Hamiltonian Systems with Lévy Noise: Symplecticity, Hamilton’s Principle and Averaging Principle

Pingyuan Wei\textsuperscript{a}, Ying Chao\textsuperscript{a}, Jinqiao Duan\textsuperscript{b,∗}

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

Abstract

The 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 behaviour 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 deterministic system of differential equations 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.

Keywords: Stochastic Hamiltonian systems; Lévy noise; symplecticity; Hamilton’s principle; averaging principle.

2010 MSC: 60H10, 60J51, 58J37

1. Introduction

Certain nonlinear systems have “geometric” structures, such as the Hamiltonian structure \[\mathbf{1}2\mathbf{3}\]. Hamiltonian systems of ordinary differential equations (ODEs) widely

\[\text{This work was partly supported by the NSF grant 1620449, and NSFC grants 11531006 and 11771449.}
\[\text{Corresponding author.}
\[Email addresses: weipingyuan@hust.edu.cn (Pingyuan Wei), yingchao1993@hust.edu.cn (Ying Chao), duan@iit.edu (Jinqiao Duan)

Preprint submitted to January 21, 2019
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, MacKay \([10]\), Misawa \([11]\), Wu \([12]\), Zhu-Huang \([13]\). In particular, Milstein et al. \([14, 15]\) proved the symplecticity for stochastic Hamiltonian systems with Brownian noise, and Wang et al. \([16]\) proposed a version of Hamilton’s principle for the same systems to construct variational integrators; Pavon \([17]\) established variational principles in stochastic mechanics; Li \([18]\) 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 \([19]\) and Givon et al. \([20]\).

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 \([8, 21, 22]\). 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.
The study on stochastic systems driven by such processes have received increasing attentions recently, especially on developing proper averaging principles for these systems. For example, Albeverio et al. [23, 24] established ergodicity of Lévy-type operators and SDEs driven by jump noise with non-Lipschitz coefficients; Högele-Ruffino [25] focused on averaging along foliated Lévy diffusions which generalized the approach by Li [18], and Högele-da Costa [26] 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. [27] and Bao et al. [28]. 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 [29, 30] 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 [31].

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 $\varepsilon$ ($\varepsilon$ 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 $t/\varepsilon$. 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 an 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 $\varepsilon$ 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 [18], as well as Högele-Ruffino [25]. 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. 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. A Hamiltonian vector field

Recall that a symplectic structure on a smooth $2n$-dimensional manifold $M$ is a closed nondegenerate differential two-form $\omega^2$. Here using the symbol with superscript 2 to avoid confusion with the chance variable of sample space. The pair $(M, \omega^2)$ is called a symplectic manifold, or we usually say that “$M$ is a symplectic manifold”. Darboux’s theorem asserts that symplectic manifold $M$ is locally $\mathbb{R}^{2n}$ with the standard symplectic form $\omega^2_0 = \sum dp_i \wedge dq_i$. Such coordinates $(q, p) = (q_1, ..., q_n, p_1, ..., p_n)$ are called symplectic, canonical, or Darboux coordinates.

Note that $\omega^2$ associates to any vector field $v$ on $M$ the differential one-form: $f(\cdot) = \omega^2(\cdot, v)$. Since $\omega^2$ is non-degenerate, and the dimensions of the vector spaces $TM_x$ and $T^*M_x (x \in M)$ coincide, there exists an inverse operator $J$ such that $f(\cdot) = \omega^2(\cdot, Jf)$.

For a smooth function $H: M \to \mathbb{R}$, it determines the one-form $dH$. The vector field $v_H = JdH$ on $M$ is called the Hamiltonian vector field with Hamiltonian $H$, that is, $dH(\cdot) = \omega^2(\cdot, v_H)$.

The space of smooth functions on $M$ has a Lie algebra structure given by the Poisson bracket. The Poisson bracket of two smooth functions $F_1, F_2$ is defined by $\{F_1, F_2\} = dF_2(v_{F_1}) = \omega^2(v_{F_1}, v_{F_2})$. Moreover, if $\{F, H\} = 0$ we say that $F$ is a first integral of $H$.

2.2. A Lévy motion

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space, where $\mathcal{F}_t$ is a nondecreasing family of sub-$\sigma$-fields of $\mathcal{F}$ satisfying the usual conditions. An $\mathcal{F}_t$ adapted process $L_t = L(t)$ taking values in $\mathbb{R}^d$ with $L(0) = 0$ a.s. is called a Lévy motion if it is stochastically continuous and has independent and stationary increments.

A Lévy process $L_t$ can be characterized by a drift vector $\gamma \in \mathbb{R}^d$, an $d \times d$ non-negative-definite symmetric covariance matrix $A$ and a Borel measure $\nu$ defined on $\mathbb{R}^d \setminus \{0\}$. We call $(\gamma, A, \nu)$ the generating triplet of the Lévy motions $L_t$. Moreover, we have the Lévy-Itô decomposition for $L_t$ as follows:

$$L_t = \gamma t + B_A(t) + \int_{|z|<1} z\tilde{N}(t, dz) + \int_{|z|\geq 1} zN(t, dz), \quad (2.1)$$
where $N(dt, dz)$ is the Poisson random measure, $\tilde{N}(dt, dz) = N(dt, dz) - \nu(dz)dt$ is the compensated Poisson random measure, $\nu(S) = \mathbb{E}N(1, S)$ is the jump measure, and $B_A(t)$ is an independent standard $d$-dimensional Brownian motion with covariance matrix $A$. 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.

### 2.3. A Stochastic Hamiltonian system with Lévy noise

We shall consider a stochastic Hamiltonian system driven by non-Gaussian Lévy noise, which is described by the following SDE in the Marcus form on a smooth $2n$-dimensional manifold $M$

$$dX = V_0(X)dt + \sum_{k=1}^d V_k(X) \odot dL^k(t), \quad X(t_0) = x \in M,$$

where $V_0$ and $V_k$ are $2n$-dimensional Hamiltonian vector fields with Hamiltonians $H_0$ and $H_k (k = 1, 2, \ldots, d)$, respectively, and $L(t) = (L^1(t), L^2(t), \ldots, L^d(t))$ is a $d$-dimensional Lévy process with respect to the given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ with characteristic triplet $(\gamma, A, \nu)$.

The equation (2.2) can be interpreted as

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

where “$\odot$” stands for Marcus integral $^{[29, 30]}$ defined by

$$\int_0^t V_k(X_{s-}) \odot dL^k(s) = \int_0^t V_k(X_{s-}) \odot dL^k(s) + \int_0^t V_k(X_{s-})dL^k(s) + \sum_{l=1}^d \sum_{0 \leq s \leq t} \left[ \phi(\Delta L^l(s), V_k(X_{s-}), X_{s-} - V_k(X_{s-})\Delta L^k(s)) \right]$$

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

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

Note that, by Lévy-Itô decomposition $^{[21]}$, the system (2.2) with Lévy generating triplet being $(0, I, 0)$ is a stochastic Hamiltonian system with Brownian noise $^{[14, 15, 16]}$, and the system (2.2) without Lévy term is a deterministic Hamiltonian
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 (2.2) and (2.3) in the Itô form [21, 29]. This can be carried out by employing the Lévy-Itô decomposition. It’s convenient to write the $d$-dimensional Brownian term in the form:

$$B_A(t) = \sigma B(t),$$

where $B(t)$ is a $d'$-dimensional standard Brownian motion and $\sigma$ 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 $1 \leq i \leq 2n, t \geq 0$,

$$dX_i^t = V_0^t(X_t)dt + \sum_{k=1}^d \gamma^k V_k^t(X_t)dt + \sum_{k=1}^d V_k^t(X_t)dB_k(t) + \frac{1}{2} \sum_{k=1}^d V_k^t(X_t)\nu dt + \int_{|z|<1} [\phi(z)(X_{t-}) - X_i^t]N(dt, dz) + \int_{|z|>1} [\phi(z)(X_{t-}) - X_i^t] N(dt, dz)$$

$$+ \frac{1}{2} \sum_{k=1}^d V_k^t(X_t)\nu dt.$$

(2.6)

We define $\hat{DV}(x)$ to be the vector in $M$ whose $i$-th component is $\max_{1 \leq k \leq d} |V_k \cdot \nabla V_i(x)|$ for $1 \leq i \leq 2n, x \in M$. If we assume that $V_0, V_k (1 \leq k \leq d)$ and $\hat{DV}(x)$ are locally Lipschitz continuous functiona and satisfy the “one sided linear growth” condition in the following sense:

A1. (Local Lipschitz condition) For any $x \in M$, there exists a neighborhood $M_0$ of $x$ such that $V_0|_{M_0}$, $V_k|_{M_0}$ and $\hat{DV}(x)|_{M_0}$ are Lipschitz continuous, i.e. there is a constant $N_1 > 0$ such that, for all $x_1, x_2 \in M_0$,

$$|V_0(x_1) - V_0(x_2)|^2 + \max_{1 \leq k \leq d} |V_k(x_1) - V_k(x_2)|^2 + |\hat{DV}(x_1) - \hat{DV}(x_2)|^2 \leq N_1|x_1 - x_2|^2,$$

A2. (One sided linear growth condition) There exists a constant $N_2 > 0$ such that, for all $x \in M$,

$$\sum_{k=1}^d V_k^2(x) + 2x \cdot V_0(x) \leq N_2(1 + |x|^2),$$

then there exists a unique global solution to (2.6), and the solution process is adapted and càdlàg, referring to [23, Theorem 3.1] and [21, Lemma 6.10.3]; see also [32].
Generator: Let $\Psi_t := (\Psi(t, \omega, x), t \geq 0)$ be the solution flow of the SDE (2.6) with starting point $x$ and $(T_t, t \geq 0)$ be the semigroup associated with $\Psi_t$. Let $C_b(M)$ be the linear space of all continuous bounded Borel measurable functions from $M$ to $\mathbb{R}^d$. If each $V_k \in C^1_b(M)$ ($1 \leq k \leq d$), then the diffusion process $X_t$ defined by (2.6) have generating operators $A : C^2_b(M) \rightarrow C_b(M)$ satisfying, for each $f \in C^2_b(M)$, $x \in M$,

$$(Af)(x) = (L_0 f)(x) + \sum_{k=1}^d \gamma^k (L_k f)(x) + \frac{1}{2} \sum_{k=1}^d (L_k L_k f)(x) + \int_{\mathbb{R}^d \setminus \{0\}} [f(\phi(z)x) - f(x) - \sum_{k=1}^d z^k (L_k f)(x) \mathbf{1}_{|z|<1}(z)]v(dz).$$

(2.7)

where $L_0$ and $L_k$ indicates Lie differentiation in the direction of $V_0$ and $V_k$, respectively. More precisely, we have $L f = df(V_0) = \omega^2(v_f, V_0)$ and $L_k f = df(V_k) = \omega^2(v_f, V_k)$ here.

Applying Itô’s formula to the solution flow $\Psi$, we have, for any $f \in C^2_b(M)$, $x \in M$, $t \geq 0$,

$$f(\Psi_t(x)) = f(x) + \int_0^t (Af)(\Psi_s(x))ds + \sum_{k=1}^d \int_0^t L_k(\Psi_s(x))dB^k(s) + \int_0^t \int_{|z|<1} [f(\phi(z)(\Psi_s(x)) - f(\Psi_s(x))]N(ds, dz) + \int_0^t \int_{|z|\geq 1} [f(\phi(z)(\Psi_s(x)) - f(\Psi_s(x))]N(ds, dz).$$

(2.8)

If $V_k$ is such that $A : C^2_0(M) \rightarrow C_0(M)$, then $(T_t, t \geq 0)$ is a Feller semigroup. For details see [21].

