IOP PUBLISHING NONLINEARITY

Nonlinearity **21** (2008) 803–822 doi:10.1088/0951-7715/21/4/008

An averaging principle for a completely integrable stochastic Hamiltonian system

Xue-Mei Li

Department of Mathematics, University of Warwick, Coventry CV4 7AL, UK

Received 20 June 2007, in final form 25 February 2008
Published 17 March 2008
Online at stacks.iop.org/Non/21/803

Recommended by L Ryzhik

Abstract
We investigate the effective behaviour of a small transversal perturbation of order ϵ to a completely integrable stochastic Hamiltonian system, by which we mean a stochastic differential equation whose diffusion vector fields are formed from a completely integrable family of Hamiltonian functions H_i, $i = 1, \ldots, n$. An averaging principle is shown to hold and the action component of the solution converges, as $\epsilon \to 0$, to the solution of a deterministic system of differential equations when the time is rescaled at $1/\epsilon$. An estimate for the rate of the convergence is given. In the case when the perturbation is a Hamiltonian vector field, the limiting deterministic system is constant in which case we show that the action component of the solution scaled at $1/\epsilon^2$ converges to that of a limiting stochastic differentiable equation.

Mathematics Subject Classification: 60H10, 58J65, 58J37

1. Introduction

The Model. A smooth $2n$-dimensional manifold M is said to be a symplectic manifold if it is equipped with a symplectic structure, that is, a closed differential two-form ω which is non-degenerate in the sense that for each $x \in M$, $\omega(v, w) = 0$ for all $w \in T_x M$ implies $v = 0$. Equivalently M admits a set of coordinates mapping such that the coordinate changing maps are symplectic on $\boldsymbol{R}^{2n}$ with the standard symplectic form $\omega_0 = \sum \mathrm{d}p_i \wedge \mathrm{d}q_i$.

A family of n smooth Hamiltonians $\{H_k\}$ on a $2n$-dimensional symplectic manifold is said to form a (completely) integrable system if they are pointwise Poisson commuting and if the corresponding Hamiltonian vector fields X_{H_k} are linearly independent at almost all points. Given such an integrable family and a C^1 locally Hamiltonian vector field V commuting with the family of vector fields X_{H_k} in the sense of $\omega(X_{H_k}, V) = 0$, consider the following model, which we call a *completely integrable stochastic symplectic/Hamiltonian system*:

$$\mathrm{d}x_t = \sum_{k=1}^{n} X_{H_k}(x_t) \circ \mathrm{d}B_t^k + V(x_t)\,\mathrm{d}t. \tag{1.1}$$

 © 2008 IOP Publishing Ltd and London Mathematical Society Printed in the UK

Here $(B_t^k, k = 1, \ldots, n)$ are pairwise independent Brownian motions on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with the standard assumptions on the filtration and $\circ$ stands for Stratonovitch integration. We have suppressed the chance element ω here as is convention. Note that the customary symbol for the symplectic form is unfortunately the same as that for the chance variable, however confusion should not arise as the chance variable will from now on not be explicitly expressed unless indicated otherwise. We call, respectively, X_{H_k} the diffusion vector fields and, V, the drift vector field for the stochastic differential equation.

In the integrable stochastic Hamiltonian system case, the diffusion vector fields span a sub-bundle of the tangent bundle, at least locally. The purpose of this paper is to investigate the effect of a small perturbation to random systems of this type. A solution to an integrable Hamiltonian system preserves the energies H_k, just as does a solution to any other stochastic Hamiltonian system and there are corresponding invariant manifolds (level sets). The Markov solution restricts to each compact level set and the restriction has generator

$$\mathcal{L}_0 = \sum_{k=1}^{n} \frac{1}{2} \boldsymbol{L}_{X_{H_k}} \boldsymbol{L}_{X_{H_k}} + \boldsymbol{L}_V.$$

Here $\boldsymbol{L}_V$ indicates Lie differentiation in the direction of V. If the integrable stochastic Hamiltonian system is perturbed by a vector field ϵK for $\epsilon > 0$ and K a C^1 vector field not necessarily taking values in the span of $\{X_{H_k}, k = 1, 2, \ldots, n\}$ the solution to the resulting equation

$$\begin{aligned} \mathrm{d}y_t^\epsilon &= \sum_{i=1}^{n} X_{H_k}(y_t^\epsilon) \circ \mathrm{d}B_t^k + V(y_t^\epsilon)\,\mathrm{d}t + \epsilon K(y_t^\epsilon)\,\mathrm{d}t, \\ y_t^\epsilon &= y_0 \end{aligned} \tag{1.2}$$

will not conserve the energies. On the other hand letting $\epsilon \to 0$, the deviation from level sets of the energies will be small. Consider the solution $y^\epsilon(t/\epsilon)$ scaled in time by $1/\epsilon$, which has a generator given by $\frac{1}{\epsilon}\mathcal{L}_0 + \boldsymbol{L}_K$. Note that the motion splits into two parts with the fast component an elliptic diffusion on the invariant torus and the slow motion governed by the transversal part of the vector field K. The evolution of $y^\epsilon(t/\epsilon)$ is the skew product of the diffusion of order 1 across the level sets and the fast elliptic diffusion of order ϵ^{-1} along the level sets. The motion on the level sets (thinking of the level sets as the standard n-torus), which would be quasi-periodic if there were no diffusion terms, is ergodic. The evolution of the action component of $y^\epsilon(t/\epsilon)$ will not depend on the angular variable in the limit as $\epsilon \to 0$ and is shown to be described by a system of n ordinary differential equations whose right-hand sides can be deduced from $\omega(K, X_{H_i})$, $i = 1, \ldots, n$. Here ω is the symplectic 2-form. The convergence rate is shown to be of order $\epsilon^{\frac{1}{3}}$.

