. Lévy noise may also 'add' energy to the system. To illustrate this point, we consider the following linear stochastic oscillator.

Example 3.1 (*Linear Stochastic Oscillator with Lévy Noise*)**.**

$$dx = ydt, \qquad x(t_0) = x_0, \tag{3.14}$$

$$dy = -xdt - \sigma dL_t, \quad y(t_0) = y_0. \tag{3.15}$$

which is a stochastic Hamiltonian system with $H(x,y) = \frac{1}{2}(x^2 + y^2)$ and $H_1(x,y) = \sigma y$ ($\sigma > 0$ is a constant). Rewrite it in 2-dimensional vector form and multiply both sides with the integrating factor e^{tJ}, where $J = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$. It is not hard to show that this equation has the unique solution

$$x(t) = x(0)\cos t + y(0)\sin t + \int_0^t \sigma\sin(t-s)dL_s, \tag{3.16}$$

$$y(t) = -x(0)\sin t + y(0)\cos t + \int_0^t \sigma\cos(t-s)dL_s. \tag{3.17}$$

For simplicity, we take the initial conditions $x_0 = 1$, $y_0 = 0$ and the drift of Lévy motion $\gamma = 0$. In the sense of Lévy–Itô decomposition, solution (3.16)–(3.17) involves a 'large jumps' term. By using interlacing [22, Page 365], it makes sense to begin by omitting this term and concentrate on the study of the corresponding interlacing solution

$$\begin{aligned} x(t) = {} & \cos t + \int_0^t \sigma\sin(t-s)dB_s \\ & + \int_{|z|<1} \sigma z\sin(t-s)\widetilde{N}(ds,dy), \end{aligned} \tag{3.18}$$

$$\begin{aligned} y(t) = {} & -\sin t + \int_0^t \sigma\cos(t-s)dB_s \\ & + \int_{|z|<1} \sigma z\cos(t-s)\widetilde{N}(ds,dy). \end{aligned} \tag{3.19}$$

By Itô isometry and the properties of compensated Poisson integral [22], we can find that the second moment of this solution satisfies

$$\mathbb{E}(x(t)^2 + y(t)^2) = 1 + \sigma^2 t + \sigma^2 t\int_{|z|<c} |z|^2\nu(dz), \tag{3.20}$$

where $\int_{|z|<c} |z|^2\nu(dz) < \infty$ by the definition of Lévy motion.

It means that the Hamiltonian here grows linearly with respect to time t. This is quite different from the case of deterministic Hamiltonian systems, for which the Hamiltonian is preserved for all t.

Remark 3.1. An alternative view of stochastic Hamilton system is that we can regard it as an open Hamiltonian system within the external world: the stochastic part in (2.1) characterizes the complicated interaction between the "deterministic" Hamiltonian system with the Hamiltonian H_0 and the chaotic environment [12].

For stochastic Hamiltonian system with Lévy noise (3.12)–(3.13), according to (3.4), the work done by Lévy noise is formally

$$W_{stoch} = -\sum_{k=1}^{d} H_k \diamond \dot{L}^k(t). \tag{3.21}$$

Based on (3.6), we infer the action integral of motion as follows

$$\begin{aligned}\mathcal{S}_{stoch}[\gamma] = \int_{t_0}^{t_1} (L - W_{stoch})dt = \int_{t_0}^{t_1} L(\gamma(t), \dot{\gamma}(t), t)dt \\ - \sum_{k=1}^{d} \int_{t_0}^{t_1} H_k(\gamma(t), t) \diamond dL^k(t),\end{aligned} \tag{3.22}$$

where $\gamma = \{(Q(t), P(t)) : t_0 \leqslant t \leqslant t_1\}$.

Moreover, by (3.8)–(3.9), the Euler–Lagrange equations of motion for the stochastic Hamiltonian system with Lévy noise (3.12)–(3.13) have the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{P}} - \frac{\partial L}{\partial P} = \sum_{k=1}^{d} \frac{\partial H_k}{\partial P} \diamond \dot{L}^k(t), \tag{3.23}$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{Q}} - \frac{\partial L}{\partial Q} = \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q} \diamond \dot{L}^k(t). \tag{3.24}$$

We call $\mathcal{S}_{stoch}$ the stochastic action integral and call (3.23)–(3.24) the stochastic Euler–Lagrange equations.

Theorem 3.2 (*Hamilton's Principle*)**.** *The paths that are realized by the stochastic dynamical system represented by stochastic Euler–Lagrange equations (3.23)–(3.24) are those for which the stochastic action integral (3.22) is stationary for fixed endpoints $\gamma(t_0) = (Q_0, P_0)$ and $\gamma(t_1) = (Q_1, P_1)$.*

Proof. The action $\mathcal{S}_{stoch}[\gamma]$ is stationary if it does not vary when the curve is slightly changed, $\gamma(t) \to \gamma(t) + \delta\gamma(t)$. The change in the action upon doing this can be formally expanded in $\delta\gamma$,

$$\mathcal{S}_{stoch}[\gamma + \delta\gamma] - \mathcal{S}_{stoch}[\gamma] = \int_{t_0}^{t_1} \frac{\delta \mathcal{S}_{stoch}}{\delta\gamma} \delta\gamma(t)dt + o(\delta\gamma), \tag{3.25}$$

where $\delta\mathcal{S}_{stoch}/\delta\gamma$ is called the Fréchet or functional derivative of $\mathcal{S}_{stoch}$.

Applying the chain rule for the Marcus integral, we calculate the derivative,

$$\begin{aligned}\delta\mathcal{S}_{stoch} &= \int_{t_0}^{t_1} (\frac{\partial L}{\partial Q}\delta Q + \frac{\partial L}{\partial P}\delta P + \frac{\partial L}{\partial \dot{Q}}\delta\dot{Q} + \frac{\partial L}{\partial \dot{P}}\delta\dot{P})dt \\ &\quad - \sum_{k=1}^{d} \int_{t_0}^{t_1} (\frac{\partial H_k}{\partial Q}\delta Q + \frac{\partial H_k}{\partial P}\delta P) \diamond \dot{L}^k(t)dt \\ &= \left[\frac{\partial L}{\partial \dot{Q}}\delta Q\right]_{t_0}^{t_1} + \left[\frac{\partial L}{\partial \dot{P}}\delta P\right]_{t_0}^{t_1} + \int_{t_0}^{t_1} (\frac{\partial L}{\partial Q} \\ &\quad - \frac{d}{dt}\frac{\partial L}{\partial \dot{Q}} - \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q} \diamond \dot{L}^k(t))\delta Q dt \\ &\quad + \int_{t_0}^{t_1} (\frac{\partial L}{\partial P} - \frac{d}{dt}\frac{\partial L}{\partial \dot{P}} - \sum_{k=1}^{d} \frac{\partial H_k}{\partial P} \diamond \dot{L}^k(t))\delta P dt.\end{aligned}$$

The boundary terms vanish because the endpoints of $\gamma(t)$ are fixed: $\delta Q(t_0) = \delta Q(t_1) = \delta P(t_0) = \delta P(t_1) = 0$. As discussed in Wang et al. [17], the desired result follows. □

Example 3.2. Consider the linear stochastic oscillators with Lévy noise (3.14)–(3.15). We show that Eqs. (3.14)–(3.15) are equivalent to the stochastic Euler–Lagrange equations of motion with Lévy noise (3.23)–(3.24). Indeed, by the relation between Lagrangian and Hamiltonian, we have

$$L(x, y, \dot{x}, \dot{y}) = x \cdot \dot{y} - H(x, y) = x \cdot \dot{y} - \frac{1}{2}(y^2 + x^2).$$

According to (3.23)–(3.24), the Euler–Lagrange equations of motion of the linear stochastic oscillators have the form

$$\begin{cases} \dfrac{d}{dt}\dfrac{\partial L}{\partial y} - \dfrac{\partial L}{\partial x} = -\sigma \dot{L}_t, \\ \dfrac{d}{dt}\dfrac{\partial L}{\partial \dot{y}} - \dfrac{\partial L}{\partial y} = 0. \end{cases} \tag{3.26}$$

since $H_1 = \sigma x$. With initial conditions $x(0) = x_0$, $y(0) = y_0$, (3.26) are equivalent to the Hamiltonian equations of motion (3.14)–(3.15).

Consider the stochastic action integral $\mathcal{S}$ in (3.22) as a function of the two endpoints $(Q(t_0), \dot{Q}(t_0)) = (Q_0, \dot{Q}_0)$ and $(Q(t_1), \dot{Q}(t_1)) = (Q_1, \dot{Q}_1)$. We have the following theorem which plays an important role in constructing some numerical methods [15–17,39].

Theorem 3.3 (*Characterization of Stochastic Action Integral*)**.** *The stochastic action integral $\mathcal{S}_{stoch}$ satisfies*

$$d\mathcal{S}_{stoch} = -P_0^T dQ_0 + P_1^T dQ_1. \tag{3.27}$$

Furthermore, if the Lagrangian L and the functions H_k ($k = 1, \dots, d$) are sufficiently smooth with respect to P and Q, then the mapping

$$(P_0, Q_0) \mapsto (P_1, Q_1)$$

defined by Eq. (3.27) is symplectic.

The proof is given in the Appendix.

4. An averaging principle for integrable stochastic Hamiltonian systems

We now return to the stochastic Hamiltonian systems with Lévy noise (2.1) on a $2n$-dimensional smooth manifold M (for simplicity, set $n = d$ in the rest of this discussion). As mentioned earlier, such systems are themselves nonconservative systems with the perturbation of Lévy noise. Then an interesting question to raise is: if there is even a small external perturbation in this stochastic system, just as the deterministic Hamiltonian case and the stochastic Hamiltonian case with Brownian noise referring to the study of Freidlin–Wentzell [7], Li [19] and so on, what the effective dynamic behavior would be? To answer this question, we consider the (completely) integrable stochastic Hamiltonian systems with Lévy noise.

Recall that on a $2d$-dimensional smooth manifold, a family of d smooth Hamiltonians $\{H_k\}_{k=1}^d$ is said to form a (completely) integrable system if they are pointwise Poisson commuting and if the corresponding Hamiltonian vector fields V_k are linearly independent at almost all points.

We call systems (2.1) *(completely) integrable stochastic Hamiltonian systems with Lévy noise*, if they satisfy the following condition:

- **A3** *Completely integrability: $\{H_k\}_{k=1}^d$ is an integrable family, and Hamiltonian vector field V_0 with Hamiltonian H_0 is commuting with the family of vector fields V_k. That is, $dH_j(V_i) = \omega^2(V_i, V_j) = 0$ for $i, j = 0, 1, 2, \dots, d$.*

For the sake of convenience and readability, in the sense of Lévy–Itô decomposition and Marcus integral (2.3), we consider the following integrable stochastic Hamiltonian system with Lévy noise, which satisfies assumptions **A1**–**A3**,

$$
\begin{aligned}
dX_t &= V_0(X_t)dt + \sum_{k=1}^{d} V_k(X_t)\circ dB^k(t) + \sum_{k=1}^{d} V_k(X_t)\diamond dL^k(t),\\
X(t_0) &= x \in M. \qquad (4.1)
\end{aligned}
$$

Where $B(t)$ is a d-dimensional independent standard Brownian motion, $L(t)$ is a d-dimensional independent Lévy motion with the generating triplet $(0, 0, \nu)$ which is a pure jump process.

4.1. Invariant manifolds and invariant measure for integrable stochastic Hamiltonian systems

Due to the system has d first integrals $H_1, \ldots, H_d$ in involution, we consider the joint integral level

$$M_h = \{x \in M : H_i(x) = h_i = const,\ i = 1, 2, ..d\}. \qquad (4.2)$$

The Liouville–Arnold theorem [1] indicates that if the functions H_i on M_h are independent, then each compact connected component of M_h is diffeomorphic to a d-dimensional torus $\mathbb{T}^d$. It remains to use the geometric fact: in this integrable system there are convenient, so-called, action–angle coordinates (I, θ) (I are the actions and θ are the angles) such that $\omega^2 = dI \wedge d\theta$ (symplecticity), $H = H(I)$ (i.e., I are first integrals).