3. Symplecticity and stochastic Hamilton’s principle

In this section we present several facts about stochastic Hamiltonian systems with Lévy noise, such as the property of preserving symplectic structure and stochastic Hamilton’s principle, which will help us to better understand such a system and 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, 9, 5]. When stochastic Hamilton systems are treated as a nonconservative systems (see next subsection), we can show that they also have this intrinsic property.

For simplicity, we consider system (2.2) in the canonical form. With $X = (Q, P), \ X_0 = (q, p), \ V = (\frac{\partial H}{\partial P}, -\frac{\partial H}{\partial Q})$ and $V_k = (\frac{\partial H_k}{\partial P}, -\frac{\partial H_k}{\partial Q}),$ we rewrite the system as follows

\[dQ = \frac{\partial H}{\partial P}(Q, P)dt + \sum_{k=1}^{d} \frac{\partial H_k}{\partial P}(Q, P) \diamond dL_k(t), \ Q(t_0) = q, \ 
\]

\[dP = -\frac{\partial H}{\partial Q}(Q, P)dt - \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q}(Q, P) \diamond dL_k(t), \ P(t_0) = p. \ 
\]

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) such that the transformation $(p, q) \rightarrow (P, Q)$ preserves symplectic structure as follows:

\[dP \wedge dQ = dp \wedge dq, \]

\[i.e., \ \sum_{i=1}^{n} dp_i \wedge dq_i = \sum_{i=1}^{n} dp_i \wedge dq_i. \ 
\]

Geometrically, (3.3) means that the sum of the oriented areas of projections is an integral invariant [1, 13]. Consequently, for such systems, all external powers of the two-form are also invariant, and the case of $n$-th external 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 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 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, 16]. 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 \( F \) be a nonconservative generalized force. The work done by this nonconservative generalized force is defined as

\[
W = - F \cdot r, \tag{3.4}
\]

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

\[
\delta W = F \cdot \delta r = F \cdot \frac{\partial r}{\partial q} \delta q.
\]

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 \( y = \{q(t) : t_0 \leq t \leq t_1\} \) as a temporally parameterized curve in the configuration space. Under the influence of \( F \), the action integral of this curve is defined by

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

Hamilton’s principle of nonconservative mechanical systems asserts that \( \delta S = 0 \) is equal to the following Euler-Lagrange equation holds:

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

Here the Lagrangian \( L \) is considered as a function with independent variables \( q, \dot{q} \) and \( t \).
It is known to \cite{6} that the Euler-Lagrange equations of motion have the property of redundancy. As the value of Lagrangian is invariant to variable transformations, Lagrangian $\mathcal{L}$ can be transformed from the variable set $\{q\}$ to a redundant variable set $\{Q, P\}$ by

$$L(q, \dot{q}, t) = L(q(Q, P, t), \dot{q}(Q, P, Q, P, t), t) = L(Q, P, \dot{Q}, \dot{P}, t).$$

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

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{P}} - \frac{\partial \mathcal{L}}{\partial P} = F \cdot \frac{\partial \mathbf{r}}{\partial P},$$

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{Q}} - \frac{\partial \mathcal{L}}{\partial Q} = F \cdot \frac{\partial \mathbf{r}}{\partial Q}.$$  \hspace{1cm} (3.8, 3.9)

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

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

$$\dot{P} = -\frac{\partial H}{\partial Q} + \frac{\partial \mathbf{r}}{\partial Q} \cdot F.$$  \hspace{1cm} (3.10, 3.11)

Lévy noise as a kind of random fluctuating force, can be treated as a special non-conservative force \cite{21,13}. We rewrite a stochastic Hamiltonian system with Lévy noise (3.1 - 3.2) in the following form

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

$$\dot{P} = -\frac{\partial H}{\partial Q} - \frac{\partial \bar{H}}{\partial Q} \circ \dot{L}(t).$$  \hspace{1cm} (3.12, 3.13)

where $\bar{H} = (H_1, H_2, ..., H_d)$. It is natural to compare (3.10 - 3.11) with (3.12 - 3.13). Formally, the associations between $F$ 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).
Remark 3.1. 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)

\begin{align}
\frac{dx}{dt} &= y(t), \quad x(t_0) = x_0, \quad (3.14) \\
\frac{dy}{dt} &= -x(t) - \sigma dL_t, \quad y(t_0) = y_0. \quad (3.15)
\end{align}

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 \( e^{\int J dt} \), where \( J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \). It’s not hard to show that this equation has the unique solution

\begin{align}
x(t) &= x(0) \cos t + y(0) \sin t + \int_0^t \sigma \sin(t - s)dL_s, \quad (3.16) \\
y(t) &= -x(0) \sin t + y(0) \cos t + \int_0^t \sigma \cos(t - s)dL_s. \quad (3.17)
\end{align}

Notice that this SDE involving the term of large jumps by Lévy-Itô decomposition. It is easy to handle using interlacing [21, Page 392], and it makes sense to begin by omitting this term and concentrate on the study of the equation driven by continuous noise interspersed with small jumps. To this end, we consider the corresponding modified SDE and take the initial conditions \( x_0 = 1, \ y_0 = 0 \) and the drift of Lévy motion \( \gamma = 0 \), then

\begin{align}
x(t) &= \cos t + \int_0^t \sigma \sin(t - s)dB_s + \int_{|z|<1} \sigma z \sin(t - s)\tilde{N}(ds, dy), \quad (3.18) \\
y(t) &= -\sin t + \int_0^t \sigma \cos(t - s)dB_s + \int_{|z|<1} \sigma z \cos(t - s)\tilde{N}(ds, dy). \quad (3.19)
\end{align}

By Itô isometry and the properties of compensated Poisson integral [21], we can find that the second moment of the solution satisfies

\begin{align}
\mathbb{E}(x^2(t) + y^2(t)) &= 1 + \sigma^2 t + \sigma^2 t \int_{|z|<c} |z|^2 \nu(dz), \quad (3.20)
\end{align}

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 3.2.** An alternative view of stochastic Hamilton system is that we can regard it as an open Hamilton system within the external world: the stochastic part in (2.2) characterizes the complicated interaction between the "deterministic" Hamiltonian system with the Hamiltonian \( H_0 \) and the chaotic environment [11].

For stochastic Hamiltonian system with Lévy noise (3.12 - 3.13), according to (3.21), the work done by Lévy noise is formally

\[
W_{\text{stoch}} = - \sum_{k=1}^{d} H_k \circ \dot{L}^k(t). 
\]

(3.21)

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

\[
S_{\text{stoch}}[\gamma] = \int_{t_0}^{t_1} (L - W_{\text{stoch}})dt = \int_{t_0}^{t_1} L(\gamma(t), \dot{\gamma}(t), t)dt - \sum_{k=1}^{d} \int_{t_0}^{t_1} H_k(\gamma(t), t) \circ dL^k(t),
\]

(3.22)

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

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

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{P}} - \frac{\partial L}{\partial P} = \sum_{k=1}^{d} \frac{\partial H_k}{\partial P} \circ \dot{L}^k(t), \quad (3.23)
\]

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{Q}} - \frac{\partial L}{\partial Q} = \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q} \circ \dot{L}^k(t). \quad (3.24)
\]

We call \( S_{\text{stoch}} \) 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 $S_{\text{stoch}}[\gamma]$ is stationary if it does not vary when the curve is slightly changed, $\gamma(t) \rightarrow \gamma(t) + \delta \gamma(t)$. The change in the action upon doing this can be formally expanded in $\delta \gamma$,

$$S_{\text{stoch}}[\gamma + \delta \gamma] - S_{\text{stoch}}[\gamma] = \int_{t_0}^{t_1} \frac{\delta S_{\text{stoch}}}{\delta \gamma} \delta \gamma(t) dt + o(\delta \gamma), \quad \text{(3.25)}$$

where $\frac{\delta S_{\text{stoch}}}{\delta \gamma}$ is called the Fréchet or functional derivative of $S_{\text{stoch}}$.

Noticing that Marcus integral enables the application of the differential chain rule of the Newton-Leibniz type, and integrating by parts, we calculate the derivative,

$$\delta S_{\text{stoch}} = \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial Q} \delta Q + \frac{\partial L}{\partial P} \delta P + \frac{\partial L}{\partial \dot{Q}} \delta \dot{Q} + \frac{\partial L}{\partial \dot{P}} \delta \dot{P} \right) dt$$

$$- \sum_{k=1}^{d} \int_{t_0}^{t_1} \left( \frac{\partial H_k}{\partial Q} \delta Q + \frac{\partial H_k}{\partial P} \delta P \right) \delta \dot{L}_k(t) dt$$

$$= \left[ \frac{\partial L}{\partial Q} \delta Q \right]_{t_0}^{t_1} + \left[ \frac{\partial L}{\partial P} \delta P \right]_{t_0}^{t_1} + \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial Q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{Q}} - \sum_{k=1}^{d} \frac{\partial H_k}{\partial Q} \delta \dot{L}_k(t) \right) \delta Q dt$$

$$+ \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial P} - \frac{d}{dt} \frac{\partial L}{\partial \dot{P}} - \sum_{k=1}^{d} \frac{\partial H_k}{\partial P} \delta \dot{L}_k(t) \delta P dt. \right.$$

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 [16], the desired result follows. □

Example 3.2. Consider the linear stochastic oscillators with Lévy noise (3.14 - 3.15). We show that the equations (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

$$\begin{cases} \frac{d}{dt} \frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} = -\sigma \dot{L}_1, \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} = 0. \end{cases} \quad \text{(3.26)}$$

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 $S$ 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.

**Theorem 3.3. (Characterization of stochastic action integral)** The stochastic action integral $S_{\text{stoch}}$ satisfies

$$dS_{\text{stoch}} = -p_0^T dq_0 + p_1^T dq_1.$$  

(3.27)

Furthermore, if the Lagrangian $L$ and the functions $H_k$ ($k = 1, \ldots, d$) are sufficiently smooth with respect to $p$ and $q$, then the mapping $(p_0, q_0) \mapsto (p_1, q_1)$ defined by equation (3.27) is symplectic.

The proof is given in the Appendix.

**Remark 3.3.** The Theorems 3.1, 3.2 and 3.3 show that it is reasonable to consider such SDEs driven by with Lévy noise with Hamiltonian structure. What’s more, the problem of preserving phase-volume or preserving symplectic structure are considered widely in numerical modeling and practical application; see Hairer et al. [33], Milstein et al. [15, 14], Fox-de la Llave [34]. And a systematic approach of producing symplectic numerical methods based on Hamilton’s principle of stochastic Hamiltonian systems with Brownian noise has been studied in Wang et al. [18]. These are also the reasons of studying the symplecticity and Hamilton’s principle of stochastic Hamiltonian systems with Lévy noise.

4. **An averaging principle for a completely integrable stochastic Hamiltonian system**

We now return to the stochastic Hamiltonian system with Lévy noise (2.2) on a $2n$-dimensional smooth manifold $M$ (for simplicity, set $n = d$ in the rest of this discussion). As mentioned above, this stochastic Hamiltonian system is itself a nonconservative system, and the perturbation of nonconservative force (Lévy noise) is interior for this stochastic system as the noise would add energy to the system and it make sense
to preserve symplectic structure and develop the stochastic Hamilton’s principle. Then 
a interesting question to raise is: if there is even a small external perturbation in this 
stochastic system, just as the deterministic case and the stochastic case with Brownian 
noise referring to the study of Freidlin-Wentzell [7], Li [18] and so on, what the effective 
dynamic behaviour would be? To answer this question, we consider a completely 
integrable stochastic Hamiltonian system with Lévy noise.