Furthermore if the vector field K is given by a Hamiltonian function, the average of $\omega(K, X_{H_i})$ over the torus vanishes and we look at the second order scaling to see an interesting limit. The action component of $y^\epsilon(t\epsilon^{-2})$ involves a martingale term in the limit and the asymptotic law of $y^\epsilon(t\epsilon^{-2})$ across the level sets is shown to be given by a stochastic differential equation. It remains open to find an estimate for the rate of the convergence of the law of $y^\epsilon(t\epsilon^{-2})$ to the law of the limiting diffusion.

Main results. Suppose that $\omega(V, X_{H_i}) = 0$ and V commutes with each vector field X_{H_i}. Let $y_0 \in M$ be a regular point of H with a neighbourhood U_0 the domain of an action-angle coordinate map. Let T^ϵ be the first time that the solution $y_{t/\epsilon}$ starting from y_0 exits U_0. Set $H^\epsilon(t) = (H_1(y_{t/\epsilon}^\epsilon), \ldots, H_n(y_{t/\epsilon}^\epsilon))$. Then H^ϵ converges to the solution of the following system

of deterministic equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{H}_i(t) = \int_{M_{\bar{H}(t)}} \omega(X_{H_i}, K)(\bar{H}(t), z)\,\mathrm{d}\mu_{\bar{H}_t}(z),$$

with a corresponding initial condition: if T^0 is the first time that $\bar{H}(t)$ exits from U_0 then for all $t < T_0$, $\beta > 1$, there exists a constant $C_2 > 0$ such that

$$\left(\boldsymbol{E}\left(\sup_{s\leqslant t} \|H^\epsilon(s\wedge T^\epsilon) - \bar{H}(s\wedge T^\epsilon)\|^\beta\right)\right)^{\frac{1}{\beta}} \leqslant C_2\epsilon^{1/3}.$$

Furthermore if $r > 0$ is such that $U \equiv \{x : \|H(x) - H(y_0)\| \leqslant r\} \subset U_0$ define

$$T_\delta = \inf_t\{|\bar{H}_t - H(y_0)| \geqslant r - \delta\}.$$

Then for any $\beta > 1$, $\delta > 0$ and a constant C depending on T_δ,

$$P(T^\epsilon < T_\delta) \leqslant C(T_\delta)\delta^{-\beta}\epsilon^{\beta/3}.$$

In the case of K being a smooth local Hamiltonian vector field $\int_{M_{\bar{H}(t)}} \omega(X_{H_i}, K)(\bar{H}(t), z)\times \mathrm{d}\mu_{\bar{H}_t}(z) = 0$ and hence we could look at the second scaling. The law of the stochastic process $H(y^\epsilon_{t/\epsilon^2})$ stopped at S^ϵ, the first time that the process $y^\epsilon_{t/\epsilon^2}$ exits from U_0, converges to that of $H(z_{t\wedge S^\epsilon})$ where z_t is the solution to the following stochastic differential equation:

$$\mathrm{d}z^j_t = \sum_i \sigma^j_i(z_t)\circ \mathrm{d}B^i_t + b_j(z_t)\,\mathrm{d}t.$$

Here (σ^j_i) is the square root of the matrices (a_{ij}),

$$a_{ij}(a) = -\int_{M_a} \omega(K, X_{H_j})\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))(a, z)\,\mathrm{d}\mu_a(z)$$

and

$$b_j(a) = \frac{1}{2}\int_{M_a} \boldsymbol{L}_K\mathcal{L}_0^{-1}(\omega(X_{H_j}, K))(a, z)\,\mathrm{d}\mu_a(z).$$

We give here a somewhat trivial example of a stochastic integrable system of equations on $\boldsymbol{R}^4$ with the standard symplectic structure:

$$\begin{aligned}
\mathrm{d}x_1(t) &= x_3(t)\,\mathrm{d}B_t,\\
\mathrm{d}x_2(t) &= x_4(t)\,\mathrm{d}B_t + x_4\,\mathrm{d}W_t,\\
\mathrm{d}x_3(t) &= -x_1(t)\,\mathrm{d}B_t,\\
\mathrm{d}x_4(t) &= -x_2(t)\,\mathrm{d}B_t - x_2(t)\,\mathrm{d}W_t,
\end{aligned}$$

where B_t and W_t are independent 1-dimensional Brownian motions. Take the perturbation vector to be

$$K_1 = \left(0, \frac{x_2}{x_2^2 + x_4^2}, 0, 0\right),$$

$$K_2 = \left(\frac{x_3}{(x_1^2 + x_3^2)^2}, 0, \frac{x_1}{(x_1^2 + x_3^2)^2}, 0\right)$$

or

$$K_3 = \left(\frac{x_3}{(x_1^2 + x_3^2)^{\frac{3}{2}}}, 0, -\frac{x_1}{(x_1^2 + x_3^2)^{\frac{3}{2}}}, 0\right).$$

In the case of K_1 we have a non-trivial average; in the case of K_2 and K_3 we have trivial averages at first scaling and can proceed to the second scaling and obtain a stochastic differential equation (SDE) in the limit.