We next show that a solution to these SDEs preserves the energies H_i and there are corresponding invariant manifolds (level sets). Let $\Psi_t := (\Psi(t, \omega, x), t \geqslant 0)$ be the solution flow of the SDE (4.1) with starting point x and $(T_t, t \geqslant 0)$ be the semigroup associated with Ψ_t. Applying the chain rule for the Stratonovich integral and Marcus integral, and using the assumption **A3** of completely integrability, we have

Lemma 4.1. *The solution flow $\Psi_t := (\Psi(t), t \geqslant 0)$ of SDE (4.1) preserves the invariant manifolds M_h, i.e. for $1 \leqslant i \leqslant d$,*

$$
\begin{aligned}
dH_i(X_t) =& dH_i(V_0(X_t))dt + \sum_{k=1}^{d} dH_i(V_k(X_t))\circ dB^k(t)\\
&+ \sum_{k=1}^{d} dH_i(V_k(X_t))\diamond dL^k(t) = 0.
\end{aligned}
$$

Indeed, for each x in M, we have $h = (H_1(x), \ldots, H_n(x))$, thus it determines an invariant manifold, which we write also as $M_{H(x)}$. Note that the d vector fields $\{V_k\}_{k=1}^{d}$ are tangent to $M_{H(x)}$ and the symplectic form ω^2 vanishes on the invariant manifolds M_h. The Markovian solution to SDEs (4.1) restricts to each invariant manifold and the generator $\mathcal{A}$ of restriction is the sum of a second-order elliptic differential operator and a (compensated) integral of difference operator, i.e.,

$$
\begin{aligned}
(\mathcal{A}f)(x) =& (\mathcal{L}_0 f)(x) + \frac{1}{2}\sum_{k=1}^{d}(\mathcal{L}_k\mathcal{L}_k f)(x)\\
&+ \int_{\mathbb{R}^d\setminus\{0\}} [f(\phi(z)x) - f(x)\\
&- \sum_{k=1}^{d} z^k(\mathcal{L}_k f)(x)\mathbf{1}_{\{|z|<1\}}(z)]\nu(dz) \qquad (4.3)
\end{aligned}
$$

for every function $f \in C_b^2(M)$. Here we denote as $\mathcal{L}_0$, $\mathcal{L}_k$ the Lie differentiation in the direction of V_0, V_k, respectively, and $C_b^2(M)$ the collection of all bounded Borel measurable C^2 functions on M. More precisely, we have $\mathcal{L}f = df(V_0) = \omega^2(v_f, V_0)$ and $\mathcal{L}_k f = df(V_k) = \omega^2(v_f, V_k)$.

We remark that an invariant probability measure for (4.1) is by definition a Borel probability measure on M such that

$$\int_M (T_t g)(x)\mu(dx) = \int_M g(x)\mu(dx)$$

for all $t > 0$, $g \in C^1(M)$. Based on the celebrated Krylov–Bogoliubov method, we have the following lemma.

Lemma 4.2 ([25, Theorem 4.5])**.** *If M is locally compact in the relative topology and assumptions* **A1** *and* **A2** *hold, then the system (4.1) has at least one invariant measure.*

For simplicity, throughout this paper, we assume that:

A4 *The invariant manifolds are compact, the map $H : x \in M \to (H_1(x), \ldots, H_d(x)) \in \mathbb{R}^d$ is proper, and its set of critical points has measure zero.*

Under our assumption, for almost every point h_0 in $\mathbb{R}^d$, there is a neighborhood N of h_0 such that $H^{-1}(h)$ is a smooth submanifold for all $h \in N$ and that there is a diffeomorphism from $H^{-1}(N)$ to $N \times H^{-1}(h_0)$. We call such h_0 a regular value of H, and call the point y in M a critical point if $H(y)$ is not regular. By Morse–Sard theorem [40], the set of critical values of the function H has measure zero.

Recall that in a neighborhood of a regular point h_0 of H, every component of the level set M_{h_0} is diffeomorphic to a âĂÍd-dimensional torus $\mathbb{T}^d$, and a small neighborhood U_0 of M_{h_0} is diffeomorphic to the product space $\mathbb{T}^d \times D$, where D is a relatively compact open set in $\mathbb{R}^d$. Take an action–angle chart around M_h. The measure $(\sum_i dI^i \wedge d\theta^i)^d$ on the product space naturally splits to give us a probability measure, the Haar measure [40] $\theta_1 \wedge \cdots \wedge \theta^d$ on $\mathbb{T}^d$. We take the corresponding one on M_h and denote it by μ_h, just like the case of Brownian in [19]. With the help of action–angle transformation and the above assumptions, we thus have the following lemma.

Lemma 4.3. *Assume that assumptions* **A1**–**A4** *are in force. Let $E = span\{V_1, \ldots, V_d\}$ be a sub-bundle of the tangent bundle of rank d. Let U be a section of E commuting with all V_i ($1 \leqslant i \leqslant d$). The invariant measure for stochastic Hamiltonian system (4.1) restricted to the invariant manifold M_h is μ_h, which varies smoothly with h in sufficiently small neighborhoods of a regular value.*

Proof. Recall that M_h have the form in (4.2), we rewrite $U = \sum_{i=1}^{d} h_i V_i(x)$. For any smooth function f on M_h, we have

$$
\begin{aligned}
\int_{M_h} df(V_i)(x)\mu_h(dx) &= \int_{\mathbb{T}^d} d(f\circ\varphi)\Big(-\sum_{k=1}^{d}\frac{\partial(H_k\circ\varphi)}{\partial I_k}\frac{\partial}{\partial\theta_k}\Big)d\theta\\
&= -\sum_{k=1}^{d}\omega_k^i(I)\int_{\mathbb{T}^d}\Big(\frac{\partial}{\partial\theta_k}(f\circ\varphi)\Big)d\theta = 0,
\end{aligned}
$$

where φ^{-1} is the action–angle coordinate map (see the next subsection for details), (I, θ) are the corresponding action–angle coordinates. Thus U is divergence free, i.e. $\mathrm{div}_E U = 0$, in the sense of

$$\int_{M_h} df(U)(x)\mu_h(dx) = -\int_{M_h}\mathrm{div}_E U(x)\mu_h(dx) = 0. \qquad (4.4)$$

Therefore, restricted to the torus, the invariant measure of SDE (4.1) is the same as that of the corresponding SDE without a drift. From the action–angle transformation we find that the measure μ_h is the desired object. $\square$

4.2. The perturbed system and statement of an averaging principle

We next study the situation where an integrable stochastic Hamiltonian system is perturbed by a transversal smooth vector field and the stochastic differentials. Let y_0 be a regular point of H in M with a neighborhood U_0 the domain of an action–angle coordinate map:

$\varphi^{-1}: U_0 \to \mathbb{T}^d \times D$

where $\mathbb{T}^d$ is a d-dimensional torus and D is a relatively compact open set of $\mathbb{R}^n$. Note that the action coordinate of a point $x \in U_0$ can be denoted with the help of the projection $\pi : U_0 \to D$ by $\varphi^{-1}(x) = (\theta^*, \pi(x))$ for some $\theta^* \in \mathbb{T}^d$. We consider the perturbed system corresponding to (4.1):

$$\begin{aligned} dY_t^\varepsilon =& V_0(Y_t^\varepsilon)dt + \sum_{k=1}^d V_k(Y_t^\varepsilon) \circ dB_t^k + \sum_{k=1}^d V_k(Y_t^\varepsilon) \diamond dL^k(t) \\ &+ \varepsilon\Big(K(Y_t^\varepsilon)dt + \sum_{k=1}^d F_k(\pi(Y_t^\varepsilon)) \circ d\tilde{B}_t^k + \sum_{k=1}^d G_k(\pi(Y_t^\varepsilon)) \diamond d\tilde{L}_t^k\Big) \end{aligned} \tag{4.5}$$

with initial condition $Y_0^\varepsilon = y_0$. Where K is a smooth and global Lipschitz continuous vector field, transversal in the sense that $\omega^2(V_k, K)$, $k = 0, 1, \ldots, d$, are not all identically zero; $\tilde{B}(t)$ is a d-dimensional independent standard Brownian motion; $\tilde{L}(t)$ is a d-dimensional independent pure jump Lévy motion with the generating triplet $(0, 0, \nu')$. Moreover, F, G are smooth vector fields such that F, $\hat{D}F$, G and $\hat{D}G$ are globally Lipschitz continuous.

We denote by Y_t^ε the solution to (4.5) and by $X_t = Y_t^0$ the solution to (4.1) with initial value y_0. In the action–angle coordinate, $X_t = \varphi(\theta_t, I_t)$, $\theta \in \mathbb{T}^d$, $I \in D$ and $Y_t^\varepsilon = \varphi(\theta_t^\varepsilon, I_t^\varepsilon)$, $\theta^\varepsilon \in \mathbb{T}^d$, $I^\varepsilon \in D$. Let $\tilde{H}_k = H_k(\varphi(\theta_t, I_t))$ be the induced Hamiltonian on $\mathbb{T}^d \times D$, then, for $i = 1, \ldots, d$,

$$\dot{\theta}_k^i = \frac{\partial \tilde{H}_k}{\partial I_i} =: \omega_k^i(I),$$

$$\dot{I}_k^i = -\frac{\partial \tilde{H}_k}{\partial \theta_i} = 0,$$

with ω_k^i smooth functions. Indeed, the corresponding induced Hamiltonian vector field $\tilde{V}_k := V_{\tilde{H}_k} = -\sum_{i=1}^d (\partial(H_k \circ \varphi)/\partial I_i)(\partial/\partial\theta_i)$.

For the perturbed SDE (4.5), we write the induced perturbation vector field of K as (K_θ, K_I) on $\mathbb{T}^d \times D$ with $K_\theta = (K_\theta^1, \ldots, K_\theta^d)$ and $K_I = (K_I^1, \ldots, K_I^d)$ the angle and action component, respectively, and we do the same thing for F and G. By the chain rule for Stratonovitch integral as well as that for Marcus integral, we have the following form of the SDE on $\mathbb{T}^d \times D$:

$$\begin{aligned} d\theta_t^\varepsilon =& \omega_0(I_t^\varepsilon)dt + \sum_{k=1}^d \omega_k(I_t^\varepsilon) \circ dB_t^k + \sum_{k=1}^d \omega_k(I_t^\varepsilon) \diamond dL^k(t) \\ &+ \varepsilon\Big(K_\theta(\theta_t^\varepsilon, I_t^\varepsilon)dt + \sum_{k=1}^d F_{\theta,k}(I_t^\varepsilon) \circ d\tilde{B}_t^k + \sum_{k=1}^d G_{\theta,k}(I_t^\varepsilon) \diamond d\tilde{L}_t^k\Big), \end{aligned} \tag{4.6}$$

$$dI_t^\varepsilon = \varepsilon\Big(K_I(\theta_t^\varepsilon, I_t^\varepsilon)dt + \sum_{k=1}^d F_{I,k}(I_t^\varepsilon) \circ d\tilde{B}_t^k + \sum_{k=1}^d G_{I,k}(I_t^\varepsilon) \diamond d\tilde{L}_t^k\Big). \tag{4.7}$$

Note that subjected to a small perturbation, the system splits into two parts with fast rotation along the nonperturbed trajectories and slow motion across them, so it is a situation where the averaging principle is to be expected to hold.

For this purpose, we further adopt the following assumptions:

A5 *There is a constant $p \geqslant 2$ such that the Lévy measures ν (of L_t) and $\tilde{\nu}$ (of $\tilde{L}_t$) satisfy*

$$\int_{\mathbb{R}^d} |z|^p \nu(dz) < \infty, \quad \text{and} \quad \int_{\mathbb{R}^d} |z|^{2p} \tilde{\nu}(dz) < \infty.$$

A6 *For any continuous function f on the compact manifold converging to infinity when t converges to infinity, $\frac{1}{t}\int_s^{s+t} f(X_r)dr \to \int_{M_h} f(z)\mu_h(z)$ when $t \to \infty$, in L^p ($p \geqslant 2$), and the rate of convergence, denoted by $\eta(t)$, is a positive, bounded, decreasing function from $[0, \infty)$ to $[0, \infty)$ with $\eta(t) \searrow 0$ as $t \to \infty$.*

Some comments on these two assumptions have to be made: Note that the invariant manifold here is actually d-dimensional torus, which is compact and bounded. It is necessary and reasonable to put forward assumption **A5** referring to [29]. This assumption indicates the polynomial moments of $L(t)$ and $\tilde{L}(t)$ exist, and will play an important role in estimating some terms of the Marcus equation in the next subsection. Note that the motion on the torus, which would be quasi-periodic if there are no diffusion terms, is ergodic. Indeed, there is no standard rate of convergence for general Markovian systems in the ergodic theorem; see e.g. Krengel [41], Kakutani and Petersen [42]. It is natural to deal with an averaging principle in the terms of the function η following the approach in Freidlin–Wentzell [7]. We thus have the ergodicity assumption **A6**. More information on rates of convergence for Lévy noise driven systems can be found in Kulik [43] and Högele–de Costa [29], and a detailed example will be shown in subsection 4.4.