4.1. Completely integrable stochastic Hamiltonian system with Lévy noise

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

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\) be a filtered probability space satisfying usual conditions. 
Given an integrable family \(\{H_k\}_{k=1}^d\) and a Hamiltonian vector field \(V_0\) with Hamiltonian 
function \(H_0\) commuting with \(V_k\). We consider the following model:

\[
dX_t = V_0(X_t)dt + \sum_{k=1}^d V_k(X_t) \circ dB^k(t) + \sum_{k=1}^d V_k(X_t) \circ dL^k(t). 
\tag{4.1}
\]

Here \(B(t)\) 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.

We proceed under the following assumption:

A3 The exponential moment of \(L(t)\) exists with constant \(\kappa > N_1\), that is, the jump 
measure \(\nu\) satisfy 
\[
\int_{\mathbb{R}^d \setminus \{0\}} e^{\kappa |z|} \nu(dz) < \infty.
\]

A4 Completely integrability: \(\{H_i, H_j\} = \omega^2(V_i, V_j) = 0\) for \(i, j = 0, 1, 2, \ldots, d\).

Remark 4.1. Some comments on our model and our assumptions have to be made:
The noises we consider here are composed of standard Brownian motion \(B(t)\) and pure 
jump Lévy motion \(L(t)\) which are mutually independent, aiming to compare our results with the Brownian case [18]. Indeed, we can understand this special form in the 
sense of Lévy-Itô decomposition (2.7) and Marcus integral (2.4). The assumption A3 is
necessary since the increment of \( \phi(\Delta L, V, \cdot) \) in [24] depends exponentially on the random size of \( \Delta L \) [35, Lemma 3.1], and it will be used in the estimation of Lemma 4.3. Moreover, this existence condition on exponential moment can be replaced by one on polynomial moments of order bigger than 2, referring to [26]. The assumption A4 will play a core role in the rest of this section, especially to ensure the existence of invariant manifolds and invariant measure of this model. In addition, \( V_0 \) can be weakened by locally Hamiltonian vector which is not given by a Hamiltonian function as in [18].

Under these assumptions, the diffusion vector fields of this integrable stochastic Hamiltonian system span a sub-bundle of the tangent bundle, at least locally. The purpose of this section is to investigate the effect of a small perturbation to stochastic systems of this type. Next, we will show that a solution to this SDE preserves the energies \( H_k \) and there are corresponding invariant manifolds (level sets).

### 4.2. Invariant manifolds and invariant measure

Due to the system has \( d \) first integrals \( H_1, ..., H_d \) in involution. We consider the joint integral level

\[
M_h = \{ x \in M : H_i(x) = h_i = \text{const}, \ i = 1, 2, ..., d \}.
\]

The Liouville-Arnold theorem 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 \( (I, \theta) \) (\( I \) are the actions and \( \theta \) are the angles) such that \( \omega^2 = dI \wedge d\theta \) (symplecticity), \( H = H(I) \) (i.e., \( I \) are first integrals).

For such a \( h \), \( M_h \) is invariant under the flows of each \( H_i \). And for each \( x \) in \( M \), we have \( h = (H_1(x), ..., H_d(x)) \), thus it determines an invariant manifold, which we write also as \( M_{H(x)} \).

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

\[
dH_i(X_t) = dH_i(V_0(X_t))dt + \sum_{k=1}^d dH_i(V_k(X_t)) \circ dB_k(t) + \sum_{k=1}^d dH_i(V_k(X_t)) \circ dL_k(t) = 0.
\]
\textbf{Proof.} By applying Itô's formula, for $1 \leq i \leq d,$
\begin{align*}
dH_i(x_i) &= \omega^2(V_0, V_i)(x_i) + \frac{1}{2} \sum_{k=1}^{d} \omega^2(V_k, \omega^2(V_k, V_i))(x_i) + \sum_{k=1}^{d} \omega^2(V_k, V_i)(x_i) dB^k(t) \\
&\quad + \int_{\|z\|<1} [H_i(\phi(z))(x_{i-})] \tilde{N}(dt, dz) + \int_{\|z\|>1} [H_i(\phi(z))(x_{i-}) - H_i(x_{i-})] N(dt, dz) \\
&\quad + \int_{\mathbb{R}^d \setminus \{0\}} [H_i(\phi(z)x) - H_i(x_{i-}) - \sum_{k=1}^{d} z_k \omega^2(V_k, V_i)(x_{i-}) I_{\|z\|<1}(z)] \nu(dz) \\
&= \omega^2(V_0, V_i)(x_i) + \sum_{k=1}^{d} \omega^2(V_k, V_i)(x_i) \circ dB^k(t) + \sum_{k=1}^{d} \omega^2(V_k, V_i)(x_i) \circ dL^k(t).
\end{align*}
Thus the assertion is obtained from the assumption A3 of completely integrability. \(\square\)

Next, we deal with the ergodicity of SDE (4.1). To ensure there is a unique invariant probability measure on $M_h,$ throughout this paper, we further assume:

\textbf{A4} The invariant manifolds are compact, the map $H : x \in M \mapsto (H_1(x), ..., H_d(x)) \in \mathbb{R}^d$ is proper and its set of critical points has measure zero. Note that the $d$ vector fields $\{V_k\}_{k=1}^{d}$ are tangent to $M_{H(x)}$ and the symplectic form $\omega^2$ vanishes on the invariant manifolds $M_h.$ Therefore the SDE (4.1) is elliptic with a compensated integral term \([21]\) when restricted to each compact invariant manifold and the Markovian solution is ergodic. Let us use $\mu_h$ for the unique invariant probability measure on $M_h$ which can be considered as the uniform probability on the torus. Moreover, there exists a positive, bounded, decreasing function $\eta(t) : [0, \infty) \to [0, \infty)$ such that, for any function $f$ on a compact manifold converging to infinity as $t$ converges to infinity, $\frac{1}{t} \int_s^{t+s} f(X_r) dr$ converges to $\int_{M_h} f(z) \mu_h(dz)$ when $s \to \infty$ with rate of convergence bounded by $\eta(t)$ and the convergence is uniform on compact time intervals in $L^p$ for all $p > 1.$

\textbf{Remark 4.2.} Indeed, when the Markovian solution restricts to individual invariant manifold, we have the generator of restriction by \([2,7]\)

\begin{align*}
(\mathcal{R}f)(x) &= (\mathcal{L}_0 f)(x) + \frac{1}{2} \sum_{k=1}^{d} (\mathcal{L}_k \mathcal{L}_k f)(x) \\
&\quad + \int_{\mathbb{R}^d \setminus \{0\}} [f(\phi(z)x) - f(x) - \sum_{k=1}^{d} z_k^2 (\mathcal{L}_k f)(x) I_{\|z\|<1}(z)] \nu(dz) \quad (4.3)
\end{align*}
for every function \( f \in C^2_b(M) \).

**Remark 4.3.** Under our assumption, for almost every point \( h_0 \) in \( \mathbb{R}^d \), there is a neighbourhood \( G \) of \( h_0 \) such that \( H^{-1}(h) \) is a smooth sub-manifold for all \( h \in G \) and that there is a diffeomorphism from \( H^{-1}(G) \) to \( G \times H^{-1}(h_0) \). Such \( h_0 \) is called a regular value of \( H \). A point \( y \) in \( M \) is said to be a critical point if \( H(y) \) is not regular. By Morse-Sard theorem [36], the set of critical values of the function \( H \) has measure zero.

Recall that in a neighbourhood of a regular point \( h_0 \) of \( H \), every component of the level set \( M_{h_0} \) is diffeomorphic to a \( d \)-dimensional torus \( T^d \), and a small neighbourhood \( U_0 \) of \( M_{h_0} \) is diffeomorphic to the product space \( T^d \times D \), where \( D \) is a relatively compact open set in \( \mathbb{R}^d \). Take an action-angle chart around \( M_{h_0} \). The measure \( (\sum_i dl^i \wedge d\theta^i)^d \) on the product space naturally splits to give us a probability measure, the Haar measure \( \theta_1 \wedge ... \wedge \theta^d \) on \( T^d \). We take the corresponding one on \( M_{h_0} \) and denote it by \( \mu_{h_0} \).

With the help of action-angle transformation and assumption A3, we thus have the following lemma.

**Lemma 4.2.** Let \( E = \text{span}[V_1, ..., V_d] \) be a sub-bundle of the tangent bundle of rank \( d \). Let \( U \) be a section of \( E \) commuting with all \( V_i \) \( (1 \leq i \leq d) \). The invariant measure for stochastic Hamiltonian system (4.1) restricted to the invariant manifold \( M_{h_0} \) is \( \mu_{h_0} \), which varies smoothly with a sufficiently small neighbourhoods of a regular value.

**Proof.** Recall that \( M_{h_0} \) have the form in (4.2), we rewrite \( U = \sum_{i=1}^d h_i V_i(x) \). By the assumption A3 of completely integrability, for any smooth function \( f \) on \( M_{h_0} \), we have

\[
\int_{M_{h_0}} d\mu_f(x) = \int_{M_{h_0}} (H_i, f)(x)\mu_{h_0}(dx) = \int_{\mathbb{T}^d} d(f \circ \varphi)(-\sum_{k=1}^d \frac{\partial (H_k \circ \varphi)}{\partial I_k} \frac{\partial}{\partial \theta_k}) d\theta = -\sum_{k=1}^d \omega_k(I) \int_{\mathbb{T}^d} \frac{\partial (f \circ \varphi)}{\partial \theta_k} d\theta = 0,
\]

where \( \varphi \) is the action-angle coordinate map (see the next subsection for detail), \( (I, \theta) \) are the corresponding action-angle coordinates. Thus \( U \) is divergence free, i.e. \( \text{div}_E U = 0 \), in the sense of

\[
\int_{M_{h_0}} d\mu_f(U(x)\mu_{h_0}(dx) = -\int_{M_{h_0}} \text{div}_E U(x)\mu_{h_0}(dx) = 0. \tag{4.4}
\]
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 \( \mu_h \) is the desired object. \( \square \)

**Remark 4.4.** Note that the invariant measure of SDE (4.1) is strongly dependent on the assumption of completely integrability. More general research on ergodicity and unique invariant measure of SDEs with Lévy noise can be found in Albeverio et al. [24], Brzeźniak et al. [23] and references therein.

### 4.3. The perturbed system and statement of an averaging principle

We next study the situation where an integrable stochastic Hamiltonian system is perturbed by a vector field \( \varepsilon K \) for small parameter \( \varepsilon > 0 \), and \( K \) is a transversal smooth vector field which is the case if \( \omega^2(V_k, K) \), \( k = 0, 1, \ldots, d \) are not all identically zero.

Consider the perturbed system corresponding to (4.1):

\[
\begin{align*}
\mathrm{d}Y^\varepsilon_t &= V_0(Y^\varepsilon_t)\mathrm{d}t + \sum_{i=1}^d V_k(Y^\varepsilon_t) \circ \mathrm{d}B^k(t) + \sum_{i=1}^d V_k(Y^\varepsilon_t) \circ \mathrm{d}L^k(t) + \varepsilon K(Y^\varepsilon_t)\mathrm{d}t \\
&\quad (4.5)
\end{align*}
\]

with initial condition \( Y^\varepsilon_0 = y_0 \). Here \( y_0 \) is 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 \( d \)-torus and \( D \) is a relatively compact open set of \( \mathbb{R}^n \).