Remark on the model. This work is based on the framework of Arnold on an averaging principle for an integrable Hamiltonian system, as a stochastic Hamiltonian system can be considered as a family of ordinary differential equations with time dependent random vector fields (whose corresponding Hamiltonians are in general neither bounded from below nor differentiable in time). Averaging of stochastic systems was pioneered by Khasminskii [16], Papanicolaou *et al* [22], and has been an active research field on which there is a rich literature. The structure of the main averaging results is close to that described in the excellent survey of Papanicolaou [21]. We refer to Abraham–Marsden [3], Arnold [4], Hofer–Zehnder [15] and McDuff–Salamon [20] as references for Hamiltonian systems on symplectic manifolds, to Givon–Kupferman–Stuart [13] for some physical models behind these problems and for recent progress in the direction of deterministic averaging, to Freidlin–Wentzell [9] and Sowers–Namachchivaya [24] for random perturbations to systems with one degree of freedom, and to Eizenberg–Freidlin [8], Borodin–Freidlin [6], Freidlin–Wentzell [10], Sowers [23], Koralov [18], Khasminskii–Krylov [17], and Khasminskii–Yin [19] for recent related work on perturbations of stochastic systems as well as Arnold–Imkeller–Namachchivaya [1] for a discussion on asymptotic expansion of a damped oscillator of one degree of freedom with small noise perturbation. For the Lagrangian mechanics and variational principle in a stochastic framework we would like to refer to Bismut's work [5]. However in this paper we do not investigate the stochastic mechanics related to the SDEs.

The main novelty of this work is the model itself. We also obtained a rate of convergence. The generator of our perturbed system is

$$\mathcal{L}_0^\epsilon = \sum_{k=1}^{n} \frac{1}{2} \boldsymbol{L}_{X_{H_k}} \boldsymbol{L}_{X_{H_k}} + \boldsymbol{L}_V + \epsilon L_K,$$

from which we observe the following aspects of the model: (a) The unperturbed random dynamical system is a completely integrable system. (b) The fast component of the system is the diffusion term, not the deterministic term. It is also worth noting that the limiting slow motion scaled at $1/\epsilon$ has n degrees of freedom and is described by a system of n ordinary differential equations. It is only at the second scaling, in the case of the perturbation being Hamiltonian, that we see a limiting n-dimensional non-trivial Markov process. The convergence of the slow motion $(H_1(y^\epsilon_{t/\epsilon}), \ldots, H_n(y^\epsilon_{t/\epsilon}))$ is in L^p with rate $\epsilon^{\frac{1}{3}}$. At the second scaling the slow motion converges weakly. The assumptions we make on the Hamiltonian functions are: the $\boldsymbol{R}^n$ valued function $(H_1, \ldots, H_n)$ is proper and its set of critical points has measure zero.

The following work relates particularly well to ours. We will point out the differences and similarities. In Dolgopyat [7] the following is proved: if $\hat{h}$ is a first integral of $\dot{y} = \boldsymbol{E}(F(y, \xi_1))$, where ξ_1 is a random variable of compact support, the piece-wise linear function h_ϵ given by $\hat{h}_\epsilon(\epsilon^2 n) = \hat{h}(x_n^\epsilon)$ converges weakly to the solution of a SDE, under suitable conditions. This was proved using the martingale method. Here x_n^ϵ is solution to

$$x_{n+1} = x_n + \epsilon F(x_n, \xi_n) + \epsilon^2 G(x_n, \xi_n) + \epsilon^3 H(x_n, \xi_n, \epsilon),$$

and the ξ_n's are i.i.d. random variables. The equation governing x_n are very general. If the perturbation K in (1.2) is given by a Hamiltonian function we may take $F \equiv 0$ in Dolgopyat's model and take $\hat{h} = h$, a first integral to (1.1); however it is not clear how the piece-wise linear function $\hat{h}_\epsilon$ relates to the fast motion $h(x_{t/\epsilon^2})$. In Eizenberg–Freidlin [8] and Borodin–Freidlin [6], the diffusion part of the motion belongs to the slow component and the fast motion is deterministic. More precisely the perturbed generator is $L_V + \epsilon L_1 + L_2$ for a vector field V and an interaction term L_2. Freidlin and Weber [11, 12] have a different objective. They are

mainly concerned with one conserved quantity H, not a completely integrable system. The objective there is to obtain a limiting Markov process, using weak convergence, on the graph homeomorphic to the set of connected components of the level sets of H.

2. Preliminaries

2.1. Hamiltonian and symplectic vector fields

Every symplectic manifold has a natural measure, called the Liouville measure. It is in fact $\wedge^n\omega$, differing from the volume form by a constant. Denote by $\iota_v\omega$ the inner product of a tangent vector v with ω. The map from TM to T^*M given by $v \mapsto \iota_v\omega$ is a vector bundle isomorphism, and there is a one to one correspondence between vector fields and differential 1-forms. A *symplectic vector field* V, also called a *locally Hamiltonian vector field*, is one which preserves the symplectic structure, i.e. $L_V\omega = 0$. Here L_V denotes Lie differentiation in the direction of V. Equivalently $\iota_V\omega$ is a closed differential 1-form. For every C^1 function $H : M \to \mathbb{R}$ we can associate a *Hamiltonian vector field* (also called *symplectic gradient vector field*) given by