To study slow motion governed by the transversal part of the vector field K and the stochastic differentials $F \circ \dot{\tilde{B}}_t$, $G \diamond \dot{\tilde{L}}_t$, it is convenient to rescale the time, see Lemma 4.4 for detail. Denote $Y_{t/\varepsilon}^\varepsilon$ the process scaled in time by $1/\varepsilon$ which coincides, in the sense of probability distributions [7], with Y_t^ε. Then, the evolution of $Y_{t/\varepsilon}^\varepsilon$ is the skew product of the fast diffusion of order $\frac{1}{\varepsilon}$ along the invariant manifold and the slow diffusion of order 1 across the invariant manifold. We finally obtain a new dynamical system in the limit as ε goes to zero: Compared with the motion in the transversal direction, the motion along the torus is significantly faster, thus as the randomness in the fast component is averaged out by the induced invariant measure, the evolution of the action component of $Y_{t/\varepsilon}^\varepsilon$ will have a limit.

The main theorem on averaging principle for (completely) integrable stochastic Hamiltonian system is formulated below, and the detailed proof is shown in next subsection.

Theorem 4.1 (*Averaging Principle*)**.** *Consider the perturbed SDE (4.5) with initial value $Y_0^\varepsilon = y_0$ and satisfying assumptions **A1**–**A6** for some $p \geqslant 2$. Set $H_i^\varepsilon(t) = H_i(Y_{t/\varepsilon}^\varepsilon)$, for $i = 1, 2, \ldots, d$. Define exit time $\tau^\varepsilon := \inf\{t \geqslant 0 : Y_{t/\varepsilon}^\varepsilon \notin U_0\}$ as the first time that the solution $Y_{t/\varepsilon}^\varepsilon$ exits from U_0.*

Let $\bar{H}(t) = (\bar{H}_1(t), \ldots, \bar{H}_d(t))$ be the solution to the following system of d deterministic differential equations

$$\begin{aligned} d\bar{H}_i(t) =& \int_{M_{\bar{H}(t)}} \omega^2(V_i, K)(\bar{H}(t), z)\mu_{\bar{H}_t}(dz)dt + \sum_{k=1}^d F_k(\bar{H}(t)) \circ d\tilde{B}_t^k \\ &+ \sum_{k=1}^d G_k(\bar{H}(t)) \diamond d\tilde{L}^k(t) \end{aligned} \tag{4.8}$$

with initial value $\bar{H}(0) = H(y_0)$. Define exit time $\tau^0 := \inf\{t \geqslant 0 : \bar{H}(t) \notin U_0\}$ as the first time that $\bar{H}(t)$ exits from U_0.

Then we have that:

(1) For any sufficiently small $\varepsilon > 0$ and $t < \tau_0$, there exist constants $k_1, k_2, k_3 > 0$ such that

$$\left(\mathbb{E}\Big[\sup_{s\leqslant t}|H^\varepsilon(s\wedge\tau^\varepsilon)-\bar{H}(s\wedge\tau^\varepsilon)|^p\Big]\right)^{\frac{1}{p}} \leqslant k_1 t\big(\varepsilon^{1-k_2 t}+\eta(t|\ln\varepsilon|)\big)\exp(k_3 t). \tag{4.9}$$

(2) If there exists a $r > 0$ such that $U_r := \{x \in M : |H(x)-H(y_0)| \leqslant r\} \subset U_0$. Define exit time $\tau_\delta := \inf\{t \geqslant 0 : |\bar{H}_t - H(Y_0)| \geqslant r-\delta\}$ for $\delta > 0$. Then for any $\delta > 0$, constant $k_2 > 0$ given above, and constants k_4, k_5 depending on τ_δ,

$$\mathbb{P}(\tau^\varepsilon < \tau_\delta) \leqslant k_4\delta^{-p}{\tau_\delta}^p\big(\varepsilon^{1-k_2\tau_\delta}+\eta(\tau_\delta|\ln\varepsilon|)\big)^p\exp(k_5\tau_\delta). \tag{4.10}$$

Remark 4.1. This result includes the case of pure Gaussian noise and case of pure jump noise, where the former situation has been considered, cf. Li [19, Theorem 3.3.]. Indeed, Hamiltonian vector V_0 in (4.1) can be weakened to be a locally Hamiltonian vector which is not given by a Hamiltonian function as in [19]. The main difference between Gaussian situation and the situation we considered here comes from the estimation for Lévy noise term. However, if deterministic part of the perturbation is a (local) Hamiltonian vector field with $\omega^2(V_i, K) = 0$, or the multiplicative coefficients of the stochastic differentials are not only depend on the slow component, the situation will become more complex. To deal with these problems on multiplicative Lévy noise is still remain to solve.

4.3. Proof of the averaging principle

In this subsection we always assume that assumptions **A1**–**A6** are in force for some $p \geqslant 2$. We first get the information on the order of which the first integrals for the perturbed system change over a time interval by next lemma.

Lemma 4.4. *Let $\tau^\varepsilon = \inf\{t \geqslant 0 : Y_t^\varepsilon \notin U_0\}$. For any Lipschitz test function $f : M \to \mathbb{R}$ and $p \geqslant 2$, we have*

$$\left[\mathbb{E}\big(\sup_{s\leqslant t\wedge\tau^\varepsilon}|f(Y_s^\varepsilon)-f(X_s)|^p\big)\right]^{\frac{1}{p}} \leqslant C_1\varepsilon e^{C_2 t}, \tag{4.11}$$

where C_1, C_2 are constants depending on the Lipschitz coefficient of f, on the upper bounds of the norms of vector fields K, F, G, V_k, $k = 0, \ldots, d$ and their derivatives with respect to the action–angle coordinate on $T^d \times D$.

Proof. In action–angle coordinates, we rewrite the flows as $X_t = \varphi(\theta_t, I_t)$ and $Y_t^\varepsilon = \varphi(\theta_t^\varepsilon, I_t^\varepsilon)$. And the corresponding SDEs on $T^d\times D$ under the action–angle coordinate map are shown in (4.6)–(4.7). Since D is relatively compact, $\partial(f\circ\varphi)/\partial\theta$ and $\partial(f\circ\varphi)/\partial I$ are bounded on $T^d \times D$. We thus obtain

$$\begin{aligned}|f(Y_t^\varepsilon)-f(X_t)| &= |f\circ\varphi(\theta_t^\varepsilon, I_t^\varepsilon)-f\circ\varphi(\theta_t, I_t)|\\ &\leqslant c_0|(\theta_t^\varepsilon-\theta_t, I_t^\varepsilon-I_t)| \leqslant c_0|\theta_t^\varepsilon-\theta_t|\\ &\quad+c_0|I_t^\varepsilon-I_t|,\end{aligned} \tag{4.12}$$

for some constant $c_0 > 0$.

Estimate of the action component $|I_t^\varepsilon - I_t|$. Note that the facts that Eq. (4.6) satisfies the chain rule in the sense of Stratonovitch and Marcus, and $\langle Dh(x), u\rangle = p|x|^{p-2}\langle x, u\rangle$ for the function $h(x) = |x|^p$, we obtain, for $s < \tau^\varepsilon$,

$$\begin{aligned}
|I_s^\varepsilon - I_s|^p =& \varepsilon p\Bigg(\int_0^s |I_r^\varepsilon - I_r|^{p-2}\big\langle I_r^\varepsilon - I_r, K_I(\theta_t^\varepsilon, I_r^\varepsilon)\big\rangle dr\\
&+\int_0^s |I_r^\varepsilon - I_r|^{p-2}\sum_{k=1}^d\big\langle I_r^\varepsilon - I_r, F_{I,k}(I_r^\varepsilon)\circ dB_r^k\big\rangle\\
&+\int_0^s |I_{r-}^\varepsilon - I_{r-}|^{p-2}\sum_{k=1}^d\Big\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)\diamond d\tilde{L}_r^k\Big\rangle\Bigg)\\
\leqslant& \varepsilon p\int_0^s |I_r^\varepsilon - I_r|^{p-2}\Big|\Big\langle I_r^\varepsilon - I_r, K_I(\theta_t^\varepsilon, I_r^\varepsilon)+\frac{d}{2}\hat{D}_I F_I(I_r^\varepsilon)\Big\rangle\Big|dr \qquad (\Sigma_1)\\
&+\varepsilon p\sum_{k=1}^d\int_0^s |I_r^\varepsilon - I_r|^{p-2}\big|\big\langle I_r^\varepsilon - I_r, F_{I,k}(I_r^\varepsilon)dB_r^k\big\rangle\big| \qquad (\Sigma_2)\\
&+\varepsilon p\sum_{k=1}^d\int_0^s |I_{r-}^\varepsilon - I_{r-}|^{p-2}\Big(\Big|\Big\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)d\tilde{L}_r^k\Big\rangle\Big|\Big) \qquad (\Sigma_3)\\
&+\varepsilon p\sum_{k=1}^d\sum_{0\leqslant r\leqslant s}\int_0^s |I_{r-}^\varepsilon - I_{r-}|^{p-1}dt\big|\phi(\Delta L^k(r),\\
&G_{I,k}(I_{r-}^\varepsilon), I_{r-}^\varepsilon)\\
&- I_{r-}^\varepsilon - G_{I,k}(I_{r-}^\varepsilon)\Delta L^k(r)\big| \qquad (\Sigma_4)\\
=& \Sigma_1+\Sigma_2+\Sigma_3+\Sigma_4,
\end{aligned} \tag{4.13}$$

where $(\hat{D}_I F_I)_i = \max_{1\leqslant k\leqslant d}|F_{I,k}\cdot\nabla_I F_{I,k}^i|$ comes from the Stratonovitch correction. By assumption the induced vector fields and their derivatives are bounded on $T^d \times D$. A direct computation gives

$$\Sigma_1 \leqslant \varepsilon p\big(\sup_{T^d\times D}|K_I|+\frac{d}{2}\sup_{T^d\times D}|\hat{D}_I F_I|\big)\int_0^s |I_r^\varepsilon - I_r|^{p-1}dr. \tag{4.14}$$

Note that the term Σ_3 has the representation with respect to the compensated Poisson random measure $\tilde{N}'$ associated to $\tilde{L}(t)$ [22,32], we have

$$\begin{aligned}
\Sigma_3 =& \varepsilon p\sum_{k=1}^d\int_0^s\int_{\mathbb{R}^d\setminus\{0\}}|I_{r-}^\varepsilon - I_{r-}|^{p-2}\Big|\Big\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)z\Big\rangle\\
&\times\tilde{N}(dr,dz)\Big|\\
&+\varepsilon p\sum_{k=1}^d\int_0^s\int_{|z|>1}|I_{r-}^\varepsilon - I_{r-}|^{p-2}\Big|\Big\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)z\Big\rangle\\
&\times\nu'(dz)dr\Big|\\
\leqslant& \varepsilon p\sum_{k=1}^d\int_0^s\int_{\mathbb{R}^d\setminus\{0\}}|I_{r-}^\varepsilon - I_{r-}|^{p-2}\Big|\Big\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)z\Big\rangle\Big|\\
&\times\tilde{N}(dr,dz)\\
&+\varepsilon pd\sup_{T^d\times D}|G_I|\int_{|z|>1}|z|\nu'(dz)\int_0^s|I_{r-}^\varepsilon - I_{r-}|^{p-1}dr.
\end{aligned} \tag{4.15}$$

As smooth vector fields G_I and $\hat{D}G_I$ are globally Lipschitz continuous. For the last term, by exploiting that $\int_{|z|>1}|z|^4\nu'(dz) < \infty$, we have the following estimation referring to [29, Lemma 3.1]:

$$\begin{aligned}
\Sigma_4 \leqslant& \varepsilon^2 c_1(p, G_I, \hat{D}G_I)\sum_{k=1}^d\sum_{0\leqslant r\leqslant s}|I_{r-}^\varepsilon - I_{r-}|^{p-1}|\Delta\tilde{L}^k(r)|^4\\
\leqslant& \varepsilon^2 c_1\left(\int_0^s\int_{\mathbb{R}^d\setminus\{0\}}|I_{r-}^\varepsilon - I_{r-}|^{p-1}|z|^4\tilde{N}'(dr,dz)\right.\\
&\left.+\int_0^s|I_{r-}^\varepsilon - I_{r-}|^{p-1}dr\right).
\end{aligned} \tag{4.16}$$

Combining the estimates (4.14)–(4.16), we can find that

$$\begin{aligned}|I_s^\varepsilon - I_s|^p \leqslant & c_2\varepsilon \int_0^s |I_{r-}^\varepsilon - I_{r-}|^{p-1}dr + \varepsilon p\sum_{k=1}^d \int_0^s |I_r^\varepsilon - I_r|^{p-2} \\ & \times \left|\left\langle I_r^\varepsilon - I_r, F_{I,k}(I_r^\varepsilon)\right\rangle\right| dB_r^k \\ & + \varepsilon p \sum_{k=1}^d \int_0^s \int_{\mathbb{R}^d\setminus\{0\}} |I_{r-}^\varepsilon - I_{r-}|^{p-2} \\ & \times \left|\left\langle I_{r-}^\varepsilon - I_{r-}, G_{I,k}(I_{r-}^\varepsilon)z\right\rangle\right| \tilde{N}(dr,dz) \\ & + \varepsilon^2 c_3 \int_0^s \int_{\mathbb{R}^d\setminus\{0\}} |I_{r-}^\varepsilon - I_{r-}|^{p-1}|z|^4 \tilde{N}'(dr,dz). \end{aligned} \tag{4.17}$$