In the action-angle coordinate, \( X_t = \varphi(\theta, I_t) \), \( \theta \in \mathbb{T}^d \), \( I_t \in D \), and \( \mathrm{d}I \land \mathrm{d}\omega \) defines a symplectic structure on \( D \times \mathbb{T}^d \). Let \( \tilde{H}_k = H_k(\varphi(\theta, I)) \) be the induced Hamiltonian on \( \mathbb{T}^d \times D \), then, for \( i = 1, \ldots, d \),

\[
\begin{align*}
\dot{\theta}_k &= \frac{\partial \tilde{H}_k}{\partial I_i} = \omega^k(I), \\
\dot{I}_k &= -\frac{\partial \tilde{H}_k}{\partial \theta_i} = 0,
\end{align*}
\]

with \( \omega^k \) smooth functions. Rewrite the perturbation vector field as \( (K_\theta, K_I) \) on \( \mathbb{T}^d \times D \) with \( K_\theta = (K_{\theta}, \ldots, K_{\theta}^d) \) and \( K_I = (K_I^1, \ldots, K_I^d) \) the angle and action component, respec-
tively. We have the following form of the SDE on $\mathbb{T}^d \times D$:

\begin{align*}
\frac{d\theta}{\varepsilon} &= \omega_0(I_{\varepsilon}^\theta) \, dt + \sum_{k=1}^{d} \omega_k(I_{\varepsilon}^\theta) \circ dB^k(t) + \sum_{k=1}^{d} \omega_k(I_{\varepsilon}^\theta) \circ dL^k(t) + \varepsilon K_0(I_{\varepsilon}^\theta, \theta_{\varepsilon}^\theta) \, dt, \quad (4.6) \\
\frac{dI_{\varepsilon}^\theta}{\varepsilon} &= \varepsilon K_1(I_{\varepsilon}^\theta, \theta_{\varepsilon}^\theta) \, dt. \quad (4.7)
\end{align*}

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’s a situation where the averaging principle is to be expected to hold.

**Remark 4.5.** To study slow motion governed by the transversal part of the vector field $K$ it is convenient to rescale the time. Denote $Y_{t/\varepsilon}^\varepsilon$ the process scaled in time by $1/\varepsilon$ which coincides, in the sense of probability distributions [7], with $Y_{t}^\varepsilon$. As both of Stratonovitch integral and Marcus integral satisfy the Newton-Leibniz chain rule under a change of variable, $Y_{t/\varepsilon}^\varepsilon$ has a generator given by $\frac{1}{\varepsilon}A + L_K$. Then, the evolution of $Y_{t/\varepsilon}^\varepsilon$ is the skew product of the fast diffusion of order $1/\varepsilon$ along the invariant manifold and the slow diffusion of order 1 across the invariant manifold. The invariant manifolds here is actually $d$-dimensional torus and the motion on the torus, which would be quasi-periodic if there is no diffusion terms, is ergodic. Indeed, we obtain a new dynamical system in the limit as $\varepsilon$ 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 detail proof is shown in next subsection.

**Theorem 4.1.** (Averaging Principle) Consider the perturbed SDE (4.5) with initial value $Y_{t=0}^\varepsilon = y_0$. Set $H_i^\varepsilon(t) = H_i(Y_{t/\varepsilon}^\varepsilon)$, for $i = 1, 2, \ldots, d$. Define exit time $\tau^\varepsilon := \inf\{t \geq 0 : Y_{t/\varepsilon}^\varepsilon \notin U_0\}$ as the first time that the solution $Y_{t/\varepsilon}^\varepsilon$ starting form $y_0$ exists 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{equation}
\frac{d}{dt}\bar{H}(t) = \int_{M_{0_t}} \omega^2(V_i, K)(\bar{H}(t), z)\mu_{\bar{H}}(dz), \quad (4.8)
\end{equation}

at
with initial value \( \tilde{H}(0) = H(y_0) \). Define exit time \( \tau^0 := \inf\{t \geq 0 : \tilde{H}(t) \notin U_0\} \) as the first time that \( \tilde{H}(t) \) exists from \( U_0 \).

Then we have that:

1. For all \( t < \tau_0 \), \( p \geq 2 \) and \( \beta \in (0, 1) \), there exist constants \( C_1, C_2 > 0 \) and a continuous function \( \zeta(t, \varepsilon) = \varepsilon t \ln \varepsilon |t| (1 + t + |t|^2 \ln |t| |t| + t\eta(t \ln |t| |t|) \) which converges to zero when \( t \) or \( \varepsilon \) goes to zero, such that
   \[
   \left( \mathbb{E} \left[ \sup_{s \leq t} |H^\varepsilon(s \wedge \tau^\varepsilon) - \tilde{H}(s \wedge \tau^\varepsilon)|^p \right] \right)^{\frac{1}{p}} \leq C_1 \zeta(t, \varepsilon)e^{C_2 t}.
   \]  
   (4.9)

2. If there exists a \( r > 0 \) such that \( U_r := \{ x \in M : |H(x) - H(y_0)| \leq r \} \subset U_0 \), and define exit time \( \tau_\delta := \inf\{t \geq 0 : |\tilde{H}_t - H(Y_0)| > r - \delta\} \) for \( \delta > 0 \). Then for any \( p \geq 2 \) and constant \( C_3, C_4 \) depending on \( \tau_\delta \),
   \[
   \mathbb{P}(\tau^\varepsilon < \tau_\delta) \leq C_3 \delta^{-p} h(\tau_\delta, \varepsilon)p^p e^{C_4 \rho_\tau \varepsilon}.
   \]  
   (4.10)

**Remark 4.6.** This result includes the case of pure Gaussian noise and case of pure jump noise, where the former situation has been considered, cf. Li [18, Theorem 3.3.]. The main difference between Gaussian situation and the situation we considered here comes from the estimation for Lévy noise term. However, if perturbation is a (local) Hamiltonian vector field with \( \omega^2(V_i, K) = 0 \), we have to look at the further scaling. To deal with this problem on multiplicative Lévy noise is still remain to solve.

### 4.4. Proof of the averaging principle

Denote by \( X_t \) and \( Y_t^\varepsilon \) the solution to integrable stochastic Hamiltonian system (4.1) and corresponding perturbed system (4.5) respectively. We first get the information on the order of which the first integrals for the perturbed system change over a time interval \( t \) by next lemma.

**Lemma 4.3.** Let \( \tau^\varepsilon = \inf\{t \geq 0 : Y_t^\varepsilon \notin U_0\} \). For any Lipschitz test function \( f : M \rightarrow \mathbb{R} \) and \( p \geq 2 \), we have
   \[
   \left( \mathbb{E} \left[ \sup_{s \leq t \wedge \tau^\varepsilon} |f(Y_s^\varepsilon) - f(X_s)|^p \right] \right)^{\frac{1}{p}} \leq C \varepsilon (t + t^2),
   \]  
   (4.11)
where $C$ is a constant depending on upper bounds of the norms of the perturbing vector field $K$, on the (locally) Lipschitz coefficient of $f$ and on the derivatives of $V_0, V_1, ..., V_d$ with respect to the action-angle coordinate system on $T^d \times D$.

**Proof.** In action-angle coordinates, we rewrite the flows as $X_t = \varphi(\theta, I)$ and $Y_t^\varepsilon = \varphi(\theta^\varepsilon, I^\varepsilon_t)$ and denote the representation of $f$ in $T^d \times D$ by $\tilde{f}$. Since $D$ is relatively compact, $\partial \tilde{f} / \partial \theta$ and $\partial \tilde{f} / \partial I$ are bounded on $T^d \times D$. We thus obtain

$$|f(Y_t^\varepsilon) - f(X_t)| = |f \circ \varphi(\theta^\varepsilon, I^\varepsilon_t) - f \circ \varphi(\theta, I)|$$

$$\leq c_0|\theta^\varepsilon - \theta| + c_0|I^\varepsilon_t - I|,$$

for some constant $c_0 > 0$.

The corresponding SDE on $T^d \times D$ under the action-angle coordinate map is shown in (4.11)–(4.10). Then, the estimate of the action coordinate $|I^\varepsilon_t - I|:

$$\sup_{s \leq t < \tau^\varepsilon} |I^\varepsilon_t - I| \leq e \sup_{s \leq t < \tau^\varepsilon} \left| \int_0^s |K_t(\theta^\varepsilon_t, I^\varepsilon_t)|dr \right| \leq e \sup_{T^d \times D} |K_t|.$$

As for the estimate of the angle coordinate,

$$\theta^\varepsilon_t - \theta^\varepsilon_s = \int_s^t (\omega^\varepsilon_k(I^\varepsilon_r) - \omega^\varepsilon_k(I_r))dr + \sum_{k=1}^d \int_s^t (\omega^\varepsilon_k(I^\varepsilon_r) - \omega^\varepsilon_k(I_r)) \circ dB^k(r)$$

$$+ \sum_{k=1}^d \int_s^t (\omega^\varepsilon_k(I^\varepsilon_r) - \omega^\varepsilon_k(I_r)) \circ dL^k(r) + c_0 |\theta^\varepsilon_t - \theta|,$$

for $s < \tau^\varepsilon$, $1 \leq i \leq d$, where the stratonovich correction term is equal to 0 as $\omega_k(I)$ does
not depend on θ. For $p \geq 2$, we obtain

$$
|\theta^p - \theta_0|^p \leq c_1(d, p) \left( \int_0^\infty (\omega_0(t^p) - \omega_0(t)) dt \right)^p + \sum_{k=1}^d c_1(d, p) \left( \int_0^\infty (\omega_k(t^p) - \omega_k(t)) dB(t) \right)^p \\
+ \sum_{k=1}^d c_1(d, p) \left( \left| \int_0^\infty (\omega_k(t^p) - \omega_k(t)) dL^k(t) \right|^p + \left| \sup_{0 \leq r < t} \left[ (\phi(\Delta L^k(r), \omega_k(t^p) - \omega_k(t)) \right) \right|^p \right) \\
+ c_1(d, p) \bar{e} \left( \sup_{T^p < T} |K_0| \right)^p \\
= c_1(d, p)(\Lambda_0 + \sum_{k=1}^d \Lambda_{1k} + \sum_{k=1}^d (\Lambda_{2k} + \Lambda_{3k}) + \Lambda_4),
$$

(4.14)

where the $\Lambda_i$, $i = 0, 1k, 2k, 3k, 4$, denote the above terms respectively. Now we present some useful estimates for $\Lambda_i$. For simplicity, in the following we use the notation when there is no ambiguity: $\tilde{\omega}_k = \omega_k(t^p) - \omega_k(t)$, for $k = 0, 1, 2, ..., d$. Indeed, we have

$$
\sup_{r \in [0, T^p]} |\tilde{\omega}_k| \leq \sup_{r \in [0, T^p]} |d\omega_k| \cdot \sup_{r \in [0, T^p]} |t^p - I_r| \leq c_2(d\omega_k, K_1)\bar{e}t.
$$

(4.15)