$$\iota_{X_H}\omega = \mathrm{d}H.$$

The canonical symplectic structure on $\boldsymbol{R}^{2n}$ with coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ is $\omega = \sum_{i=1}^n \mathrm{d}q_i \wedge \mathrm{d}p_i$. Darboux's theorem asserts that any symplectic manifold is locally $\boldsymbol{R}^{2n}$ with its canonical symplectic structure. If the first de Rham cohomology $\mathbb{H}^1(M; \boldsymbol{R})$ vanishes, as in the case of $\boldsymbol{R}^{2n}$, every locally Hamiltonian vector field is given by a Hamiltonian function. There are locally Hamiltonian vector fields which are not given by a Hamiltonian function. Take the two torus T^2 with coordinates x and y periodic in x and y. The canonical symplectic structure on $\boldsymbol{R}^2$ induces the symplectic structure on T^2: $\omega = \mathrm{d}x \wedge \mathrm{d}y$. A vector field $X(x, y) = a(x, y)(\partial/\partial x) + b(x, y)(\partial/\partial y)$ with $(\partial a/\partial x) + (\partial b/\partial y) \equiv 0$ is clearly locally Hamiltonian: the 1-form $\iota_X\omega = a(x, y)\,\mathrm{d}y - b(x, y)\,\mathrm{d}x$ is closed. If $a = 1, b = 0$, the vector field is not given by a Hamiltonian function on T^2.

The space of smooth functions on M has a Lie algebra structure given by the Poisson bracket. The Poisson bracket of two smooth functions is denoted by $\{F_1, F_2\}$ and $\{F_1, F_2\} = \mathrm{d}F_1(X_{F_2}) = \omega(X_{F_1}, X_{F_2})$. The vector field corresponding to the Poisson bracket is precisely the Lie bracket of the Hamiltonian vector fields X_{F_1} and X_{F_2}. Two Hamiltonian functions are *Poisson commuting* or *in involution* if their Poisson bracket vanishes, in which case their corresponding Hamiltonian flows commute. If $\{F, H\} = 0$ we say that F is a *first integral* of H. Two Hamiltonian functions are said to be linearly independent at x if their associated Hamiltonian vector fields are linearly independent at that point. A family of n Hamiltonian functions is said to form an *integrable system* if the Hamiltonian functions are pairwise Poisson commuting and if they are linearly independent on a set of full measure.

2.2. An example of a stochastic Hamiltonian system on $\boldsymbol{R}^{2n}$

The Hamiltonian vector field given by a Hamiltonian function H is given by $X_H = J\,\mathrm{d}H$, where J is the canonical complex structure:

$$J = \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix},$$

where **1** denotes the n by n identity matrix. The corresponding Hamiltonian system thus takes the familiar form

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad 1 \leqslant i \leqslant n,$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \qquad 1 \leqslant i \leqslant n.$$

For simplicity write $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$. An important class of examples of Hamiltonian functions on $\boldsymbol{R}^{2n}$ is those of the form $H(p, q) = \frac{1}{2}|p|^2 + V(q)$ for some potential function V. If V is quadratic, e.g. $V(q) = \frac{1}{2}a^2|q|^2$, we have the standard harmonic oscillator. The Poisson bracket in $\boldsymbol{R}^{2n}$ is of the following form:

$$\{H, F\} = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i} \frac{\partial F}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial F}{\partial p_i} \right).$$

A example of an integrable stochastic Hamiltonian system is given by

$$\mathrm{d}q_i(t) = \frac{\partial K}{\partial p_i}\mathrm{d}t + \sum_{k=1}^{n} \frac{\partial H_k}{\partial p_i} \circ \mathrm{d}B_t^k,$$

$$\mathrm{d}p_i(t) = -\frac{\partial K}{\partial q_i}\mathrm{d}t + \sum_{k=1}^{n} \frac{\partial H_k}{\partial q_i} \circ \mathrm{d}B_t^k,$$

where

$$H_1 = \frac{1}{2}\sum_{i=1}^{n} a_i^2 q_i^2 + \frac{1}{2}\sum_{i=1}^{n} p_i^2$$

$$H_k = \frac{1}{2}a_k q_k^2 + \frac{1}{2}\frac{p_k^2}{a_k}, \qquad 2 \leqslant k < n,$$

and K is a smooth function which commutes with all H_i, e.g. if K is a smooth function of H_i.

2.3. The invariant manifolds and integrable symplectic Hamiltonian systems

Let $\{H_k\}_{k=1}^n$ be an integrable family of smooth Hamiltonian functions, i.e. they are Poisson commuting and so the H_k's are first integrals of each other and they are independent on a set of full measure. For $a = (a_1, \ldots, a_n) \in \boldsymbol{R}^n$ denote by M_a the level set of the first integrals $\{H_k\}$:

$$M_a = \cap_{i=1}^n \{x : H_i(x) = a_i\}.$$

The Liouville–Arnold theorem states that if $\{H_k\}_{k=1}^n$ are independent on M_a then M_a is a smooth manifold and furthermore it is diffeomorphic to an n-dimensional torus if it is compact and connected. For such value a, M_a is invariant under the flows of each H_k and each x in M determines an invariant manifold through the value $a = (H_1(x), \ldots, H_n(x))$, which we write also as $M_{H(x)}$.