In order to calculate estimate of the expectation of the supremum for the equation above, it is natural to use Itô isometry for the Brownian term and use Kunita's first inequality ([22, Page 265]) or other maximal inequality for the compensated Poisson terms. We refer to Högele–da Costa [29] for a standard argument on such a estimate. One difference with [29] is that there is an extra Brownian term here. Indeed, with the help of Itô isometry, we obtain

$$\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon} |\Sigma_2|^2\Big] \leqslant \varepsilon c_4(p,d)\big(\sup_{T^d\times D}|F_I|\big)^2 \int_0^t \mathbb{E}\big[|I_s^\varepsilon - I_s|^{2(p-1)}\big]ds. \tag{4.18}$$

Therefore, the estimate for (4.17) is quite similar to estimate (44) in [29] and yields a constant c_5 such that

$$\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon} |I_s^\varepsilon - I_s|^p\Big] \leqslant c_5\varepsilon^p(1+t^{2p+1}). \tag{4.19}$$

Estimate of the angle component $|\theta_t^\varepsilon - \theta_t|$. For $s<\tau^\varepsilon$, applying the chain rule again, we have

$$\begin{aligned}|\theta_s^\varepsilon - \theta_s|^p \leqslant & \; p\int_0^s |\theta_r^\varepsilon - \theta_r|^{p-2}\left\langle \theta_r^\varepsilon - \theta_r, \omega_0^i(I_r^\varepsilon) - \omega_0^i(I_r)\right\rangle dr & (\Lambda_1) \\ & + p\sum_{k=1}^d \int_0^s |\theta_r^\varepsilon - \theta_r|^{p-2} \\ & \times \left\langle \theta_r^\varepsilon - \theta_r, (\omega_k^i(I_r^\varepsilon) - \omega_k^i(I_r))dB^k(r)\right\rangle & (\Lambda_2) \\ & + p\sum_{k=1}^d \int_0^s |\theta_r^\varepsilon - \theta_r|^{p-2} \\ & \times \left\langle \theta_r^\varepsilon - \theta_r, (\omega_k^i(I_r^\varepsilon) - \omega_k^i(I_r))dL^k(r)\right\rangle & (\Lambda_3) \\ & + p\sum_{k=1}^d \sum_{0\leqslant r\leqslant s} \int_0^s |\theta_r^\varepsilon - \theta_r|^{p-1}\big|\phi(\Delta L^k(r), \omega_k^i(I_{r-}^\epsilon), I_{r-}^\epsilon) \\ & - \phi(\Delta L^k(r), \omega_k^i(I_{r-}), I_{r-}) \\ & - \big(I_{r-}^{\epsilon,i} - I_{r-}^i\big) - \big(\omega_k^i(I_{r-}^\epsilon) - \omega_k^i(I_{r-})\big)\Delta L^k(r)\big| & (\Lambda_4) \\ & + \varepsilon p\int_0^s |\theta_r^\varepsilon - \theta_r|^{p-2}\left|\left\langle \theta_r^\varepsilon - \theta_r, K_\theta(\theta_t^\varepsilon, I_r^\varepsilon)\right\rangle\right| dr & (\Lambda_5) \\ & + \varepsilon p\sum_{k=1}^d \int_0^s |\theta_r^\varepsilon - \theta_r|^{p-2}\left|\left\langle \theta_r^\varepsilon - \theta_r, F_{\theta,k}(I_r^\varepsilon)dB_r^k\right\rangle\right| & (\Lambda_6) \\ & + \varepsilon p\sum_{k=1}^d \int_0^s |\theta_{r-}^\varepsilon - \theta_{r-}|^{p-2}\left|\left\langle \theta_{r-}^\varepsilon - \theta_{r-}, G_{\theta,k}(I_{r-}^\varepsilon)d\tilde{L}_r^k\right\rangle\right| & (\Lambda_7) \\ & + \varepsilon p\sum_{k=1}^d \sum_{0\leqslant r\leqslant s} \int_0^s |I_{r-}^\varepsilon - I_{r-}|^{p-1}dt\big|\phi(\Delta L^k(r), \\ & G_{\theta,k}(I_{r-}^\varepsilon), I_{r-}^\varepsilon) \\ & - I_{r-}^\varepsilon - G_{\theta,k}(I_{r-}^\varepsilon)\Delta L^k(r)\big| & (\Lambda_8) \\ = & \Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4+\Lambda_5+\Lambda_6+\Lambda_7+\Lambda_8. & (4.20)\end{aligned}$$

Here we can replace the Stratonovitch integrations by Itô integrations, as both $\omega_k(I)$ and $F_\theta(I)$ do not depend on θ and the Stratonovitch correction terms vanish. We next estimate each summand on the right hand side of equation above. Note that, for $k = 0, 1, 2, \ldots, d$,

$$|\omega_k(I_r^\varepsilon) - \omega_k(I_r)| \leqslant \sup_{T^d\times D}|d\omega_k|\cdot|I_r^\varepsilon - I_r|. \tag{4.21}$$

The first term Λ_1 can be dealt with by Lipschitz estimate. Indeed, by Young's inequality and (4.19), clearly we have

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_1\Big] & \leqslant c_6(p,d\omega_0)\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon} |\theta_s^\varepsilon - \theta_s|^{p-1}|I_s^\varepsilon - I_s|ds\Big] \\ & \leqslant c_6\int_0^{t\wedge\tau^\varepsilon} \mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^p|\Big]ds + c_{13}\varepsilon^p t^{p+1}.\end{aligned} \tag{4.22}$$

For the stochastic Itô terms, we use the different kinds of maximal inequalities and the embedding $L^2\subset L^1$. Itô isometry yields

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_2\Big] & \leqslant c_7(p,d,d\omega_k)\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon} |\theta_s^\varepsilon - \theta_s|^{2p}|I_s^\varepsilon - I_s|^2ds\Big]^{\frac{1}{2}} \\ & \leqslant c_7\Big(\int_0^t \mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^p|\Big]ds\Big)^{\frac{1}{2}} + c_{15}\varepsilon^p t^{p+1}.\end{aligned} \tag{4.23}$$

Kunita's first inequality ([22, Page 265]) with the exponent 2 yields

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_3\Big] \leqslant & c_8(p,d,d\omega_k)\Bigg(\mathbb{E}\Bigg[\int_0^{t\wedge\tau^\varepsilon}\int_{\mathbb{R}^d\setminus\{0\}} |\theta_s^\varepsilon - \theta_s|^{2p} \\ & \times |I_s^\varepsilon - I_s|^2|z|^2\nu(dz)ds\Bigg]^{\frac{1}{2}} \\ & + \int_{|z|>1}|z|\nu(dz)\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon}|\theta_s^\varepsilon - \theta_s|^{p-1}|I_s^\varepsilon - I_s|ds\Big]\Bigg) \\ \leqslant & c_9\Bigg(\Big(\int_0^t \mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^p|\Big]ds\Big)^{\frac{1}{2}} \\ & + \int_0^t \mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^p|\Big]ds + \varepsilon^p t^{p+1}\Bigg)\end{aligned} \tag{4.24}$$

For canonical Marcus terms, we adapt the methods developed in [29, Section 3]. In fact, the term Λ_4 can be estimated in terms of the quadratic variation of L_t as shown in [29]. We rewrite the result in terms of the compensated Poisson random measure $\tilde{N}$ and then obtain

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_4\Big] \leqslant & c_{10}\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\sum_{k=1}^d\sum_{0<s\leqslant t}|\theta_s^\varepsilon - \theta_s|^{p-1}|I_s^\varepsilon - I_s||\triangle L^k(s)|^2\Big] \\ \leqslant & c_{11}\Big(\int_{\mathbb{R}^d\setminus\{0\}}|z|^4\nu(dz)\int_0^{t\wedge\tau^\varepsilon}\mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^{2(p-1)} \\ & \times |I_r^\varepsilon - I_r|^2\Big]ds\Big)^{\frac{1}{2}} \\ & + \int_0^{t\wedge\tau^\varepsilon}\mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon - \theta_r|^{p-1}|\Big]ds\end{aligned}$$

$$\leqslant c_{12}\left(\left(\int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds\right)^{\frac{1}{2}} + \int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds+\varepsilon^p t^{p+1}\right) \quad (4.25)$$

Observe that the terms Λ_5–Λ_8 are structurally identical to Σ_1–Σ_4 and they can be estimated analogously by replacing $K_I+\frac{d}{2}\hat{D}_I F_I$, $\tilde{K}_I$ and $\tilde{G}_I$ by K_θ, $\tilde{K}_\theta$ and $\tilde{G}_\theta$, respectively. Hence

$$\mathbb{E}\big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_5\big]\leqslant c_{13}\varepsilon\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^{p-1}|ds\Big] \leqslant c_{14}\varepsilon\int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds+c_{14}\varepsilon^p t, \quad (4.26)$$

$$\mathbb{E}\big[\sup_{s\leqslant t\wedge\tau^\varepsilon}\Lambda_6\big]\leqslant c_{15}\varepsilon\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^{2(p-1)}ds\Big]^{\frac{1}{2}} \leqslant c_{16}\varepsilon\left(\int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds\right)^{\frac{1}{2}}+c_{16}\varepsilon^p t, \quad (4.27)$$

and

$$\mathbb{E}\big[\sup_{s\leqslant t\wedge\tau^\varepsilon}(\Lambda_7+\Lambda_8)\big]\leqslant c_{17}\varepsilon\left(\mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^{2(p-1)}ds\Big]^{\frac{1}{2}} + \mathbb{E}\Big[\int_0^{t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^{p-1}|ds\Big]\right) \leqslant c_{18}\left(\left(\int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^p|\big]ds\right)^{\frac{1}{2}} + \int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds+\varepsilon^p t\right). \quad (4.28)$$

Taking the supremum and expectation in inequality (4.20) and combining the estimates (4.22)–(4.28), we obtain

$$\mathbb{E}\big[\sup_{s\leqslant t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^p\big]\leqslant c_{19}\left(\left(\int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds\right)^{\frac{1}{2}} + \int_0^t \mathbb{E}\big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|\theta_r^\varepsilon-\theta_r|^p|\big]ds + t\varepsilon^p(1+t^p)\right). \quad (4.29)$$

That is, for $u(t):=\mathbb{E}\big[\sup_{s\wedge\tau^\varepsilon\leqslant t}|\theta_s^\varepsilon-\theta_s|^p\big]$, $p(t)=t\varepsilon^p(1+t^p)$ and the concave invertible function $f(x)=\sqrt{x}+x$ we have

$$u(t)\leqslant c_{19}f\left(\int_0^t u(s)ds\right)+c_{19}p(t). \quad (4.30)$$

By the nonlinear extension of the Gronwall–Bihari inequality (see [44]), or a nonlinear comparison principle in [29], we finally have

$$\mathbb{E}\big[\sup_{s\leqslant t\wedge\tau^\varepsilon}|\theta_s^\varepsilon-\theta_s|^p\big]\leqslant c_{20}\varepsilon^{2p}\exp(c_{31}t). \quad (4.31)$$

Eventually, the desired result follows from Minkowski's inequality and the estimates (4.19) and (4.31),

$$\left[\mathbb{E}(\sup_{s\leqslant t\wedge\tau^\varepsilon}|f(Y_s^\varepsilon)-f(X_s)|^p)\right]^{\frac{1}{p}}\leqslant C_1\varepsilon\exp(C_2t). \quad \square \quad (4.32)$$

This lemma shows that, over a time interval t, the first integrals of the perturbed system change by an order $\varepsilon\exp(C_2t)$, and the slow component thus accumulate over a time interval of the size t/ε. Next, we would like to show that the randomness in the fast component could be averaged out by the induced invariant measure, and we can obtain a new dynamical system as ε goes to zero.

For convenience, we adopt the following notation. Let $g: M\to\mathbb{R}$ be a continuous function, and $\tilde{g}:\mathbb{T}^d\times D\to\mathbb{R}$ be its representation in action–angle coordinate. We define the average of g over the torus as $Q^g: D\subset\mathbb{R}^d\to\mathbb{R}$, i.e.,

$$Q^g(h)=\int_{\mathbb{T}^d}\tilde{g}(h,z)\mu(dz) \quad (4.33)$$

We remark that this average can be also understood in the sense of μ_h by taking the canonical transformation map $\pi': M_h\to\mathbb{T}^d$. Indeed, the induced measure $\pi'(\mu_h)$ is the Lebesgue measure μ on the torus and the average can be written as $Q^g(h)=\int_{M_h}g(h,z)\mu_h(dz)$ formally.