Firstly, Hölder inequality yields:

$$
\mathbb{E} \left[ \sup_{s \in [0, T^p]} \Lambda_0 \right] = \mathbb{E} \left[ \sup_{s \in [0, T^p]} \left| \int_0^\infty \tilde{\omega}_0 dr \right|^p \right] \leq \mathbb{E} \left[ \sup_{s \in [0, T^p]} s^{p-1} \int_0^\infty |\tilde{\omega}_0|^p dr \right] \\
\leq c_3(p, d\omega_0, K_1)\bar{e}^p r^p.
$$

(4.16)

By using Itô isometry for the term $\Lambda_{1k}$,

$$
\mathbb{E} \left[ \sup_{s \in [0, T^p]} \sum_{k=1}^d \Lambda_{1k} \right] \leq \sum_{k=1}^d \mathbb{E} \left[ \int_0^{\wedge \wedge T^p} |\tilde{\omega}_k|^2 dr \right] \leq \sum_{k=1}^d \mathbb{E} \int_0^{\wedge \wedge T^p} |\tilde{\omega}_k|^p dr \\
\leq c_4(d, p, d\omega_k, K_1)\bar{e}^p t^p.
$$

(4.17)

Due to $L(t)$ is a Lévy motion with exponential moment, there exist moments of order $p \geq 1$. And the term $\Lambda_{2k}$ has the following representation [21] with respect to the compensated Possion random measure associated to $L(t)$,

$$
\Lambda_{2k} = \int_0^T \int_{\mathbb{R} \setminus \{0\}} \tilde{\omega}_k z \tilde{N}(dr, dz) + \int_0^T \int_{|z| \geq 1} \tilde{\omega}_k z v(dz) dr,
$$

here $\tilde{\omega}_k = \omega_k(t^p) - \omega_k(t)$. By Kunita’s first inequality ([21], Page 265) and Young
inequality, we obtain

$$\mathbb{E}[\sup_{\lambda \in \Lambda_{11}} \sum_{k=1}^{d} T_{k}]$$

\leq c_5(p) \left[ \sum_{k=1}^{d} \mathbb{E} \left[ \left( \int_{0}^{T_{k}} \int_{\mathbb{R}^{d} \backslash \{0\}} |z|^2 |\tilde{\omega}_k|^2 \nu(dz) dr \right)^{p/2} \right] \right] + c_6(p) \left[ \sum_{k=1}^{d} \mathbb{E} \left[ \left( \int_{0}^{T_{k}} \int_{\mathbb{R}^{d} \backslash \{0\}} |\tilde{\omega}_k|^2 \nu(dz) dr \right)^{p/2} \right] \right]$$

\leq c_5(p) \left( \sum_{k=1}^{d} \mathbb{E} \left[ \left( \int_{0}^{T_{k}} \int_{\mathbb{R}^{d} \backslash \{0\}} |\tilde{\omega}_k|^2 \nu(dz) dr \right)^{p/2} \right] \right) + c_6(p) \left( \sum_{k=1}^{d} \mathbb{E} \left[ \left( \int_{0}^{T_{k}} |\tilde{\omega}_k|^2 \nu(dz) dr \right)^{p/2} \right] \right)

\leq c_5(p)c_7(p,c_7(d,p,d\omega_k,K_r)e^{-\rho p+1} \int_{0}^{T_{k}} |\tilde{\omega}_k|^2 \nu(dz))$$

\leq c_6(p)c_8(d,p,d\omega_k,K_r)e^{-\rho p+1} \left( \int_{0}^{T_{k}} |\tilde{\omega}_k|^2 \nu(dz) \right)^{p/2}$$

By the definition of $\phi$ in (2.5), Taylor theorem and Gronwall’s lemma, we have the following estimation refer to [35, 25].

$$\sum_{0 \leq r \leq s} \left[ \phi(\Delta L^k_r(\omega_k(I_{k}^r), I_{k}^r), I_{k}^r) - \phi(\Delta L^k_r(\omega_k(I_{k}^r), I_{k}^r)) \right] - (I_{k}^r - I_{k}) \leq c_9 \sum_{0 \leq r \leq s} \left[ \phi(\Delta L^k_r(\omega_k(I_{k}^r), I_{k}^r)) - \phi(\Delta L^k_r(\omega_k(I_{k}^r), I_{k}^r)) \right]$$

$$\leq c_{10} \sum_{0 \leq r \leq s} |I_{k}^r - I_{k}||\Delta L^k_r| e^{|\lambda_k| |\Delta L_k|}.$$
binning with the assumption A2 of existence of exponential moments, we have

\[
\mathbb{E}[\sup_{s \in \mathbb{R}^d} |A_{3k}|] \leq c_{12}(p, d, dw_k, K_1)e^{p}p^\frac{d}{p} \left\{ \mathbb{E} \left[ \sup_{s \in \mathbb{R}^d} \left| \int_0^\tau \int_{\mathbb{R}^d} |y|^2 e^{2|y|^2 (\nu(dy))} \right|^p \right] \right. \\
\leq c_{12}(p, d, dw_k, K_1)e^{p}p^\frac{d}{p} \left[ c_{13}(p) \left( \int_0^\tau \int_{\mathbb{R}^d} |y|^4 e^{2|y|^2 (\nu(dy))}dr \right)^{p/2} + c_{13}(p) \left( \int_0^\tau \int_{\mathbb{R}^d} |y|^2 e^{2|y|^2 (\nu(dy))}dr \right)^p \right] \\
\leq c_{14}(p, d, dw_k, K_1)e^{p}p^\frac{d}{p} \left. \left( t^p + t^{p+1} + t^{2p} \right) \right].
\]

(4.19)

At last,

\[
\mathbb{E}[\sup_{s \in \mathbb{R}^d} A_4] = e^{p}p^\frac{d}{p} \left( \sup_{T \in \mathbb{R}} |K_0| \right)^p. \tag{4.20}
\]

Taking the supremum and expectation in inequality (4.14) and combining the estimates (4.16), (4.17), (4.18), (4.19) and (4.20) with the fact \( p + 1 \leq \frac{3p}{2} \leq 2p \) for \( p \geq 2 \), we obtain

\[
\mathbb{E}[\sup_{0 < t < T} |\theta_1 - \theta_0|^p] \leq c_0'e^{p}p^\frac{d}{p} + c_1'e^{p}p^\frac{d}{p}t^p + c_2'e^{p}p^\frac{d}{p}(t^p + t^{p+1} + t^{2p}) \\
+ c_3'e^{p}p^\frac{d}{p}(t^p + t^{p+1} + t^{2p}) + c_4'e^{p}p^\frac{d}{p} \leq c'e^{p}(t + T^2)^p.
\]

(4.21)

Eventually, the desired result follows from Minkowski’s inequality and the estimates (4.13) and (4.21).

\[
\mathbb{E} \left( \sup_{s \in \mathbb{R}^d} |f(Y_s) - f(X_s)|^p \right)^{\frac{1}{p}} \leq C \varepsilon (t + T^2) \tag{4.22}
\]

We thus find that the slow component accumulate over a time interval of the size \( t/\varepsilon \). 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 dynamic system as \( \varepsilon \) goes to zero. For convenience, we adopt the following notation: Let \( g : M \to \mathbb{R} \) be a continuous function on \( M \), we define its average over \( M \) as \( Q^g : D \subset \mathbb{R}^d \to \mathbb{R} \), i.e.,

\[
Q^g(h) = \int_{M_h} g(h, z)\mu_h(dz)
\]  

(4.23)

Then, the estimation of averaging error is given in the following lemma.

**Lemma 4.4.** Suppose that \( g \) is continuous on \( U_0 \). Set \( H^\varepsilon_i(s) = H_i(Y^\varepsilon_{t_i/\varepsilon}) \) and \( \forall \varepsilon \) defined by

\[
\delta^\varepsilon(g, t) = \int_{s/\varepsilon}^{s/\varepsilon + \tau^\varepsilon} g(Y^\varepsilon_{t/\varepsilon})dr - \int_{s/\varepsilon}^{s/\varepsilon + \tau^\varepsilon} Q^g(H^\varepsilon_i(r))dr
\]

(4.24)

the averaging error. Then \( \delta^\varepsilon(g, t) \to 0 \), as \( t \) or \( \varepsilon \to 0 \).

More precisely, let \( \eta(t) \) be the rate of convergence of ergodic unperturbed dynamic which is a function of \( C_b([0, \infty), [0, \infty)) \) satisfying \( \eta(t) \searrow 0 \) as \( t \to \infty \). There exists a constant \( C_0 > 0 \) such that, for all \( p \geq 2, \beta \in (0, 1) \), we have the following estimate

\[
(\mathbb{E}[\sup_{s \leq t} |\delta^\varepsilon(g, s)|^p])^{1/p} \leq C_0 \zeta(t, \varepsilon)
\]

(4.25)

where \( \zeta(t, \varepsilon) := \varepsilon t |\ln \varepsilon|^\beta (1 + t^2 |\ln \varepsilon|^\beta) + t |\ln \varepsilon|^\beta \) is continuous for \( t \geq 0, \varepsilon > 0 \) and converges to zero when \( t \) or \( \varepsilon \to 0 \).

**Proof.** We recall that \( \tau^\varepsilon \) denotes the exit time of \( Y^\varepsilon \) from \( U_0 \). The proof consists of considering a convenient partition of the interval \((0, t/\varepsilon \wedge \tau^\varepsilon)\) where we can get the estimates by comparing in each subinterval the average of the flow of the original system with the average of the perturbed flow.

For \( \varepsilon \) small enough and \( t \geq 0 \) we define the partition

\[
t_0 = 0 < t_1 < \ldots < t_N = \frac{t}{\varepsilon} \wedge \tau^\varepsilon
\]

with the following assignment of increments:

\[
\Delta_t = t |\ln \varepsilon|^\beta , \beta \in (0, 1).
\]

The grid points of the partition are given by \( t_n = n\Delta_t \) for \( 0 \leq n \leq N_\varepsilon \) with \( N_\varepsilon = [\varepsilon^{-1}|\ln \varepsilon|^{-\beta}] \) where the bracket function \([\cdot]\) denotes the integer part of the value.
Initially we represent the left hand side as the sum:

\[ \int_0^{t \wedge \tau^e} g(Y^e_{ij})dt = \epsilon \int_0^{t \wedge \tau^e} g(Y^e_t)dt = \epsilon \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} g(Y^e_t)dt \epsilon \int_{J_n}^{t \wedge \tau^e} g(Y^e_t)dt. \tag{4.26} \]

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 \( \Theta_t \) the shift operator on the canonical probability space, i.e., \( \Theta_t(\omega)(-\tau) = \omega(-\tau) - \omega(\tau) \). Then,

\[
|\delta^\tau(\epsilon, t)| \leq \epsilon \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} |g(Y^e_t) - g(\Psi_{t-l}(\Theta_l(\omega), Y^e_t))|dt |\epsilon| \int_{J_n}^{t \wedge \tau^e} Q^e(H^e(\epsilon t))dt| + \epsilon \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} Q^e(H^e(\epsilon t)) - \int_0^{t \wedge \tau^e} Q^e(H^e(r))dr|dt|g(Y^e_t)\epsilon| \int_{J_n}^{t \wedge \tau^e} g(Y^e_t)dt| = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4. \tag{4.27}\]