An application of Itô's formula below shows that the solution flow $\{F_t(x) : t \geqslant 0\}$ of (1.1) preserves the invariant manifolds $\{M_a\}$:

$$\mathrm{d}H_i(x_t) = \sum_k \mathrm{d}H_i(X_{H_k}(x_t)) \circ \mathrm{d}B_t^k + \mathrm{d}H_i(V(x_t))\,\mathrm{d}t = 0, \qquad 1 \leqslant i \leqslant n.$$

For simplicity we assume throughout the paper the following:

- The invariant manifolds are compact, which is the case if the map $x \in M \mapsto (H_1(x), \ldots, H_n(x)) \in \boldsymbol{R}^n$ is proper. Note that the n vector fields $\{X_{H_i(x)}\}$ are tangent to $M_{H(x)}$ and the symplectic form ω vanishes on the invariant manifolds M_a. Therefore the stochastic differential equation (1.1) is elliptic when restricted to individual invariant manifolds and the Markovian solution is ergodic. Denote by μ_a the unique invariant probability measure on M_a; it can be considered as the uniform probability measure on the torus.

2.4. The invariant measure and the divergence operator for semi-elliptic stochastic symplectic systems

Let $\{A^0, A^1, \ldots, A^n\}$ be smooth symplectic vector fields with $[A^i, A^j] = 0$ for all i, j. Assume that $\{A^1, \ldots, A^n\}$ spans a sub-bundle E of the tangent bundle TM of rank n. Consider the following stochastic differential equation:

$$\mathrm{d}x_t = \sum_{i=1}^{n} A^i(x_t) \circ \mathrm{d}B_t^i + A^0(x_t)\,\mathrm{d}t. \tag{2.1}$$

If there is a global solution flow $\{F_t(x_0, \omega) : t \geqslant 0\}$ to equation (2.1), then the solution flows are stochastic symplectomorphisms, i.e. $\omega = F_t^*\omega$, where ω is the symplectic form, not the chance variable.

For each $x \in M$, define a linear map $A(x) : \boldsymbol{R}^{2n} \to T_xM$ by

$$A(x)(e) = \sum_{i=1}^{n} A^i(x)\langle e, e^i\rangle, \qquad e \in \boldsymbol{R}^{2n}$$

where $\{e^i\}$ is an orthonormal basis of $\boldsymbol{R}^{2n}$. The linear map is onto E_x with kernel $\{0\} \times \boldsymbol{R}^n$ and gives a positive symmetric bilinear form on E by making $\{A^i(x)\}$ an orthonormal basis:

$$\langle A^i(x), A^j(x)\rangle = \delta_{ij}.$$

Then $A(x)$ is an isomorphism from $\boldsymbol{R}^n \times \{0\}$ to E_x. This defines a metric on E: for $u = \sum_i u_i A^i$ and $v = \sum_i v_i A^i$,

$$\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$$

and for a function f we define its gradient $\nabla_E f = \sum_i \mathrm{d}f(A^i)A^i$.

The symplectic structure ω restricts to E defines a complex structure on M as follows: we first give the tangent bundle TM any Riemannian metric which agrees with the one constructed on E using the linear map A. Define $J_x : T_xM \to T_xM$ by

$$\omega(J_x u, v) = \langle u, v\rangle_x.$$

To see that this identity defines J_x uniquely suppose that for $u \in T_xM$ there are u_1 and u_2 satisfies $\omega(u_i, v) = \langle u_i, v\rangle_x$, $i = 1, 2$. Then $\omega(u_1 - u_2, v) = 0$ for all v. Thus $u_1 = u_2$. Existence can be easily seen as direct calculations can be done in $\boldsymbol{R}^{2n}$.

Next take $A^i = X_{H_i}$ in (2.1), $i = 1, 2, \ldots, n$, to be the Hamiltonian vector fields for an integrable family of Hamiltonian functions $\{H_i\}$ and $A^0 = V$. We arrive back to the integrable stochastic symplectic equation (1.1) where V is a symplectic vector field commuting with all

X_{H_i}. Under our assumption that

$$H : x \to (H_1(x), \ldots, H_n(x))$$

is a proper map, then for almost every point a_0 in $\boldsymbol{R}^n$ it is either trivial or a local trivial fibration in the sense that there is a neighbourhood V of a_0 such that $H^{-1}(a)$ is a smooth sub-manifold for all $a \in V$ and that there is a diffeomorphism from $H^{-1}(V)$ to $V \times H^{-1}(a_0)$. Such a_0 is called a regular value of H. Denote by Σ_H the set of values in $\boldsymbol{R}^n$ which are not regular. A point y in M is said to be a critical point if $H(y) \in \Sigma_H$. By Sard's theorem the set of critical values of the function H has measure zero. The $2n$-differential form ω^n, as a measure, has a decomposition which gives a measure on each invariant manifold M_a for regular a value. The decomposition can be chosen in the following way. First recall that on a neighbourhood of a regular point a_0 of H, every component of the level set M_{a_0} is diffeomorphic to an n-torus and a small neighbourhood U_0 of M_{a_0} is diffeomorphic to the product space $T^n \times D$, where D is a relatively compact open set in $\boldsymbol{R}^n$, see, e.g. [2]. More precisely if V is an open set of $\boldsymbol{R}^n$ such that $H^{-1}(V)$ does not contain any critical points of H then it is diffeomorphic to $D \times T^n$. Take an action-angle chart around M_a which is diffeomorphic to $D \times T^n$ for some open set D. The measure $(\sum_i \mathrm{d}I^i \wedge \mathrm{d}\theta^i)^n$ on the product space naturally splits to give us a probability measure, the Haar measure $\mathrm{d}\theta^1 \wedge \cdots \mathrm{d}\theta^n$ on T^n. We take the corresponding one on M_a and denote it by μ_a. Let U be a section of E. Define the divergences $\mathrm{div}_E^a U$ to be the functions such that

$$\int_{M_a} \mathrm{d}f(U)\,\mathrm{d}\mu_a = -\int_{M_a} f\,\mathrm{div}_E^a U\,\mathrm{d}\mu_a$$

for all smooth functions f on M_a. Note that $\mathrm{div}_E X_{H_i} = 0$, since

$$\int_{M_a} \mathrm{d}f(X_{H_i})\,\mathrm{d}\mu_a = \int_{M_a} \{H_i, f\}\,\mathrm{d}\mu_a = 0$$

for all smooth functions f (see the beginning of section 4 for a calculation). Thus if $U = \sum_i a_i X_{H_i}$, where a_i are constant on M_a and is thus divergence free.