Lemma 4.5 (*Estimation of the Averaging Error*)**.** *Suppose that g is continuous on U_0. Set $H_i^\varepsilon(s)=H_i(Y_{t/\varepsilon}^\varepsilon)$ and $H^\varepsilon(s)=(H_1^\varepsilon(s),\ldots,H_d^\varepsilon(s))$. For $\tau^\varepsilon=\inf\{t\geqslant 0: Y_{t/\varepsilon}^\varepsilon\notin U_0\}$, we denote by*

$$\delta^g(\varepsilon,t)=\int_{s\wedge\tau^\varepsilon}^{(s+t)\wedge\tau^\varepsilon}g(Y_{r/\varepsilon}^\varepsilon)dr-\int_{s\wedge(\tau^\varepsilon/\varepsilon)}^{(s+t)\wedge(\tau^\varepsilon/\varepsilon)}Q^g(H^\varepsilon(r))dr \quad (4.34)$$

the averaging error. Then, for any given $t>0$ and sufficiently small $\varepsilon>0$, there are constants $k_1,k_2>0$ such that

$$\left(\mathbb{E}[\sup_{s\leqslant t}|\delta^g(\varepsilon,s)|^p]\right)^{\frac{1}{p}}\leqslant k_1t\left(\varepsilon^{1-k_2t}+\eta(t|\ln\varepsilon|)\right). \quad (4.35)$$

where $\eta(t)$ is the rate of convergence for ergodicity assumption **A6**.

Proof. The main idea is to use the approximate result in Lemma 4.4 on sufficiently small intervals and to apply the ergodicity assumption to replace time average by space average. We refer to Li [19] for a nice proof in the Brownian case and Högele–Ruffino [27], Gargate–Ruffino [28] and Högele–da Costa [29] for the extensions of this proof method. For sufficiently small $\varepsilon>0$ and $t\geqslant 0$ we define the partition

$$t_0=0<t_1<\cdots<t_{N_\varepsilon}\leqslant\frac{t}{\varepsilon}\wedge\tau^\varepsilon$$

with the following assignment of increments:

$$\triangle_\epsilon t=t|\ln\varepsilon|.$$

The grid points of the partition are given by $t_n^\varepsilon=n\triangle_\varepsilon t$ for $0\leqslant n\leqslant N_\varepsilon$ with $N_\varepsilon=\lfloor(\varepsilon|\ln\varepsilon|)^{-1}\rfloor$ where the bracket function $\lfloor\cdot\rfloor$ denotes the integer part of the value.

Initially we represent the left hand side as the sum:

$$\int_0^{t\wedge\varepsilon\tau^\varepsilon}g(Y_{t/\varepsilon}^\varepsilon)dr=\varepsilon\int_0^{\frac{t}{\varepsilon}\wedge\tau^\varepsilon}g(Y_r^\varepsilon)dr =\varepsilon\sum_{n=0}^{N_\varepsilon-1}\int_{t_n}^{t_{n+1}}g(Y_r^\varepsilon)dr+\varepsilon\int_{t_n}^{\frac{t}{\varepsilon}\wedge\tau^\varepsilon}g(Y_r^\varepsilon)dr. \quad (4.36)$$

Suppose that $\Psi:=\Psi_t=(\Psi(t,\omega,x),t\in\mathbb{R}^+)$ the solution flow of the unperturbed stochastic differential equation (4.1) with initial point x and Θ_t the shift operator on the canonical probability space, i.e., $\Theta_t(\omega)(-)=\omega(-+t)-\omega(t)$. Then,

$$|\delta^g(\varepsilon,t)|\leqslant\varepsilon\Big|\sum_{n=0}^{N_\varepsilon-1}\int_{t_n}^{t_{n+1}}g(Y_r^\varepsilon)dr-g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))dr\Big| +\varepsilon\Big|\sum_{n=0}^{N_\varepsilon-1}\int_{t_n}^{t_{n+1}}g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon)) -\triangle_\varepsilon tQ^g(H^\varepsilon(\varepsilon t_n))dr\Big|$$

$$
\begin{aligned}
&+\varepsilon|\sum_{n=0}^{N_\varepsilon-1}\triangle_\varepsilon t Q^g(H^\varepsilon(\varepsilon t_n))-\int_0^{t\wedge\varepsilon\tau^\varepsilon}Q^g(H^\varepsilon(r))dr|\\
&+\varepsilon|\int_{t_n}^{\frac{t}{\varepsilon}\wedge\tau^\varepsilon}g(Y_r^\varepsilon)dr|\\
=&\Pi_1+\Pi_2+\Pi_3+\Pi_4.
\end{aligned}
\tag{4.37}
$$

We proceed showing that the preceding four terms tend to zero uniformly on compact intervals. In the proof below c stands for an unspecified constant. Using the Markov property, Lemma 4.4 and Hölder's inequality,

$$
\begin{aligned}
&\left(\mathbb{E}\sup_{s\leqslant t}|\Sigma_1|^p\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon\sum_{n=0}^{N_\varepsilon-1}\left(\mathbb{E}\Big[\int_{t_n}^{t_{n+1}}\sup_{t_n\leqslant s\leqslant r}|g(Y_r^\varepsilon)dr-g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))|dr\Big]^p\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon\sum_{n=0}^{N_\varepsilon-1}(\triangle_\varepsilon t)^{\frac{p-1}{p}}\left(\mathbb{E}\Big[\int_{t_n}^{t_{n+1}}\sup_{t_n\leqslant s\leqslant r}|g(Y_r^\varepsilon)dr\right.\\
&\quad\left.-g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))|^pdr\Big]\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon N_\varepsilon(\triangle_\varepsilon t)C_1\varepsilon e^{C_2\triangle_\varepsilon t}=\varepsilon\lfloor(\varepsilon|\ln\varepsilon|)^{-1}\rfloor\cdot t|\ln\varepsilon|\cdot C_1\varepsilon e^{-C_2t\ln\varepsilon}\\
&\leqslant ct\varepsilon^{1-C_2t}
\end{aligned}
\tag{4.38}
$$

We denote $\mu_{H^\varepsilon(\varepsilon t_n)}$ by the invariant measure on the invariant manifold $M_{H^\varepsilon(\varepsilon t_n)}\equiv M_{Y^\varepsilon_{t_n}}$. Ergodicity assumption **A6** and the Markov property of the flow yield,

$$
\begin{aligned}
&\left(\mathbb{E}\sup_{s\leqslant t}|\Sigma_2|^p\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon\sum_{n=0}^{N_\varepsilon-1}\left(\mathbb{E}\sup_{s\leqslant t}|\int_{t_n}^{t_{n+1}}g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))dr\right.\\
&\quad\left.-\triangle_\varepsilon tQ^g(H^\varepsilon(\varepsilon t_n))|^p\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon\triangle_\varepsilon t\sum_{n=0}^{N_\varepsilon-1}\left(\mathbb{E}\sup_{s\leqslant t}|\frac{1}{\triangle_\varepsilon t}\int_{t_n}^{t_n+\triangle_\varepsilon t}g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))dr\right.\\
&\quad\left.-Q^g(H^\varepsilon(\varepsilon t_n))|^p\right)^{\frac{1}{p}}\\
&\leqslant\varepsilon N_\varepsilon\triangle_\varepsilon t\sup_n\Big(\mathbb{E}\sup_{s\leqslant t}|\frac{1}{\triangle_\varepsilon t}\int_{t_n}^{t_n+\triangle_\varepsilon t}g(\Psi_{r-t_n}(\Theta_{t_n}(\omega),Y_r^\varepsilon))dr\\
&\quad-\int_{M_{H^\varepsilon(\varepsilon t_n)}}g(H^\varepsilon(\varepsilon t_n),z)d\mu_{H^\varepsilon(\varepsilon t_n)}(z)|^p\Big)^{\frac{1}{p}}\\
&\leqslant c\varepsilon N_\varepsilon\triangle_\varepsilon t\eta(\triangle_\varepsilon t)=\varepsilon\lfloor(\varepsilon|\ln\varepsilon|)^{-1}\rfloor\cdot t|\ln\varepsilon|\cdot\eta(t|\ln\varepsilon|)\\
&\leqslant ct\eta(t|\ln\varepsilon|).
\end{aligned}
\tag{4.39}
$$

Note that g is C^1 on U_0, both $\sup_{U_0}|g|$ and $\sup_{U_0}|dg|$ are finite. We have the following estimates:

$$
\begin{aligned}
\left(\mathbb{E}\sup_{s\leqslant t}|\Sigma_3|^p\right)^{\frac{1}{p}}&\leqslant\varepsilon\sum_{n=0}^{N_\varepsilon-1}\triangle_\varepsilon t(\mathbb{E}\sup_{\varepsilon t_n\leqslant r\leqslant\varepsilon t_n+1}|Q^g(H^\varepsilon(\varepsilon t_n))\\
&\quad-Q^g(H^\varepsilon(\varepsilon r))|^p)^{\frac{1}{p}}\\
&\leqslant c\varepsilon N_\varepsilon\triangle_\varepsilon t\cdot c\varepsilon\exp(c\triangle_\varepsilon t)\\
&\leqslant ct\varepsilon^{1-ct},
\end{aligned}
\tag{4.40}
$$

and

$$
\left(\mathbb{E}\sup_{s\leqslant t}|\Sigma_4|^p\right)^{\frac{1}{p}}\leqslant c\varepsilon t|\ln\varepsilon|.
\tag{4.41}
$$

Consequently, the desired result follows from inequality (4.37), estimates (4.38)–(4.41), and Minkowski's inequality. □

At last, we present the proof of Theorem 4.1 based on the results of Lemmas 4.4 and 4.5.

Proof of Theorem 4.1. Applying the change of variable formula [33] for Marcus SDE (4.5) and using the completely integrability assumption **A3**, we have for $t<\tau_0\wedge\tau^\varepsilon$, $1\leqslant i\leqslant d$,

$$
\begin{aligned}
H_i^\varepsilon(t)=&H_i(Y_0)+\int_0^t\omega^2(V_i,K)(Y^\varepsilon_{s/\varepsilon})ds\\
&+\int_0^t\sum_{k=1}^d\omega^2(V_i,F_k\circ\pi)(Y^\varepsilon_{s/\varepsilon})\circ d\tilde{B}_t^k\\
&+\int_0^t\sum_{k=1}^d\omega^2(V_i,G_k\circ\pi)(Y^\varepsilon_{s/\varepsilon})\diamond d\tilde{L}_t^k.
\end{aligned}
\tag{4.42}
$$

For i fixed, we write

$$
g_i=\omega^2(V_i,K)
\tag{4.43}
$$

which is C^1 on U_0. Applying (4.34) to the functions g_i, we obtain for any $t<\tau^\varepsilon$,

$$
\int_0^{t\wedge\tau^\varepsilon}g_i(Y^\varepsilon_{s/\varepsilon})ds=\int_0^{t\wedge(\tau^\varepsilon/\varepsilon)}Q^{g_i}(H^\varepsilon(s))ds+\delta^{g_i}(\varepsilon,t).
\tag{4.44}
$$

On the other hand, using the notations of the previous two lemmas, Eq. (4.8) can be written as

$$
\begin{aligned}
d\bar{H}_i(t)=&Q^{g_i}(\bar{H}_i(t))dt+\sum_{k=1}^dF_{I,k}(\bar{H}_i(t))\circ d\tilde{B}_t^k\\
&+\sum_{k=1}^dG_{I,k}(\bar{H}_i(t))\diamond d\tilde{L}_t^k,\\
\bar{H}_0(t)=&H(Y_0).
\end{aligned}
$$

Therefore, for any $t<\tau^\varepsilon$, we have

$$
\begin{aligned}
|H_i^\varepsilon(t\wedge\tau^\varepsilon)-\bar{H}_i(t\wedge\tau^\varepsilon)|\leqslant&\int_0^{t\wedge\tau^\varepsilon}|Q^{g_i}(H_i^\varepsilon(s))-Q^{g_i}(\bar{H}_i(s))|ds\\
&+\delta(g_i,\varepsilon,t)\\
&+\int_0^{t\wedge\tau^\varepsilon}|F_{I,k}(H_i^\varepsilon(s))-F_{I,k}(\bar{H}_i(s))|\circ d\tilde{B}_t^k\\
&+\int_0^{t\wedge\tau^\varepsilon}|G_{I,k}(H_i^\varepsilon(s))-G_{I,k}(\bar{H}_i(s))|\diamond d\tilde{L}_t^k
\end{aligned}
\tag{4.45}
$$