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 4.8 and Hölder’s inequality,

\[
(\mathbb{E} \sup_{x \in \mathbb{R}} |\Sigma_1|^p)^{\frac{1}{p}} \leq \epsilon \sum_{n=0}^{N(t) - 1} (\mathbb{E} \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} \sup_{l \in [s, t]} |g(Y^e_t) - g(\Psi_{t-l}(\Theta_l(\omega), Y^e_t))|dt |\epsilon| \int_{J_n}^{t \wedge \tau^e} Q^e(H^e(\epsilon t))dt|^{p/2}^{\frac{1}{2}} + \epsilon \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} Q^e(H^e(\epsilon t)) - \int_0^{t \wedge \tau^e} Q^e(H^e(r))dr|dt|g(Y^e_t)\epsilon| \int_{J_n}^{t \wedge \tau^e} g(Y^e_t)dt|^{p/2}^{\frac{1}{2}}
\]

\[
\leq \epsilon \sum_{n=0}^{N(t) - 1} (\Delta_{t} \epsilon^{p-1} \mathbb{E} \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} |g(Y^e_t) - g(\Psi_{t-l}(\Theta_l(\omega), Y^e_t))|^{p}dt|^{1/2}^{1/2}) + \epsilon \sum_{n=0}^{N(t) - 1} (\Delta_{t} \epsilon^{p-1} \mathbb{E} \sum_{n=0}^{N(t) - 1} \int_{J_n}^{J_{n+1}} |g(Y^e_t) - g(\Psi_{t-l}(\Theta_l(\omega), Y^e_t))|^{p}dt|^{1/2}^{1/2})
\]

\[
\leq \epsilon \sum_{n=0}^{N(t) - 1} (\Delta_{t} \epsilon^{p-1} \mathbb{E} \sup_{l \in [s, t]} |g(Y^e_t) - g(\Psi_{t-l}(\Theta_l(\omega), Y^e_t))|^{p}dt|^{1/2}^{1/2}) + \epsilon N \Delta_{t} \epsilon^{p-1} \epsilon (\Delta_{t} \epsilon^{p-1} + (\Delta_{t} \epsilon^{p-1}))
\]

\[
\leq c \epsilon^2 \epsilon^{-1} \ln |\epsilon|^{\frac{1}{2}} |\epsilon| \ln |\epsilon|^{\frac{3}{2}} (1 + t \ln |\epsilon|^{\frac{1}{2}})
\]

\[
= c \epsilon^2 |\epsilon| \ln |\epsilon|^{\frac{1}{2}} (1 + t \ln |\epsilon|^{\frac{1}{2}}). \tag{4.28}\]

Recall that, in Brownian case, the law of large numbers indicates, for any function \( f \) on the compact manifold converging to infinity when \( t \) converges to infinity,

\[
\frac{1}{t} \int_{s}^{t} f(x)dt \to \int_{M} f(z)dz \text{ as } t \to \infty, \text{ in } L^p \text{ } (p > 1), \text{ and the rate of convergence is}
\]
In Lévy case, such a rate of convergence should be consider as a positive, bounded, 
decreasing function from \([0, \infty)\) to \([0, \infty)\) which it’s just \(\eta(t)\) here. More information on 
rates of convergence can be found in Hölgele [25] and Kulik [37], and a detail example 
is shown in subsection 4.5.

Denote by \(\mu_{\nu}\) or \(\mu_{\nu}^\gamma\), the invariant measure on the invariant manifold \(M_{\nu} = 
M_{\nu}^\gamma\). By Markov property of the flow, we obtain

\[
\left(\mathbb{E} \sup_{s \leq t} |\Sigma_s|^p\right)^{\frac{1}{p}} \leq e \sum_{n=0}^{N-1} \left(\mathbb{E} \sup_{s \leq t} |g(\Psi_{t_n}(\Theta_n, \omega), Y_s^\nu)| \right)^{\frac{1}{p}} \\
\leq e^{-\Delta t} \sum_{n=0}^{N-1} \left(\mathbb{E} \sup_{s \leq t} |g(\Psi_{t_n}(\Theta_n, \omega), Y_s^\nu)| \right)^{\frac{1}{p}} \\
\leq eN \Delta t \sup_n \left(\mathbb{E} \|H(\nu)\| \right)^{\frac{1}{p}} \\
- \int_{\sqrt{N \Delta t}} g(H(\nu), \tau) d\mu_{\nu}^{\gamma}(\tau)^{\frac{1}{p}} \\
\leq c\epsilon N \Delta t \eta(\Delta t) \leq c\epsilon \left[\epsilon^{-1} \ln |t| \frac{1}{p} \ln |t| \ln |t| \frac{1}{p} \right] \\
= c\epsilon \eta(t) \ln |t| \frac{1}{p}. \quad (4.29)
\]

It’s not hard to find that \(Q^\nu\) is Lipschitz continuous on \(T^d \times D\). We thus have

\[
\left(\mathbb{E} \sup_{s \leq t} |\Sigma_s|^p\right)^{\frac{1}{p}} \leq e \sum_{n=0}^{N-1} \Delta t \mathbb{E} \sup_{\nu \in [0, \infty)} |Q^\nu(H(\nu)) - Q^\nu(H(\nu + \Delta t))| \right)^{\frac{1}{p}} \\
\leq c\epsilon N \Delta t \sup_n \|I^\nu - I^\nu\| \\
\leq cN \epsilon^2 \Delta t \\
\leq ct \epsilon^{-1} \ln |t| \frac{1}{p} \eps^2 (t) \ln |t| \frac{1}{p} \\
= c\epsilon t \ln |t| \frac{1}{p}. \quad (4.30)
\]

Moreover, since \(g\) is \(C^1\) on \(U_0\),

\[
\left(\mathbb{E} \sup_{s \leq t} |\Sigma_s|^p\right)^{\frac{1}{p}} \leq c\epsilon t \ln |t| \frac{1}{p}. \quad (4.31)
\]

In conclusion, note that each of the four estimates of (4.28 - 4.31) is determined by 
a continuous function which goes to zero as \(t\) or \(\epsilon \to 0\). Lemma 4.24 now follows by 
inequality (4.27) and Minkowski’s inequality. \(\Box\)

29
At last, we present the proof of Theorem 4.1 based on the results of Lemma 4.3 and Lemma 4.4.

**Proof of Theorem 4.1.** By Itô’s formula and Lemma 4.1, for \( t < \tau_0 \wedge \tau^\varepsilon, \quad 1 \leq i \leq d \),

\[
H_i^\varepsilon(t) = H_i(Y_0) + \int_0^t \omega^2(V_i, K)(Y_s^\varepsilon) ds.
\] (4.32)

For \( i \) fixed, we write

\[
g_i = \omega^2(V_i, K)
\] (4.33)

which is \( C^1 \) on \( U_0 \). And thus we only need to estimate

\[
|H_i^\varepsilon(t) - \bar{H}_i(t)| = \left| \int_0^t g_i(Y_s^\varepsilon) ds - \bar{H}_i(t) \right|.
\] (4.34)

Applying (4.24) to the functions \( g_i \), we have, for \( t < \tau^\varepsilon \),

\[
\int_0^{t \wedge \tau^\varepsilon} g(Y_s^\varepsilon) ds = \int_0^{t \wedge (\tau^\varepsilon / \varepsilon)} Q^\varepsilon(H_i^\varepsilon(s)) ds + \delta^\varepsilon(\varepsilon, t). \] (4.35)

On the other hand, using the notation of the previous lemma, the equation (4.8) can be written as

\[
\frac{d}{dt} \bar{H}_i(t) = Q^\varepsilon(\bar{H}_i(t)),
\]

\[
\bar{H}_0(t) = H(Y_0).
\]

Therefore, for any \( t < \tau^\varepsilon \), we obtain

\[
|H_i^\varepsilon(t \wedge \tau^\varepsilon) - \bar{H}_i(t \wedge \tau^\varepsilon)| \leq \int_0^{d \wedge \tau^\varepsilon} \left| Q^\varepsilon(H_i^\varepsilon(s)) - Q^\varepsilon(\bar{H}_i(s)) \right| ds + \delta(g, \varepsilon, t)
\]

\[
\leq C(g, \varphi) \int_0^{d \wedge \tau^\varepsilon} |H_i^\varepsilon(s) - \bar{H}_i(s)| ds + \delta(g, \varepsilon, t) \] (4.36)

where \( C(g, \varphi) \) is the Lipschitz constant.

By Lemma 6.3 and Gronwall’s inequality, there is a continuous function \( \zeta(t, \varepsilon) \) which converges to zero when \( t \) or \( \varepsilon \to 0 \), such that

\[
\left( \sup_{s \leq t} |H_i^\varepsilon(s \wedge \tau^\varepsilon) - \bar{H}_i(s \wedge \tau^\varepsilon)|^p \right)^{\frac{1}{p}} \leq C_1 \zeta(t, \varepsilon) e^{C_2(g, \varphi) t}.
\] (4.37)
For the second part of the theorem, we have the following estimate by the definition of $\tau^\epsilon$, $\tau_\delta$ and Chebychev’s inequality,

\[
\mathbb{P}(\tau^\epsilon < \tau_\delta) \leq \mathbb{P}(\sup_{s \in \tau^\epsilon \land \tau_\delta} |\tilde{H}(s) - H^\epsilon(s)| > \delta) \\
\leq \delta^{-p}\mathbb{E}\left[ \sup_{s \in \tau^\epsilon \land \tau_\delta} |\tilde{H}(s) - H^\epsilon(s)|^p \right] \\
\leq C_3 \delta^{-p} \xi(\tau_\delta, \epsilon)^p e^{C_4(g(\epsilon)) p \tau_\delta}, \quad (4.38)
\]

\[
4.5. \text{An example: Perturbed stochastic harmonic oscillator with Lévy noise}
\]

In this subsection, let’s present a simple illustrative example for the above averaging principle of integrable stochastic Hamiltonian system with Lévy noise. We write $(q, p) = (q_1, ..., q_d, p_1, ..., p_d)$ 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. $V(q) = \frac{1}{2} \sigma |q|^2$ with $\sigma$ a frequency, then we have the linear harmonic oscillator. Given Hamiltonian functions as follow,

\[
H_1 = \frac{1}{2} \sum_{i=1}^d p_i^2 + \frac{1}{2} \sum_{i=1}^d \sigma_i^2 p_i^2, \\
H_k = \frac{1}{2} \frac{p_i^2}{\sigma_k} + \frac{1}{2} \sigma_k p_k^2, \quad k = 2, ..., d,
\]

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

\[
dq_i(t) = \frac{\partial H_0}{\partial p_i} dt + \sum_{i=1}^d \frac{\partial H_k}{\partial p_i} \circ dB_i^k + \sum_{i=1}^d \frac{\partial H_k}{\partial p_i} \circ dL_k^i, \quad (4.39)
\]

\[
 dp_i(t) = -\frac{\partial H_0}{\partial q_i} dt - \sum_{i=1}^d \frac{\partial H_k}{\partial q_i} \circ dB_i^k - \sum_{i=1}^d \frac{\partial H_k}{\partial q_i} \circ dL_k^i, \quad (4.40)
\]

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

\[ dX_i = J\nabla H_0(X_i)dt + \sum_{i=1}^{d} J\nabla H_k(X_i) \circ dB^i_k + \sum_{i=1}^{d} J\nabla H_k(X_i) \circ dL^i_k. \]  

(4.41)

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

\[ q_i = \sqrt{2I_i} \cos \theta_i, \quad p_i = \sqrt{2\sigma_i I_i} \sin \theta_i, \]

then the induced Hamiltonians \( H'_k = H_k(\varphi(\theta, I)) = \left\{ \begin{array}{ll} \sum_{i=1}^{d} \sigma_i I_i, & k = 1 \\ I_k, & k = 2, \ldots, d \end{array} \right\} \) on \( T^d \times D \) satisfy,

\[ \dot{\theta}_k = \frac{\partial H'_k}{\partial I_k} = : \omega^i_k(I), \quad \begin{cases} \sigma_i, & k = 1; \\ 1, & k = 2, \ldots, d \text{ & } i = k; \\ 0, & \text{otherwise.} \end{cases} \]

\[ I^i_k = \frac{\partial H'_k}{\partial \theta_i} = 0. \]

Next, if this integrable stochastic Hamiltonian system (4.41) is subjected to a small non-Hamiltonian perturbation, we try to study the corresponding averaging principle.