Remark 2.1. Let U be a section of E commuting with all X_{H_i} the invariant measure for the SDE (1.1) restricted to the invariant manifold M_a is μ_a, which varies smoothly with a in sufficiently small neighbourhoods of a regular value.

Proof. The measure ω^n is an invariant measure for the SDE on M due to the fact that the solution of the SDE leaves invariant the symplectic form. More precisely, since U commutes with $\{X_{H_i}\}$ and thus can be written in the the form of $U = \sum_i a_i X_{H_i}(x)$, where a_i are constant on M_a, it is therefore divergence free. Thus the invariant measure of the SDE restricted to the torus is the same as that of the corresponding SDE without a drift. From the action-angle transformation we see that the measure μ_a is an invariant measure for the SDE restricted to M_a. This is in fact the only invariant measure for the SDE on M_a since the system is elliptic when restricted to each level set and the conclusion follows. □

3. An averaging principle

Let $\{H_i\}_{i=1}^n$ be a completely integrable system on a smooth $2n$-dimensional symplectic manifold M so that the functions $\{H_i\}$ are constants of motions of each other and that they are pairwise in involution. We assume that the $\boldsymbol{R}^n$-valued function $H = (H_1, \ldots, H_n)$ is

proper and its set of critical points has measure zero. Note that the vector fields $\{X_{H_i}\}$ form an integrable distribution and through each point of the manifold there is an integrable n-dimensional manifold.

Take an action-angle coordinate: $\phi^{-1} : U_0 \to D \times T^n$. In this coordinate, $x = \phi(I, \theta)$, $I \in D, \theta \in T^n$, and $(\phi^{-1})_*\omega = \mathrm{d}I \wedge \mathrm{d}\theta$ defines a symplectic structure on $D \times T^n$. Furthermore if $\tilde{H}_i = H_i(\phi(I, \theta))$ is the induced Hamiltonian on $D \times T^n$ then $\dot{I}_i^k = -(\partial \tilde{H}_i/\partial\theta_k) = 0$ and

$$\dot{\theta}_i^k = \frac{\partial \tilde{H}_i}{\partial I_k} = \omega_i^k(I) \tag{3.1}$$

with ω_i^k smooth functions. In fact $X_{\tilde{H}_i} = (\phi^{-1})_*(X_{H_i}) = -\sum_{k=1}^n (\partial(H_i \circ \phi)/\partial I_k)(\partial/\partial\theta_k)$. For example the integrable Hamiltonian system in section 2.2 is equivalent to the Hamiltonian system $H_1 = \sum_{i=1}^n a_i\bar{q}_i$, $a_i > 0$, and $H_k = \bar{q}_k, k = 2, \ldots, n$, through the action-angle coordinates change $(q, p) \mapsto (\bar{q}, \bar{p})$:

$$(q_1, \ldots, q_n,\ p_1, \ldots, p_n) = \left(\sqrt{\frac{2\bar{q}_1}{a_1}}\cos\bar{p}_1, \ldots, \sqrt{\frac{2\bar{q}_n}{a_n}}\cos\bar{p}_n, \sqrt{2a_1\bar{q}_1}\sin\bar{p}_1, \ldots, \sqrt{2a_n\bar{q}_n}\sin\bar{p}_n,\right).$$

The corresponding Hamiltonian system is the trivial one $\dot{\bar{p}}_i = a_i, \dot{\bar{q}}_i = 0$. Since U_0 is diffeomorphic to $D \times T^n$ there is a constant $r > 0$ such that U_0 contains the open set $\{x : \sum_i |H_i(x) - H_i(y_0)|^2 \leqslant r^2\}$.

Let K be a smooth vector field, transversal in the sense that $\omega(X_{H_i}, K)$ are not all identically zero. Denote by y_t^ϵ the solution to (1.2), the perturbation of the integrable system (1.1) starting from a given point y_0 in M. Set $x_t = y_t^0$, the solution to (1.1) with initial value y_0. If V is a vector field on M denote by $\tilde{V}$ the induced vector field on $D \times T^n$. We assume the following of the SDE (1.2):

Condition R. Suppose that $\omega(V, X_{H_i}) = 0$ and V commutes with all vector fields X_{H_i}. Let $y_0 \in M$ be a regular point of H with a neighbourhood U_0 the domain of an action-angle coordinate map: $\phi^{-1} : U_0 \to D \times T^n$, where D is an open set of $\boldsymbol{R}^n$.

We adopt the following notation: if f is a function on U_0, by $\tilde{f}$ we mean the representation of f in $D \times T^n$.