Note that the estimate of the first term is straight forward Lipschitz estimate,

$$
\begin{aligned}
&\int_0^{t\wedge\tau^\varepsilon}|Q^{g_i}(H_i^\varepsilon(s))-Q^{g_i}(\bar{H}_i(s))|ds\\
&\leqslant C(g,\varphi)\int_0^{t\wedge\tau^\varepsilon}|(H_i^\varepsilon(s))-\bar{H}_i(s)|ds.
\end{aligned}
\tag{4.46}
$$

The estimates of the Brownian term and the Lévy term can be dealt with by Kunita's second inequality [22, Page 268] or other maximal inequalities. The computation for these two terms are very similar to that in the proof of Lemma 4.4, and we refer to [29, Section 5] for a detailed procedure. Finally,

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon}|H_i^\varepsilon(s)-\bar{H}_i(s)|^p\Big]\leqslant&C_1\int_0^t\mathbb{E}\Big[\sup_{r\leqslant s\wedge\tau^\varepsilon}|H_i^\varepsilon(r)-\bar{H}_i(r)|^p\Big]ds\\
&+\mathbb{E}\Big[\sup_{s\leqslant t}|\delta(g_i,\varepsilon,t)|^p\Big]
\end{aligned}
\tag{4.47}
$$

By Lemma 4.5 and Gronwall's inequality, there is a constant $k_3 > 0$ such that

$$\left(\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon} |H_i^\varepsilon(s) - \bar{H}_i(s)|^p\Big]\right)^{\frac{1}{p}} \leqslant \mathbb{E}\Big[\sup_{s\leqslant t} |\delta(g_i, \varepsilon, t)|^p\Big]^{\frac{1}{p}} \exp(k_3 t) \leqslant k_1 t\left(\varepsilon^{1-k_2 t} + \eta(t|\ln\varepsilon|)\right)\exp(k_3 t). \tag{4.48}$$

For the second part of the theorem, we have the following estimate by the definition of τ^ε, τ_δ and Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}(\tau^\varepsilon < \tau_\delta) &\leqslant \mathbb{P}\big(\sup_{s\leqslant\tau^\varepsilon\wedge\tau_\delta} |\bar{H}(s) - H^\varepsilon(s)| > \delta\big) \\ &\leqslant \delta^{-p}\mathbb{E}\Big[\sup_{s\leqslant\tau^\varepsilon\wedge\tau_\delta} |\bar{H}(s) - H^\varepsilon(s)|^p\Big] \\ &\leqslant k_4\delta^{-p}t^p\left(\varepsilon^{1-k_2 t} + \eta(t|\ln\varepsilon|)\right)^p \exp(k_5 t). \quad \square \end{aligned} \tag{4.49}$$

4.4. An example: Perturbed stochastic harmonic oscillator with Lévy noise

In this subsection, let us present a simple illustrative example for the above averaging principle of integrable stochastic Hamiltonian system with Lévy noise. We write $(q, p) = (q_1, \dots, q_d, p_1, \dots, p_d)$ as canonical coordinates, and there is an important class of Hamiltonian functions on $\mathbb{R}^{2n}$ of the form $H(q, p) = \frac{1}{2}|p|^2 + V(q)$, i.e. Hamiltonian H is the sum of kinetic, $T = \frac{1}{2}|p|^2 = \frac{1}{2}\sum_{i=1}^d p_i^2$ and potential, $V(q)$, energies. Furthermore, if V is quadratic, e.g. $V(q) = \frac{1}{2}\varpi|q|^2$ with ϖ a frequency, then we have the linear harmonic oscillator. Given Hamiltonian functions as follow,

$$H_1 = \frac{1}{2}\sum_{i=1}^d p_i^2 + \frac{1}{2}\sum_{i=1}^d \varpi_i^2 p_i^2,$$

$$H_k = \frac{1}{2}\frac{p_k^2}{\varpi_k} + \frac{1}{2}\varpi_k p_k^2, \quad k = 2, \dots, d,$$

and a smooth function H_0 commuting with all H_k, $k = 1, \dots, d$, i.e.

$$\{H_0, H_k\} = \sum_{i=1}^d \left(\frac{\partial H_0}{\partial p_i}\frac{\partial H_k}{\partial q_i} - \frac{\partial H_0}{\partial q_i}\frac{\partial H_k}{\partial p_i}\right) = 0,$$

we have

$$dq_i(t) = \frac{\partial H_0}{\partial p_i}dt + \sum_{i=1}^d \frac{\partial H_k}{\partial p_i} \circ dB_t^k + \sum_{i=1}^d \frac{\partial H_k}{\partial p_i} \diamond dL_t^k, \tag{4.50}$$

$$dp_i(t) = -\frac{\partial H_0}{\partial q_i}dt - \sum_{i=1}^d \frac{\partial H_k}{\partial q_i} \circ dB_t^k - \sum_{i=1}^d \frac{\partial H_k}{\partial q_i} \diamond dL_t^k, \tag{4.51}$$

which is an integrable stochastic Hamiltonian system with α-stable Lévy noise. Let $J = \left[\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right]$ be a $2d \times 2d$ antisymmetric matrix, which is called Poisson matrix, this system is equivalent to

$$dX_t = J\nabla H_0(X_t)dt + \sum_{i=1}^d J\nabla H_k(X_t) \circ dB_t^k + \sum_{i=1}^d J\nabla H_k(X_t) \diamond dL_t^k. \tag{4.52}$$

For $M_h = \{x \in M : H_k(x) = h_k,\ k = 1, 2, ..d\}$, if we take an action–angle coordinates change $\varphi^{-1} : U_0 \to \mathbb{T}^d \times D, (q, p) \mapsto (\theta, I)$,

$$q_i = \sqrt{\frac{2I_i}{\varpi_i}}\cos\theta_i, \quad p_i = \sqrt{2\varpi_i I_i}\sin\theta_i, \tag{4.53}$$

then the induced Hamiltonians $H'_k = H_k(\varphi(\theta, I)) = \begin{cases} \sum_{i=1}^d \varpi_i I_i, & k = 1 \\ I_k, & k = 2, \dots, d \end{cases}$ on $\mathbb{T}^d \times D$ satisfy,

$$\dot{\theta}_k^i = \frac{\partial H'_k}{\partial I_i} =: \omega_k^i(I) = \begin{cases} \varpi_i, & k = 1; \\ 1, & k = 2, \dots, d \ \&\ i = k; \\ 0, & otherwise. \end{cases}$$

$$\dot{I}_k^i = -\frac{\partial H'_k}{\partial\theta_i} = 0.$$

Next, we investigate the effective behavior of a small transversal perturbation of order ε to this system. For simplicity, we consider the case on $\mathbb{R}^4$ with $\varpi = 1$ and L_t having second moments,

$$d\begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} p_1 & 0 \\ p_2 & p_2 \\ -q_1 & 0 \\ -q_2 & -q_2 \end{pmatrix} \circ d\begin{pmatrix} B_t^1 \\ B_t^2 \end{pmatrix} + \begin{pmatrix} p_1 & 0 \\ p_2 & p_2 \\ -q_1 & 0 \\ -q_2 & -q_2 \end{pmatrix} \diamond d\begin{pmatrix} L_t^1 \\ L_t^2 \end{pmatrix}. \tag{4.54}$$

Take the perturbation vectors to be $\varepsilon K = \left(0, \varepsilon q_2/(q_2^2 + p_2^2), 0, 0\right)^T$, $\varepsilon(E, O)^T\dot{\tilde{B}}_t$ and $\varepsilon(E, O)^T\dot{\tilde{L}}_t$, where E is the identity matrix, O is the zero matrix and $\tilde{L}_t$ is a pure jump Lévy motion with four-order moments. By action–angle coordinates change (4.53), we have, with $\Lambda = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$,

$$d\begin{pmatrix} \theta_1 \\ \theta_2 \\ I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} \Lambda \\ O \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \end{pmatrix} + \begin{pmatrix} \Lambda \\ O \end{pmatrix} d\begin{pmatrix} L_t^1 \\ L_t^2 \end{pmatrix} + \varepsilon\begin{pmatrix} 0 \\ \dfrac{1}{2I_2}\sin\theta_2\cos\theta_2 \\ 0 \\ -\cos^2\theta_2 \end{pmatrix} \times dt + \varepsilon\begin{pmatrix} E \\ O \end{pmatrix} d\begin{pmatrix} \tilde{B}_t^1 \\ \tilde{B}_t^2 \end{pmatrix} + \varepsilon\begin{pmatrix} E \\ O \end{pmatrix} d\begin{pmatrix} \tilde{L}_t^1 \\ \tilde{L}_t^2 \end{pmatrix}.$$

For unperturbed system, it is easy to get fundamental solution with initial condition $(q_0, p_0) = \varphi(\theta_0, I_0)$: $q_t = \sqrt{2I_0}\cos(\Lambda(B_t + L_t))$, $p_t = \sqrt{2I_0}\sin(\Lambda(B_t + L_t))$ with $\Lambda = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Note that

$$g_i := \omega^2(V_i, K) = V_i^T JK = \frac{q_2^2}{q_2^2 + p_2^2} \implies \tilde{g}_i = \cos^2\theta_2, \ i = 1, 2.$$

We obtain

$$Q^g(h_i) = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\cos^2\theta_2 d\theta_1 d\theta_2 = \frac{1}{2}.$$

We verify that $\frac{1}{t}\int_0^t g_i(q_s, p_s)ds \to Q^g(h_i)$ in L^2, as $t \to \infty$, with a rate of convergence $\eta(t) = \frac{1}{\sqrt{t}}$ in the Appendix. Therefore, the transversal system stated in Theorem 4.1 is $\bar{H}_i(t) = \frac{t}{2}$. The result guarantees that, on the accelerated time scale $\frac{t}{\varepsilon}$, $H_i^\varepsilon(t)$ has a local behavior close to $\frac{t}{2}$ in the sense that

$$\left(\mathbb{E}\Big[\sup_{s\leqslant t\wedge\tau^\varepsilon} \Big|H_i^\varepsilon(s) - \frac{t}{2}\Big|^2\Big]\right)^{\frac{1}{2}} \leqslant k_1 t\left(\varepsilon^{1-k_2 t} + \frac{t}{2}|\ln\varepsilon|\right)\exp(k_3 t) \tag{4.55}$$

tends to 0 when $\varepsilon \to 0$, for any fixed t and the constant $k_1, k_2, k_3 > 0$.

Acknowledgments

The authors are grateful to Qingshan Chen, Guan Huang, Haitao Xu, Guowei Yu, Jianlu Zhang, Lei Zhang and Yanjie Zhang for helpful discussions and comments. This work was partly supported by the NSFC grants 11531006 and 11771449.

Appendix. Proofs of Theorems 3.1–3.2 and calculations of example 4.4

We now prove Theorems 3.1 and 3.2 which are based on the formula of change of variables in differential forms.