For simplicity, we consider \( d = 2, \sigma = 1 \) and perturbation vector field is \( K = (0, q_2/(q_2^2 + p_2^2), 0, 0)^T \).

\[ d \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} = d \begin{pmatrix} p_1 & 0 \\ p_2 & p_2 \\ -q_1 & 0 \\ -q_2 & -q_2 \end{pmatrix} \circ d \begin{pmatrix} B^1_1 \\ B^2_1 \\ B^2_2 \\ -q_2 \\ -q_2 \end{pmatrix} \circ d \begin{pmatrix} L^1_1 \\ L^2_1 \\ L^2_2 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 \\ q_2/(q_2^2 + p_2^2) \\ 0 \end{pmatrix} dt. \]

By action-angle coordinates change,

\[ d \begin{pmatrix} q_i \\ p_i \end{pmatrix} = \begin{pmatrix} \frac{\partial q_i}{\partial \theta_i} & \frac{\partial q_i}{\partial I_i} \\ \frac{\partial p_i}{\partial \theta_i} & \frac{\partial p_i}{\partial I_i} \end{pmatrix} d \begin{pmatrix} \theta_i \\ I_i \end{pmatrix}, \]

(4.42)

\[ \Longrightarrow d \begin{pmatrix} \theta_i \\ I_i \end{pmatrix} = \begin{pmatrix} \frac{1}{2\sqrt{I_i}} \sin \theta_i & -\frac{1}{\sqrt{2I_i}} \cos \theta_i \\ -\frac{1}{\sqrt{2I_i}} \cos \theta_i & -\frac{1}{\sqrt{2I_i}} \sin \theta_i \end{pmatrix} d \begin{pmatrix} q_i \\ p_i \end{pmatrix}. \]

(4.43)
Thus, \( \varphi(t, I_0) \):

\[
\begin{pmatrix}
\vartheta_1 \\
\vartheta_2 \\
I_1 \\
I_2
\end{pmatrix}
= d
\begin{pmatrix}
1 & 0 \\
1 & 1 \\
0 & 0
\end{pmatrix}
+ d
\begin{pmatrix}
0 & 0 \\
1 & 1 \\
0 & 0
\end{pmatrix}
+ E
\begin{pmatrix}
0 & \frac{1}{2r}
\sin \vartheta_2 \cos \vartheta_2 \\
0 & -\cos^2 \vartheta_2
\end{pmatrix}
dt.
\]

For unperturbed system, it is easy to get fundamental solution with initial condition \((q_0, p_0) = \varphi(\theta_0, I_0)\):

\[
\begin{align*}
\vartheta_i &= \vartheta_0 \Lambda(B_i + L_i) \\
I_i &= I_0
\end{align*}
\]

\( i = 1, 2 \)

Thus,

\[
Q^\vartheta(h_i) = \int_0^{q_2^2 + p_2^2} \cos^2 \vartheta_2 d\vartheta_1 d\vartheta_2 = \frac{1}{2}.
\]

Next, we verify \( \int_0^t g_i(q_j, p_j)ds \rightarrow Q^\vartheta(h_i) \) in \( L^2 \), as \( t \rightarrow \infty \), with a rate of convergence \( \eta(t) \).

Recall that, by Lévy-Khintchine formula \([21, 28]\) the characteristic function for Lévy motion in \( \mathbb{R}^d \) is

\[
\mathbb{E}e^{iu(B_t)} = e^{\eta_0(u)}, u \in \mathbb{R}^d,
\]

where \( \eta_0(u) = \int_{\mathbb{R}^d, \{0\}} [e^{iu.\zeta} - 1 - i(u \cdot \zeta) \nu(d\zeta)] \) whose real part \( \Re \eta_0 \leq 0 \). And the characteristic function for standard Brownian motion in \( \mathbb{R}^d \) is

\[
\mathbb{E}e^{iu(u, B_t)} = e^{-\frac{1}{2} A(u, u)}, u \in \mathbb{R}^d.
\]

Therefore,

\[
\begin{align*}
\mathbb{E}\left[ \frac{1}{t} \int_0^t g_i(q_s, p_s)ds \right]
&= \mathbb{E}\left[ \frac{1}{t} \int_0^t \tilde{g}_i(\vartheta_s, I_s)ds \right]
= \mathbb{E}\left[ \frac{1}{t} \int_0^t \cos^2(\langle u, B_s + L_s \rangle)ds \right]

&= \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 \mathbb{E}e^{(2u, B_s)} \mathbb{E}e^{(2u, L_s)}ds + \frac{1}{2}

&= \frac{1}{2t} \int_0^t e^{-\frac{i}{2}(2u_1^2 - \Re \eta_0(2u))}ds + \frac{1}{2} = \frac{1}{2t} \mathbb{E}(1 - e^{-At}) + \frac{1}{2}.
\end{align*}
\] (4.45)
Here \( u = (1, 1)^T \in \mathbb{R}^2 \) and \( A = \frac{1}{4}|2u|^2 - \Re \eta_0(2u) > 0 \). Hence, as \( t \) goes to \( \infty \), the expectation is equal to \( \frac{1}{t} \) eventually. Next, we calculate the secondary moment as following.

\[
\mathbb{E}\left\{ \frac{1}{t} \int_0^t g(q_s, p_s)\,ds \right\}^2 = \mathbb{E}\left\{ \frac{1}{t^2} \int_0^t \cos^2(\langle u, B_s + L_s \rangle)\,ds \right\}^2 \\
= 2 \mathbb{E}\left[ \int_0^\infty \int_0^\infty \cos^2(\langle u, B_s + L_s \rangle)\cos^2(\langle u, B_r + L_r \rangle)\,ds\,dr \right] \\
= \frac{1}{4t^2} \int_0^\infty \int_0^\infty \mathbb{E}\left[ \Re e^{i(2u, B_s+L_s)+(B_r+L_r)} + \Re e^{i(2u, B_r+L_r)-(B_s+L_s)} \right. \\
\left. + 2\Re e^{i(2u, B_s+L_s)} + 2\Re e^{i(2u, B_r+L_r)} + 2 \right] ds\,dr \\
= \frac{1}{4t^2} \int_0^\infty \int_0^\infty \left[ e^{-\frac{1}{2}(2u)^2 - \Re \eta_0(2u)} + e^{-\frac{1}{2}(2u)^2 - \Re \eta_0(2u)} + 2 e^{-\frac{1}{2}(2u)^2 - \Re \eta_0(2u)} + 2 \right] ds\,dr \\
= \frac{1}{4t^2} \int_0^\infty \int_0^\infty \left[ e^{-A^2 B_s} + e^{-A^2 C(r-s)} + 2 e^{-A^2} + 2 e^{-A^2} + 2 \right] ds\,dr \\
= \frac{1}{4t^2} \int_0^\infty \frac{1}{A(A + B)} - \frac{1}{C^2} + \frac{1}{A + B} t + \frac{2}{A} t + \frac{e^{-(A+B)t}}{B(A + B)} - \frac{e^{-A^2 C}}{AB} + \frac{e^{-2A^2}}{C^2} - \frac{2te^{-At}}{A} \right]. \\
(4.46)
\]

Here we used the stationary independent increments property of the Brownian motion and Lévy motion. By Taylor expansion [22, Page 40] with \( u = (1, 1)^T \), we can find that \( B = \frac{1}{2}|4u|^2 - \frac{1}{2}|2u|^2 - \Re \eta_0(4u) + \Re \eta_0(2u) > 0 \), so we have

\[
\mathbb{E}\left[ \left| \frac{1}{t} \int_0^t g(q_s, p_s)\,ds \right|^2 \right] \to \frac{1}{t} \text{ as } t \to \infty. 
\]

Thus,

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

Moreover, combining (4.45), (4.46) 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).

With initial condition \((q_0, p_0) = \varphi(t_0, I_0)\), the transversal system stated in Theorem 4.1 is \( B(t) = \frac{1}{t} \). Therefore, the result guarantees that the radial part \( H^r(t) \) of the
accelerated time scale $\frac{1}{\varepsilon}$ has a local behavior close to $\frac{1}{\varepsilon}$ in the sense that

$$\left(\mathbb{E}\left[\sup_{s \in [0, t]} |H^t(s) - \frac{t}{2}| \right]\right)^{\frac{1}{2}} \leq C_1 |\varepsilon| \ln |\varepsilon| \left(1 + t + t^2 \ln |\varepsilon| + c t^2 \ln |\varepsilon| \right)e^{C_2 t}$$

(4.47)

tends to 0 when $\varepsilon \to 0$, for any $\beta \in (0, 1)$ and the constant $C_1, C_2 > 0$.

**Appendix: Proof of Theorem 3.1 and Theorem 3.2**

We now give the proof of the theorem 3.1 and theorem 3.2 which are based on the formula of change of variables in differential forms.

**Proof of Theorem 3.1**

Noticing that

$$dP \wedge dQ = \sum_{i=1}^{n} dP^i \wedge dQ^i$$

$$= \sum_{i=1}^{n} \sum_{l=r+1}^{n} \sum_{r=1}^{n} \left[ \frac{\partial P^i}{\partial p^r} \frac{\partial Q^r}{\partial q^l} - \frac{\partial P^i}{\partial q^l} \frac{\partial Q^r}{\partial p^r} \right] dp^r \wedge dq^l + \frac{\partial P^i}{\partial q^l} \frac{\partial Q^r}{\partial p^r} dp^r \wedge dq^l$$

$$+ \sum_{i=1}^{n} \sum_{l=r+1}^{n} \sum_{r=1}^{n} \left[ \frac{\partial P^i}{\partial q^l} \frac{\partial Q^r}{\partial p^r} - \frac{\partial P^i}{\partial p^r} \frac{\partial Q^r}{\partial q^l} \right] dp^r \wedge dq^l,$$

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

$$\sum_{i=1}^{n} \frac{D(P^i, Q^i)}{D(p^r, p^r)} = 0, \ r \neq l,$$

$$\sum_{i=1}^{n} \frac{D(P^i, Q^i)}{D(q^l, q^l)} = 0, \ r \neq l,$$

(4.48)

$$\sum_{i=1}^{n} \frac{D(P^i, Q^i)}{D(p^r, q^l)} = \delta_{il}, \ r, l = 1, ..., n.$$

Clearly,

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

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

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

Therefore, (4.48) is fulfilled if and only if

$$\sum_{i=1}^{n} d \frac{D(P^i(t), Q^i(t))}{D(p^r, p^r)} = \sum_{i=1}^{n} d \frac{D(P^i(t), Q^i(t))}{D(q^r, q^r)} = \sum_{i=1}^{n} d \frac{D(P^i(t), Q^i(t))}{D(p^r, q^l)} = 0.$$