Lemma 3.1. *Assume condition R holds for (1.2). Let τ^ϵ be the first time that the solution y_t^ϵ starting from y_0 exits U_0. Then for any smooth function f on M,*

(1) $$\left[\boldsymbol{E}\left(\sup_{s \leqslant t\wedge\tau^\epsilon} |f(y_s^\epsilon) - f(x_s)|^p\right)\right]^{\frac{1}{p}} \leqslant C_1\epsilon(t + t^2),$$

where $C_1 = C_1(V, K, H_i, f)$ depends on the upper bounds of the functions $|\mathrm{d}\tilde{f}|$, $|\partial^2\tilde{H}_k/\partial I_i\partial I_j|$, $|\mathrm{d}\tilde{V}|$, $|\tilde{K}|$ on $D \times T^n$.

(2) If $V \equiv 0$, then the estimates above, $\epsilon(t + t^2)$, can be improved to $C_1\epsilon(t + t^{\frac{3}{2}})$.

Proof. In the proof below C stands for an unspecified constant. We denote the flows in action-angle coordinates by $x_t = \phi(I_t, \theta_t)$ and $y_t^\epsilon = \phi(I_t^\epsilon, \theta_t^\epsilon)$. Set $\tilde{f} = f \circ \phi$. Then

$$\begin{aligned}|f(y_t^\epsilon) - f(x_t)| &= |\tilde{f}(I(y_t^\epsilon), \theta(y_t^\epsilon)) - \tilde{f}(I(x_t), \theta(x_t))| \\ &\leqslant C|I(y_t^\epsilon) - I(x_t)| + C|\theta(y_t^\epsilon) - \theta(x_t)|,\end{aligned}$$

using the fact that $\partial \tilde{f}/\partial I$ and $\partial \tilde{f}/\partial \theta$ are bounded on $T^n \times D$ as D is relatively compact. In the local chart, $(\partial \tilde{V}/\partial \theta_i) = 0$ and we can write $V(I,\theta) = V_j(I)(\partial H_j/\partial I)(I,\theta) = \omega_0^j(I)(\partial/\partial\theta_j)$ for some smooth functions ω_0^j on D. The perturbation vector field can be written as (K^θ, K^I), where $K_\theta = (K_\theta^1, \ldots, K_\theta^n)$ and $K_I = (K_I^1, \ldots, K_I^n)$ are, respectively, the angle and the action component of the vector field $\tilde{K}$ on $T^n \times D^n$. The result is now clear from the form of the SDE on $T^n \times D$:

$$\mathrm{d}I_t^{\epsilon,i} = \epsilon K_I^i(I_t^\epsilon, \theta_t^\epsilon)\,\mathrm{d}t,$$

$$\mathrm{d}\theta_t^{\epsilon,i} = \sum_{k=1}^n \omega_k^i(I_t^\epsilon) \circ \mathrm{d}B_t^k + \omega_0^i(I_t^\epsilon)\,\mathrm{d}t + \epsilon K_\theta^i(I_t^\epsilon, \theta_t^\epsilon)\,\mathrm{d}t,$$

where ω_k^i, $i, k = 1, \ldots, n$, are defined by (3.1). Indeed, then

$$\sup_{s \leqslant t\wedge\tau^\epsilon} |I_s^\epsilon - I_s| = \epsilon \sup_{s \leqslant t\wedge\tau^\epsilon} \Big| \int_0^s |K_I(I_s^\epsilon, \theta_s^\epsilon)|\,\mathrm{d}s \leqslant \epsilon t \sup_{D\times T^n} |K_I|,$$

and for $s < \tau^\epsilon$,

$$\theta_s^{\epsilon,i} - \theta_s^i = \sum_{k=1}^n \int_0^s (\omega_k^i(I_r^\epsilon) - \omega_i^k(I_r)) \circ \mathrm{d}B_r^k + \int_0^s (\omega_0^i(I_r^\epsilon) - \omega_0^i(I_r))\,\mathrm{d}r + \epsilon \int_0^s K_\theta^i(I_r^\epsilon, \theta_r^\epsilon)\,\mathrm{d}r.$$

As the Stratonovitch correction for the SDE vanishes, we may replace the Stratonovitch integration by Itô integration:

$$\int_0^s (\omega_k^i(I_r^\epsilon) - \omega_k^i(I_r)) \circ \mathrm{d}B_r^k = \int_0^s (\omega_k^i(I_r^\epsilon) - \omega_k^i(I_r))\,\mathrm{d}B_r^k.$$

Consequently,

$$\begin{aligned}
|\theta_i(y_s^\epsilon) - \theta_i(x_s)| &\leqslant \left| \sum_{k=1}^n \int_0^s (\omega_k^i(I_r^\epsilon) - \omega_i^k(I_r))\,\mathrm{d}B_r^k \right| \\
&\quad + \sup_{D\times T^n} |\mathrm{d}\omega_0^i| \cdot \int_0^s |I_r^\epsilon - I_r|\,\mathrm{d}r + \epsilon s \sup_{D\times T^n} |K_\theta^i| \\
&\leqslant \left| \sum_{k=1}^n \int_0^s (\omega_k^i(I_r^\epsilon) - \omega_i^k(I_r))\,\mathrm{d}B_r^k \right| \\
&\quad + \epsilon \frac{s^2}{2} \sup_{D\times T^n} |K_I| \cdot \sup_{D\times T^n} |\mathrm{d}\omega_i^0| + \epsilon s \sup_{D\times T^n} |K_\theta^i|.
\end{aligned}$$