Proof of Theorem 3.1. Noticing that

$$\begin{aligned} dP\wedge dQ &= \sum_{i=1}^{n} dP^i\wedge dQ^i \\ &= \sum_{i=1}^{n}\sum_{l=r+1}^{n}\sum_{r=1}^{n}\Big[\big(\frac{\partial P^i}{\partial p^r}\frac{\partial Q^i}{\partial p^l}-\frac{\partial P^i}{\partial p^l}\frac{\partial Q^i}{\partial p^r}\big)dp^r\wedge dp^l \\ &\quad +\big(\frac{\partial P^i}{\partial q^r}\frac{\partial Q^i}{\partial q^l}-\frac{\partial P^i}{\partial q^l}\frac{\partial Q^i}{\partial q^r}\big)dq^r\wedge dq^l\Big] \\ &\quad +\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\big(\frac{\partial P^i}{\partial p^r}\frac{\partial Q^i}{\partial q^l}-\frac{\partial P^i}{\partial q^l}\frac{\partial Q^i}{\partial p^r}\big)dp^r\wedge dq^l, \end{aligned}$$

we infer that the phase flow of (3.1)–(3.2) preserves symplectic structure if and only if

$$\begin{cases} \displaystyle\sum_{i=1}^{n}\frac{D(P^i,Q^i)}{D(p^r,p^l)}=0, \quad r\neq l, \\ \displaystyle\sum_{i=1}^{n}\frac{D(P^i,Q^i)}{D(q^r,q^l)}=0, \quad r\neq l, \\ \displaystyle\sum_{i=1}^{n}\frac{D(P^i,Q^i)}{D(p^r,q^l)}=\delta_{rl}, \quad r,l=1,\dots,n. \end{cases} \tag{A.56}$$

Clearly,

$$\frac{D(P^i(t_0),Q^i(t_0))}{D(p^r,p^l)}=\frac{D(p^i,q^i)}{D(p^r,p^l)}=0,$$
$$\frac{D(P^i(t_0),Q^i(t_0))}{D(q^r,q^l)}=\frac{D(p^i,q^i)}{D(q^r,q^l)}=0,$$
$$\frac{D(P^i(t_0),Q^i(t_0))}{D(p^r,q^l)}=\frac{D(p^i,q^i)}{D(p^r,q^l)}=\delta_{rl}.$$

Therefore, (A.56) is fulfilled if and only if

$$\begin{aligned}\sum_{i=1}^{n} d\frac{D(P^i(t),Q^i(t))}{D(p^r,p^l)} &= \sum_{i=1}^{n} d\frac{D(P^i(t),Q^i(t))}{D(q^r,q^l)} \\ &= \sum_{i=1}^{n} d\frac{D(P^i(t),Q^i(t))}{D(p^r,q^l)}=0. \end{aligned}\tag{A.57}$$

Introduce the notation

$$P_p^{ir}=\frac{\partial P^i}{\partial p^r},\quad P_q^{ir}=\frac{\partial P^i}{\partial q^r},\quad Q_p^{ir}=\frac{\partial Q^i}{\partial p^r},\quad Q_q^{ir}=\frac{\partial P^i}{\partial q^r}.$$

For a fixed r, by calculating at (P,Q) with $P=P(t)=(P^1(t;t_0,p,q),\dots,P^n(t;t_0,p,q))$ and $Q=Q(t)=(Q^1(t;t_0,p,q),\dots,Q^n(t;t_0,p,q))$ which is a solution to systems (3.1)–(3.2), we obtain $P_p^{ir}, Q_p^{ir}, i=1,\dots,n$, satisfy the following system of SDEs:

$$\begin{aligned} dP_p^{ir} &= \sum_{j=1}^{n}\big(\frac{\partial f^i}{\partial p^j}P_p^{jr}+\frac{\partial f^i}{\partial q^j}Q_p^{jr}\big)dt \\ &\quad +\sum_{k=1}^{d}\sum_{j=1}^{n}\big(\frac{\partial \sigma_k^i}{\partial p^j}P_p^{jr}+\frac{\partial \sigma_k^i}{\partial q^j}Q_p^{jr}\big)\diamond dL^k,\quad P_p^{ir}(t_0)=\delta_{ir}, \\ dQ_p^{ir} &= \sum_{j=1}^{n}\big(\frac{\partial g^i}{\partial p^j}P_p^{jr}+\frac{\partial g^i}{\partial q^j}Q_p^{jr}\big)dt \\ &\quad +\sum_{k=1}^{d}\sum_{j=1}^{n}\big(\frac{\partial \gamma_k^i}{\partial p^j}P_p^{jr}+\frac{\partial \gamma_k^i}{\partial q^j}Q_p^{jr}\big)\diamond dL^k,\quad Q_p^{ir}(t_0)=0, \end{aligned}\tag{A.58}$$

where

$$f(Q,P)=\frac{\partial H}{\partial P}(Q,P),\quad \sigma_k(Q,P)=\frac{\partial H_k}{\partial P}(Q,P),\tag{A.59}$$

$$g(Q,P)=-\frac{\partial H}{\partial Q}(Q,P),\quad \gamma_k(Q,P)=-\frac{\partial H_k}{\partial Q}(Q,P),\tag{A.60}$$

for $k=1,\dots,m$.

Then, we get

$$\begin{aligned} &dP_p^{ir}(t)Q_p^{il}(t) \\ &= \sum_{j=1}^{n}\Big[\big(\frac{\partial f^i}{\partial p^j}P_p^{jr}+\frac{\partial f^i}{\partial q^j}Q_p^{jr}\big)Q_p^{il}+\big(\frac{\partial g^i}{\partial p^j}P_p^{jr}+\frac{\partial g^i}{\partial q^j}Q_p^{jr}\big)P_p^{ir}\Big]dt \\ &\quad +\sum_{k=1}^{m}\sum_{j=1}^{n}\Big[\big(\frac{\partial \sigma_k^i}{\partial p^j}P_p^{jr}+\frac{\partial \sigma_k^i}{\partial q^j}Q_p^{jr}\big)Q_p^{il} \\ &\quad +\big(\frac{\partial \gamma_k^i}{\partial p^j}P_p^{jl}+\frac{\partial \gamma_k^i}{\partial q^j}Q_p^{jl}\big)P_p^{ir}\Big]\diamond dL^k. \end{aligned}$$

Similarly, we can also calculate $dP_p^{il}(t)Q_p^{ir}(t)$, then

$$\sum_{i=1}^{n} d\frac{D(P^i(t),Q^i(t))}{D(p^r,p^l)}=\sum_{i=1}^{n}\Big[\sum_{j=1}^{n}\Xi_1 dt+\sum_{k=1}^{m}\sum_{j=1}^{n}\Xi_2\diamond dL^k\Big],\tag{A.61}$$

where

$$\begin{aligned} \Xi_1 &= \frac{\partial f^i}{\partial p^j}P_p^{jr}Q_p^{il}+\frac{\partial f^i}{\partial q^j}Q_p^{jr}Q_p^{il}+\frac{\partial g^i}{\partial p^j}P_p^{jl}P_p^{ir}+\frac{\partial g^i}{\partial q^j}Q_p^{jl}P_p^{ir} \\ &\quad -\frac{\partial f^i}{\partial p^j}P_p^{jl}Q_p^{ir}-\frac{\partial f^i}{\partial q^j}Q_p^{jl}Q_p^{ir}-\frac{\partial g^i}{\partial p^j}P_p^{jr}P_p^{il}-\frac{\partial g^i}{\partial q^j}Q_p^{jr}P_p^{il}, \\ \Xi_2 &= \frac{\partial \sigma_k^i}{\partial p^j}P_p^{jr}Q_p^{il}+\frac{\partial \sigma_k^i}{\partial q^j}Q_p^{jr}Q_p^{il}+\frac{\partial \gamma_k^i}{\partial p^j}P_p^{jl}P_p^{ir}+\frac{\partial \gamma_k^i}{\partial q^j}Q_p^{jl}P_p^{ir} \\ &\quad -\frac{\partial \sigma_k^i}{\partial p^j}P_p^{jl}Q_p^{ir}-\frac{\partial \sigma_k^i}{\partial q^j}Q_p^{jl}Q_p^{ir}-\frac{\partial \gamma_k^i}{\partial p^j}P_p^{jr}P_p^{il}-\frac{\partial \gamma_k^i}{\partial q^j}Q_p^{jr}P_p^{il}. \end{aligned}$$

It is not difficult to find out that, a sufficient condition of $\Xi_1=0$ is

$$\frac{\partial f^i}{\partial p^j}=-\frac{\partial g^j}{\partial q^i},\quad \frac{\partial f^i}{\partial q^j}=\frac{\partial f^j}{\partial q^i},\quad \frac{\partial g^i}{\partial p^j}=\frac{\partial g^j}{\partial p^i},\tag{A.62}$$

and a sufficient condition of $\Xi_2=0$ is

$$\frac{\partial \sigma_k^i}{\partial p^j}=-\frac{\partial \gamma_k^j}{\partial q^i},\quad \frac{\partial \sigma_k^i}{\partial q^j}=\frac{\partial \sigma_k^j}{\partial q^i},\quad \frac{\partial \gamma_k^i}{\partial p^j}=\frac{\partial \gamma_k^j}{\partial p^i}.\tag{A.63}$$

Noticing that relations (A.59)–(A.60) imply (A.62)–(A.63), we obtain $\sum_{i=1}^{n} d\frac{D(P^i(t),Q^i(t))}{D(p^r,p^l)}=0$. Similarly, we prove that the conditions (A.59)–(A.60) ensure the other two terms of (A.57) as well. This completes the proof. □

Proof of Theorem 3.3. We calculate the derivatives of $\mathcal{S}$ with respect to Q_0 and Q_1:

$$\begin{aligned} \frac{\partial \mathcal{S}}{\partial Q_0} &= \int_{t_0}^{t_1}\Big(\frac{\partial L}{\partial Q}\frac{\partial Q}{\partial Q_0}+\frac{\partial L}{\partial \dot{Q}}\frac{\partial \dot{Q}}{\partial Q_0}+\frac{\partial L}{\partial P}\frac{\partial P}{\partial Q_0}+\frac{\partial L}{\partial \dot{P}}\frac{\partial \dot{P}}{\partial Q_0}\Big)dt \\ &\quad -\sum_{k=1}^{d}\int_{t_0}^{t_1}\Big(\frac{\partial H_k}{\partial Q}\frac{\partial q}{\partial Q_0}+\frac{\partial H_k}{\partial P}\frac{\partial P}{\partial Q_0}\Big)\diamond dL^k(t) \\ &= \Big[\frac{\partial L}{\partial \dot{Q}}\frac{\partial Q}{\partial Q_0}\Big]_{t_0}^{t_1}+\int_{t_0}^{t_1}\Big(\frac{\partial L}{\partial Q}-\frac{d}{dt}\frac{\partial L}{\partial \dot{Q}}-\sum_{k=1}^{d}\frac{\partial H_k}{\partial Q}\diamond \dot{L}^k(t)\Big) \\ &\quad \times\frac{\partial Q}{\partial Q_0}dt \\ &\quad +\Big[\frac{\partial L}{\partial \dot{P}}\frac{\partial P}{\partial Q_0}\Big]_{t_0}^{t_1}+\int_{t_0}^{t_1}\Big(\frac{\partial L}{\partial P}-\frac{d}{dt}\frac{\partial L}{\partial \dot{P}}-\sum_{k=1}^{d}\frac{\partial H_k}{\partial P}\diamond \dot{L}^k(t)\Big) \end{aligned}$$

$$\times \frac{\partial p}{\partial Q_0}dt$$
$$= -P_0^T, \tag{A.64}$$

where the last equality follows from the stochastic Lagrange equations (3.23)–(3.24) and the Legendre transform $P = \frac{\partial L}{\partial \dot{Q}}$.

Similarly, we have

$$\frac{\partial \mathcal{S}}{\partial Q_1} = -P_1^T. \tag{A.65}$$

Therefore,

$$d\mathcal{S} = -P_0^T dQ_0 + P_1^T dQ_1. \tag{A.66}$$

Moreover,

$$dp_1 \wedge dQ_1 = d(\frac{\partial \mathcal{S}}{\partial Q_1}) \wedge dQ_1 = \frac{\partial^2 \mathcal{S}}{\partial Q_1 \partial Q_0} dQ_0 \wedge dQ_1, \tag{A.67}$$

$$dp_0 \wedge dQ_0 = d(-\frac{\partial \mathcal{S}}{\partial Q_0}) \wedge dQ_0 = \frac{\partial^2 \mathcal{S}}{\partial Q_0 \partial Q_1} dQ_0 \wedge dQ_1. \tag{A.68}$$

Smoothness of L and the $H_k(k = 1, \ldots, d)$ in $\mathcal{S}$ ensures that $\frac{\partial^2 \mathcal{S}}{\partial Q_1 \partial Q_0} = \frac{\partial^2 \mathcal{S}}{\partial Q_0 \partial Q_1}$, which implies

$$dP_1 \wedge dQ_1 = dP_0 \wedge dQ_0. \tag{A.69}$$

The proof is thus complete. □

Detailed Calculations of Example 4.4. Recall that, by Lévy–Khintchine formula [8,22] the characteristic function for Lévy motion in $\mathbb{R}^d$ is

$$\mathbb{E}e^{i\langle u, L_t\rangle} = e^{t\eta_0(u)}, \ u \in \mathbb{R}^d,$$

where $\eta_0(u) = \int_{\mathbb{R}^d \setminus \{0\}} \left[e^{iu\cdot z} - 1 - i\mathbf{1}_{\{|z|<1\}} u \cdot z\right]\nu(dz)$ whose real part $\Re\eta_0 \leqslant 0$. And the characteristic function for standard Brownian motion in $\mathbb{R}^d$ is

$$\mathbb{E}e^{i\langle u, B_t\rangle} = e^{-\frac{1}{2}t\langle u, Iu\rangle} = e^{-\frac{1}{2}t|u|^2}, \ u \in \mathbb{R}^d.$$