(4.49)
3.2), we obtain

\[ P^i = \frac{\partial P^i}{\partial p^j}, \quad Q^i = \frac{\partial Q^i}{\partial q^j} \]

Introduce the notation

For a fixed \( r \), by calculating at \((P, Q)\) with \( P = P(t) = (P^1(t; t_0, p, q), \ldots, P^n(t; t_0, p, q)) \)
and \( Q = Q(t) = (Q^1(t; t_0, p, q), \ldots, Q^n(t; t_0, p, q)) \) which is a solution to systems 3.1-3.2, we obtain \( P^r_i, Q^r_i, i = 1, \ldots, n \), satisfy the following system of SDEs:

\[
\begin{align*}
\frac{dP^r_i}{dt} &= \sum_{j=1}^{n} \left( \frac{\partial f^i}{\partial p^j} P^r_j + \frac{\partial f^i}{\partial q^j} Q^r_j \right) dt + \sum_{k=1}^{m} \sum_{j=1}^{n} \left( \frac{\partial \sigma_k^i}{\partial p^j} P^r_j + \frac{\partial \sigma_k^i}{\partial q^j} Q^r_j \right) \circ dL^k, \quad P^r_i(t_0) = \delta_{ir}, \\
\frac{dQ^r_i}{dt} &= \sum_{j=1}^{n} \left( \frac{\partial g^i}{\partial p^j} P^r_j + \frac{\partial g^i}{\partial q^j} Q^r_j \right) dt + \sum_{k=1}^{m} \sum_{j=1}^{n} \left( \frac{\partial \gamma_k^i}{\partial p^j} P^r_j + \frac{\partial \gamma_k^i}{\partial q^j} Q^r_j \right) \circ dL^k, \quad Q^r_i(t_0) = 0,
\end{align*}
\]

where

\[
\begin{align*}
f(Q, P) &= \frac{\partial H}{\partial p}(Q, P), \quad \sigma_k(Q, P) = \frac{\partial H_k}{\partial p}(Q, P), \\
g(Q, P) &= \frac{\partial H}{\partial q}(Q, P), \quad \gamma_k(Q, P) = \frac{\partial H_k}{\partial q}(Q, P),
\end{align*}
\]

for \( k = 1, \ldots, m \).

Then, we get

\[
\frac{dP^r_i}{dt} Q^r_j(t) = \sum_{j=1}^{n} \left[ \frac{\partial f^i}{\partial p^j} P^r_j + \frac{\partial f^i}{\partial q^j} Q^r_j + \frac{\partial g^i}{\partial p^j} P^r_j + \frac{\partial g^i}{\partial q^j} Q^r_j \right] dt + \sum_{k=1}^{m} \sum_{j=1}^{n} \left( \frac{\partial \sigma_k^i}{\partial p^j} P^r_j + \frac{\partial \sigma_k^i}{\partial q^j} Q^r_j \right) \circ dL^k.
\]

Similarly, we can also calculate \( \frac{dP^r_i}{dt} Q^r_j(t) \), then

\[
\sum_{i=1}^{n} d \frac{D(P^r_i(t), Q^r_j(t))}{D(p^r, p^j)} = \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} \xi_1 dt + \sum_{k=1}^{m} \sum_{j=1}^{n} \xi_2 \circ dL^k \right],
\]

36
Noticing that relations (4.51 - 4.52) imply (4.54 - 4.55), we obtain

Similarly, we prove that the conditions (4.51 - 4.52) ensure the other two terms of (4.49) as well. This completes the proof.

\[ \text{Proof of Theorem 3.2.} \]

We calculate the derivatives of \( S \) with respect to \( q_0 \) and \( q_1 \):

\[
\frac{\partial S}{\partial q_0} = \int_0^t \left( \frac{\partial L}{\partial q} \frac{\partial q}{\partial q_0} + \frac{\partial L}{\partial \dot{q}} \frac{\partial \dot{q}}{\partial q_0} + \frac{\partial L}{\partial \dot{p}} \frac{\partial \dot{p}}{\partial q_0} + \frac{\partial L}{\partial \dot{\dot{q}}} \frac{\partial \dot{\dot{q}}}{\partial q_0} \right) dt
- \sum_{k=1}^d \int_0^t \left( \frac{\partial H_k}{\partial q} \frac{\partial q}{\partial q_0} + \frac{\partial H_k}{\partial \dot{p}} \frac{\partial \dot{p}}{\partial q_0} \right) \cdot L^k(t) dt
\]

\[
= \left[ \frac{\partial L}{\partial q} \frac{\partial q}{\partial q_0} \right]_{q_0 = 0}^{q_0 = t} + \int_0^t \left( \frac{\partial L}{\partial \dot{q}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \sum_{k=1}^d \frac{\partial H_k}{\partial q} \cdot L^k(t) \frac{\partial q}{\partial q_0} \right) dt
+ \left[ \frac{\partial L}{\partial \dot{p}} \frac{\partial \dot{p}}{\partial q_0} \right]_{q_0 = 0}^{q_0 = t} + \int_0^t \left( \frac{\partial L}{\partial \dot{p}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{p}} - \sum_{k=1}^d \frac{\partial H_k}{\partial \dot{p}} \cdot L^k(t) \frac{\partial \dot{p}}{\partial q_0} \right) dt
= -p_0^T, \tag{4.56}
\]

where the last equality follows from the stochastic Lagrange equations (3.23 - 3.24) and the Legendre transform \( p = \frac{\partial L}{\partial q} \).
Similarly, we have
\[ \frac{\partial S}{\partial q_1} = -p_1^T. \] (4.57)

Therefore,
\[ dS = -p_0^Tdq_0 + p_1^Tdq_1. \] (4.58)

Moreover,
\[ dp_1 \wedge dq_1 = d\left( \frac{\partial S}{\partial q_1} \right) \wedge dq_1 = \frac{\partial^2 S}{\partial q_1 \partial q_0} dq_0 \wedge dq_1, \] (4.59)
\[ dp_0 \wedge dq_0 = d(-\frac{\partial S}{\partial q_0}) \wedge dq_0 = \frac{\partial^2 S}{\partial q_0 \partial q_1} dq_0 \wedge dq_1. \] (4.60)

Smoothness of $L$ and the $H_k (k = 1, ..., d)$ in $S$ ensures that $\frac{\partial^2 S}{\partial q_1 \partial q_0} = \frac{\partial^2 S}{\partial q_0 \partial q_1}$, which implies
\[ dp_1 \wedge dq_1 = dp_0 \wedge dq_0. \] (4.61)

The proof is thus complete. \qed

Acknowledgements

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 NSF grant 1620449, and NSFC grants 11531006 and 11771449.

References

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 et fils, 1893.

[6] F. T. Tverer, Deriving the Hamilton Equations of Motion for a Nonconservative System Using A Variational Principle, Journal of Mathematical Physics 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, Lecture Notes in Math, Springer, Berlin, 1981.

[10] 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.

[11] T. Misawa, Conserved Quantities and Symmetries Related to Stochastic Dynamical Systems, Annals of the Institute of Statistical Mathematics 51 (4) (1999) 779–802.

[12] L. Wu, Large and Moderate Deviations and Exponential Convergence for Stochastic Damping Hamiltonian Systems, Stochastic Processes and their Applications 91 (2) (2001) 205–238.

[13] W. Zhu, Z. Huang, Stochastic Stabilization of Quasi-partially Integrable Hamiltonian Systems by Using Lyapunov Exponent, Nonlinear Dynamics 33 (2) (2003) 209–224.
[14] G. N. Milstein, Y. M. Repin, M. V. Tretyakov, Numerical Methods for Stochastic Systems Preserving Symplectic Structure, SIAM Journal on Numerical Analysis 40 (4) (2002) 1583–1604.

[15] G. N. Milstein, Y. M. Repin, M. V. Tretyakov, Symplectic Integration of Hamiltonian Systems with Additive Noise, SIAM Journal on Numerical Analysis 39 (6) (2002) 2066–2088.

[16] L. Wang, J. Hong, R. Scherer, F. Bai, Dynamics and Variational Integrators of Stochastic Hamiltonian Systems., International Journal of Numerical Analysis & Modeling 6 (4).

[17] M. Pavon, Hamilton’s Principle in Stochastic Mechanics, Journal of Mathematical Physics 36 (12) (1995) 6774–6800.

[18] X.-M. Li, An Averaging Principle for a Completely Integrable Stochastic Hamiltonian System, Nonlinearity 21 (4) (2008) 803.

[19] J. Cresson, S. Darses, Stochastic Embedding of Dynamical Systems, Journal of Mathematical Physics 48 (7) (2007) 072703.

[20] D. Givon, R. Kupferman, A. Stuart, Extracting Macroscopic Dynamics: Model Problems and Algorithms, Nonlinearity 17 (6) (2004) R55.

[21] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge University Press, 2009.

[22] K.-i. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.

[23] 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, Journal of Mathematical Analysis and Applications 371 (1) (2010) 309–322.

[24] S. Albeverio, B. Rüdiger, J.-L. Wu, Invariant Measures and Symmetry Property of Lévy Type Operators, Potential Analysis 13 (2) (2000) 147–168.
[25] M. Högele, P. Ruffino, Averaging Along Foliated Lévy Diffusions, Nonlinear Analysis: Theory, Methods & Applications 112 (2015) 1–14.

[26] M. A. Högele, P. H. da Costa, Strong Averaging Along Foliated Lévy Diffusions with Heavy Tails on Compact Leaves, Potential Analysis 47 (3) (2017) 277–311.

[27] Y. Xu, J. Duan, W. Xu, An Averaging Principle for Stochastic Dynamical Systems with Lévy Noise, Physica D: Nonlinear Phenomena 240 (17) (2011) 1395–1401.

[28] J. Bao, G. Yin, C. Yuan, et al., Two-time-scale Stochastic Partial Differential Equations Driven by α-stable Noises: Averaging Principles, Bernoulli 23 (1) (2017) 645–669.

[29] H. Kunita, Stochastic Differential Equations Based on Lévy Processes and Stochastic Flows of Diffeomorphisms, in: Real and Stochastic Analysis, Springer, 2004, pp. 305–373.

[30] S. I. Marcus, Modeling and Approximation of Stochastic Differential Equations Driven by Semimartingales, Stochastics: An International Journal of Probability and Stochastic Processes 4 (3) (1981) 223–245.

[31] K. D. Elworthy, Stochastic Differential Equations on Manifolds, Vol. 70, Cambridge University Press, 1982.

[32] F. Xi, C. Zhu, Jump Type Stochastic Differential Equations with Non-Lipschitz Coefficients: Non-confluence, Feller and Strong Feller Properties, and Exponential Ergodicity, Journal of Differential Equations.

[33] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration, Springer Ser. Comput. Math 31.

[34] A. M. Fox, R. de la Llave, Barriers to Transport and Mixing in Volume-preserving Maps with Nonzero Flux, Physica D: Nonlinear Phenomena 295 (2015) 1–10.

[35] T. G. Kurtz, E. Pardoux, P. Protter, Stratonovich Stochastic Differential Equations Driven by General Semimartingales, Purdue University, 1991.
[36] V. I. Arnold, V. V. Kozlov, A. I. Neishtadt, Mathematical Aspects of Classical and Celestial Mechanics, Vol. 3, Springer Science & Business Media, 2007.

[37] A. M. Kulik, Exponential Ergodicity of the Solutions to SDEs with a Jump Noise, Stochastic Processes and their Applications 119 (2) (2009) 602 – 632.