Therefore,

$$\mathbb{E}\Big[\frac{1}{t}\int_0^t g_i(q_s,p_s)ds\Big] = \mathbb{E}\Big[\frac{1}{t}\int_0^t \tilde{g}_i(\theta_s, I_s)ds\Big]$$
$$= \mathbb{E}\Big[\frac{1}{t}\int_0^t \cos^2(\langle u, B_s + L_s\rangle)\Big]ds$$
$$= \frac{1}{2t}\int_0^t \mathbb{E}\cos 2(\langle u, B_s + L_s\rangle)ds + \frac{1}{2}$$
$$= \frac{1}{2t}\int_0^t \Re\mathbb{E}e^{i\langle 2u, B_s\rangle}\Re\mathbb{E}e^{i\langle 2u, L_s\rangle}ds + \frac{1}{2}$$
$$= \frac{1}{2t}\int_0^t e^{-s(\frac{1}{2}|2u|^2 - \Re\eta_0(2u))}ds + \frac{1}{2} = \frac{1}{2t}\frac{1}{A}(1 - e^{-At}) + \frac{1}{2}. \tag{A.70}$$

Here $u = (1,1)^T \in \mathbb{R}^2$ and $A = \frac{1}{2}|2u|^2 - \Re\eta_0(2u) > 0$. Hence, as t goes to ∞, the expectation is equal to $\frac{1}{2}$ eventually. Next, we calculate the secondary moment as follows,

$$\mathbb{E}\Big[\Big|\frac{1}{t}\int_0^t g_i(q_s,p_s)ds\Big|^2\Big] = \mathbb{E}\Big[\frac{1}{t^2}\Big|\int_0^t \cos^2(\langle u, B_s + L_s\rangle)ds\Big|^2\Big]$$
$$= \frac{2}{t^2}\int_0^t\int_0^r \mathbb{E}\Big[\cos^2(\langle u, B_s + L_s\rangle)\cos^2(\langle u, B_r + L_r\rangle)\Big]dsdr$$
$$= \frac{1}{4t^2}\int_0^t\int_0^r \mathbb{E}\Big[\Re e^{i\langle 2u,(B_s+L_s)+(B_r+L_r)\rangle} + \Re e^{i\langle 2u,(B_s+L_s)-(B_r+L_r)\rangle}$$
$$+ 2\Re e^{i\langle 2u,(B_s+L_s)\rangle} + 2\Re e^{i\langle 2u,(B_r+L_r)\rangle} + 2\Big]dsdr$$
$$= \frac{1}{4t^2}\int_0^t\int_0^r \Big[\mathbb{E}e^{i\langle 4u, B_s\rangle}\mathbb{E}e^{i\langle 2u, B_r - B_s\rangle}\mathbb{E}e^{i\langle 4u, L_s\rangle}\mathbb{E}e^{i\langle 2u, L_r - L_s\rangle}$$
$$+ \mathbb{E}e^{i\langle 2u, B_r - B_s\rangle}\mathbb{E}e^{i\langle 2u, L_r - L_s\rangle}$$
$$+ 2\mathbb{E}e^{i\langle 2u, B_s\rangle}\mathbb{E}e^{i\langle 2u, L_s\rangle} + 2\mathbb{E}e^{i\langle 2u, B_r\rangle}\mathbb{E}e^{i\langle 2u, L_r\rangle} + 2\Big]dsdr$$
$$= \frac{1}{4t^2}\int_0^t\int_0^r \Big[e^{-(\frac{1}{2}|2u|^2 - \Re\eta_0(2u))r - (\frac{1}{2}|4u|^2 - \frac{1}{2}|2u|^2 - \Re\eta_0(4u) + \Re\eta_0(2u))s}$$
$$+ e^{-(\frac{1}{2}|2u|^2 + \Re\eta_0(2u))(r-s)}$$
$$+ 2e^{-(\frac{1}{2}|2u|^2 - \Re\eta_0(2u))s} + 2e^{-(\frac{1}{2}|2u|^2 - \Re\eta_0(2u))r} + 2\Big]dsdr$$
$$= \frac{1}{4t^2}\int_0^t\int_0^r \Big[e^{-Ar-Bs} + e^{-C(r-s)} + 2e^{-As} + 2e^{-Ar} + 2\Big]dsdr$$
$$= \frac{1}{4t^2}\Big[\frac{1}{A(A+B)} - \frac{1}{C^2} + (\frac{1}{C} + \frac{2}{A})t + t^2 + \frac{e^{-(A+B)t}}{B(A+B)} - \frac{e^{-At}}{AB}$$
$$+ \frac{e^{-Ct}}{C^2} - \frac{2te^{-At}}{A}\Big]. \tag{A.71}$$

Here we used the stationary independent increments property of the Brownian motion and Lévy motion. By Taylor expansion [23, Page 40] with $u = (1,1)^T$, we can find that $B = \frac{1}{2}|4u|^2 - \frac{1}{2}|2u|^2 - \Re\eta_0(4u) + \Re\eta_0(2u) > 0$, $C = \frac{1}{2}|2u|^2 + \Re\eta_0(2u) > 0$, so we have $\mathbb{E}\Big[\Big|\frac{1}{t}\int_0^t g_i(q_s,p_s)ds\Big|^2\Big] \to \frac{1}{4}$ as $t \to \infty$. Thus,

$$\mathbb{E}\Big[\Big|\frac{1}{t}\int_0^t g_i(q_s,p_s)ds - Q^g(h_i)\Big|^2\Big]$$
$$= \mathbb{E}\Big[\frac{1}{t^2}\Big|\int_0^t \cos^2(\langle u, B_s\rangle + \langle u, L_s\rangle)ds - \frac{1}{2}\Big|^2\Big]$$
$$= \mathbb{E}\Big[\frac{1}{t^2}\Big(\int_0^t \cos^2(\langle u, B_s\rangle + \langle u, L_s\rangle)ds\Big)^2\Big]$$
$$- \mathbb{E}\Big[\frac{1}{t}\int_0^t \cos^2(\langle u, B_s\rangle + \langle u, L_s\rangle)ds\Big] + \frac{1}{4}$$
$$\to 0, \ as\ t \to \infty.$$

Moreover, combining (A.70)–(A.71) and taking the square root, the rate of convergence is of the order $\eta(t) = \frac{c}{\sqrt{t}}$ as $t \to \infty$ (c is a constant). □

References

[1] V.I. Arnol'd, Mathematical Methods of Classical Mechanics, Vol. 60, Springer Science & Business Media, 2013.

[2] J.D. Meiss, Differential Dynamical Systems, Vol. 14, Siam, 2007.

[3] S. Wiggins, Normally Hyperbolic Invariant Manifolds in Dynamical Systems, Vol. 105, Springer Science & Business Media, 2013.

[4] F. Rezakhanlou, Hamiltonian ODE, Homogenization, and Symplectic Topology, UC Berkeley. URL https://math.berkeley.edu/~rezakhan/WKAM.pdf.

[5] H. Poincaré, Les Méthodes Nouvelles de la Mécanique Céleste: Méthodes de MM. Newcomb, Glydén, Lindstedt et Bohlin, Vol. 2, Gauthier-Villars it fils, 1893.

[6] F.T. Tveter, Deriving the Hamilton equations of motion for a nonconservative system using a variational principle, J. Math. Phys. 39 (3) (1998) 1495–1500.

[7] M.I. Freidlin, A.D. Wentzell, Random Perturbations of Dynamical Systems, Vol. 260, Springer Science & Business Media, 2012.

[8] J. Duan, An Introduction to Stochastic Dynamics, Cambridge University Press, 2015.

[9] J.-M. Bismut, Mécanique Aléatoire, in: Lecture Notes in Math, Springer, Berlin, 1981.

[10] M. Brin, M. Freidlin, On stochastic behavior of perturbed hamiltonian systems, Ergodic Theory Dynam. Systems 20 (1) (2000) 55–76.

[11] R. MacKay, Langevin equation for slow degrees of freedom of Hamiltonian systems, in: Nonlinear Dynamics and Chaos: Advances and Perspectives, Springer, 2010, pp. 89–102.

[12] T. Misawa, Conserved quantities and symmetries related to stochastic dynamical systems, Ann. Inst. Statist. Math. 51 (4) (1999) 779–802.

[13] L. Wu, Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems, Stochastic Process. Appl. 91 (2) (2001) 205–238.

[14] W. Zhu, Z. Huang, Stochastic stabilization of quasi-partially integrable Hamiltonian systems by using Lyapunov exponent, Nonlinear Dynam. 33 (2) (2003) 209–224.
[15] G.N. Milstein, Y.M. Repin, M.V. Tretyakov, Numerical methods for stochastic systems preserving symplectic structure, SIAM J. Numer. Anal. 40 (4) (2002) 1583–1604.
[16] G.N. Milstein, Y.M. Repin, M.V. Tretyakov, Symplectic integration of Hamiltonian systems with additive noise, SIAM J. Numer. Anal. 39 (6) (2002) 2066–2088.
[17] L. Wang, J. Hong, R. Scherer, F. Bai, Dynamics and variational integrators of stochastic Hamiltonian systems, Int. J. Numer. Anal. Model. 6 (4) (2009) 586–602.
[18] M. Pavon, Hamilton's principle in stochastic mechanics, J. Math. Phys. 36 (12) (1995) 6774–6800.
[19] X.-M. Li, An averaging principle for a completely integrable stochastic Hamiltonian system, Nonlinearity 21 (4) (2008) 803–822.
[20] J. Cresson, S. Darses, Stochastic embedding of dynamical systems, J. Math. Phys. 48 (7) (2007) 365–373.
[21] D. Givon, R. Kupferman, A. Stuart, Extracting macroscopic dynamics: Model problems and algorithms, Nonlinearity 17 (6) (2004) R55–127.
[22] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge University Press, 2009.
[23] K.-I. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.
[24] L. Garnett, Foliations, the ergodic theorem and brownian motion, J. Funct. Anal. 51 (3) (1983) 285–311.
[25] S. Albeverio, Z. Brzeźniak, J.-L. Wu, Existence of global solutions and invariant measures for stochastic differential equations driven by poisson type noise with non-Lipschitz coefficients, J. Math. Anal. Appl. 371 (1) (2010) 309–322.
[26] S. Albeverio, B. Rüdiger, J.-L. Wu, Invariant measures and symmetry property of Lévy type operators, Potential Anal. 13 (2) (2000) 147–168.
[27] M. Högele, P. Ruffino, Averaging along foliated Lévy diffusions, Nonlinear Anal. TMA 112 (2015) 1–14.
[28] I.I.G. Gargate, P.R. Ruffino, An averaging principle for diffusions in foliated spaces, Ann. Probab. 44 (1) (2016) 567–588.
[29] M.A. Högele, P.H. da Costa, Strong averaging along foliated Lévy diffusions with heavy tails on compact leaves, Potential Anal. 47 (3) (2017) 277–311.
[30] Y. Xu, J. Duan, W. Xu, An averaging principle for stochastic dynamical systems with Lévy noise, Physica D 240 (17) (2011) 1395–1401.
[31] J. Bao, G. Yin, C. Yuan, Two-time-scale stochastic partial differential equations driven by α-stable noises: Averaging principles, Bernoulli 23 (1) (2017) 645–669.
[32] H. Kunita, Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms, in: Real and Stochastic Analysis, Springer, 2004, pp. 305–373.
[33] T.G. Kurtz, É. Pardoux, P. Protter, Stratonovich stochastic differential equations driven by general semimartingales, in: Annales de L'IHP Probabilités et Statistiques, Vol. 31, 1995, pp. 351–377.
[34] S.I. Marcus, Modeling and approximation of stochastic differential equations driven by semimartingales, Stochastics 4 (3) (1981) 223–245.
[35] K.D. Elworthy, Stochastic Differential Equations on Manifolds, Vol. 70, Cambridge University Press, 1982.
[36] P.E. Protter, Stochastic Integration and Differential Equations, Springer, 2005.
[37] R. Abraham, J.E. Marsden, Foundations of Mechanics, Benjamin/Cummings, 1978.
[38] F. Xi, C. Zhu, Jump type stochastic differential equations with non-Lipschitz coefficients: Non-confluence, Feller and strong Feller properties, and exponential ergodicity, J. Differential Equations 266 (8) (2018) 4668–4711.
[39] A.M. Fox, R. de la Llave, Barriers to transport and mixing in volume-preserving maps with nonzero flux, Physica D 295 (2015) 1–10.
[40] V.I. Arnold, V.V. Kozlov, A.I. Neishtadt, Mathematical Aspects of Classical and Celestial Mechanics, Vol. 3, Springer Science & Business Media, 2007.
[41] U. Krengel, On the speed of convergence in the ergodic theorem, Monatsh. Math. 86 (1) (1978) 3–6.
[42] S. Kakutani, K. Petersen, The speed of convergence in the ergodic theorem, Monatsh. Math. 91 (1) (1981) 11–18.
[43] A.M. Kulik, Exponential ergodicity of the solutions to SDE's with a jump noise, Stochastic Process. Appl. 119 (2) (2009) 602–632.
[44] B.G. Pachpatte, Inequalities for Differential and Integral Equations, Academic Press, 1998.