Summing up over i, we have

$$\begin{aligned}
\boldsymbol{E} \sup_{s \leqslant t\wedge\tau^\epsilon} |\theta_s^\epsilon - \theta_s|^p &\leqslant C_1 \sup_{s\leqslant t} \left(\sum_{i,k=1}^n \boldsymbol{E} \int_0^s \left| \omega_k^i(I_r^\epsilon) - \omega_k^i(I_r) \right|^2 \right)^{p/2} + C_2(\tilde{K})\epsilon(t+t^2)^p \\
&\leqslant C_1 \left(\sum_{i,k} (|\mathrm{d}\omega_k^i| \vee 1)^p \right) \epsilon^p t^{\frac{3p}{2}} + C_2(\tilde{K})\epsilon^p (t+t^2)^p
\end{aligned}$$

by L_p inequalities for martingales. Combining the estimates we obtain

$$\boldsymbol{E} \sup_{s \leqslant t \wedge \tau^\epsilon} |f(y_s^\epsilon) - f(x_s)| \leqslant C_3 \epsilon (t + t^2)$$

for some constant C_3.

If the drift $V \equiv 0$ then $\omega_0 = 0$ and the calculation above shows that the estimate is of the order $\epsilon(t^{\frac{3}{2}} \wedge 1)$. □

If the stochastic dynamical system (1.1) is subjected to a small non-Hamiltonian perturbation, the slow variable is in the direction transversal to the energy surfaces while the stochastic components are the fast variables. The lemma shows that the first integrals of the perturbed system change by an order $\epsilon(t + t^2)$ over a time interval t and so the slow component accumulates over a time interval of the size t/ϵ and we obtain a new dynamical system in the limit: as ϵ goes to zero the motion along the torus is significantly faster compared with the motion in the transversal direction and thus the action component of $y_{t/\epsilon}^\epsilon$ has a limit as the randomness in the fast component is averaged out by the induced invariant measure, as shown below. Recall that $H(x) = (H_1(x), \ldots, H_n(x))$.

We first prove the following lemma.

Lemma 3.2. *Assume condition R holds. Let g be a real-valued function on M, which is considered in the action-angle coordinates as a function from $D \times T^n$ to $\boldsymbol{R}$. Define $Q^g : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$ to be its average over the torus:*

$$Q^g(a) = \int_{T^n} \tilde{g}(a, z)\, \mathrm{d}\mu(z). \tag{3.2}$$

Suppose that g is C^1 on U_0. Set

$$H_i^\epsilon(s) = H_i(y_{s/\epsilon}^\epsilon), \qquad H^\epsilon(s) = (H_1^\epsilon(s), \ldots, H_n^\epsilon(s)).$$

Then

$$\int_{s \wedge T^\epsilon}^{(s+t) \wedge T^\epsilon} g(y_{r/\epsilon}^\epsilon)\, \mathrm{d}r = \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} Q^g(H^\epsilon(r))\, \mathrm{d}r + \delta(g, \epsilon, t) \tag{3.3}$$

with the following rate of convergence: for any $\beta > 1$,

$$\left(\boldsymbol{E} \sup_{s \leqslant t} |\delta(g, \epsilon, s)|^\beta \right)^{\frac{1}{\beta}} \leqslant C(t) \epsilon^{\frac{1}{3}}, \tag{3.4}$$

where T^ϵ is the first time that $y_{t/\epsilon}^\epsilon$ exits from U_0 and $\tau^\epsilon = T^\epsilon/\epsilon$.

Proof. The idea is to approximate $g(y_r^\epsilon)$ by $g(y_r)$ on sufficiently small intervals and to apply the law of large numbers to each integral bearing in mind that the cost of replacing y_r^ϵ by y_r is of order $\epsilon\delta$ on an interval of size δ, assuming that $\delta \to \infty$ as $\epsilon \to 0$, and that the error term for replacing time average by space average is of order $1/\sqrt{\delta}$.

Let τ^ϵ be the first time that y_t^ϵ exits from U_0. For $q \in (0, 1)$, let $\Delta t = ((t + s)/\epsilon^q) \wedge \tau^\epsilon - (s/\epsilon^q) \wedge \tau^\epsilon$, which is of order ϵ^{-q}, and set $N \equiv N(\epsilon) = [\epsilon^{q-1}] + 1$, which is of order ϵ^{q-1}. Here $[\epsilon^{q-1}]$ is the integer part of ϵ^{q-1} and all terms may depend on the sample paths of ω. Take $t_n = (s/\epsilon) \wedge \tau^\epsilon + n\Delta t$, $1 \leqslant n \leqslant N - 1$, so that

$$\frac{s}{\epsilon} \wedge \tau^\epsilon = t_0 < t_1 < \cdots < t_{N-1} < \frac{s+t}{\epsilon} \wedge \tau^\epsilon.$$

We first make some pathwise estimates. For any C^1 function g on M,

$$\left|\int_{s\wedge T^\epsilon}^{(s+t)\wedge T^\epsilon} g(y^\epsilon_{u/\epsilon})\,\mathrm{d}u\right| = \left|\epsilon\int_{\frac{s}{\epsilon}\wedge\tau^\epsilon}^{\frac{s+t}{\epsilon}\wedge\tau^\epsilon} g(y^\epsilon_r)\,\mathrm{d}r\right|$$

$$\leqslant \epsilon\left|\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} g(y^\epsilon_r)\,\mathrm{d}r\right| + \epsilon\left|\int_{t_N}^{\frac{s+t}{\epsilon}\wedge\tau^\epsilon}\right| g(y^\epsilon_r)\,\mathrm{d}r.$$

Since g is bounded on U_0, the second term on the right-hand side of the above expression converges to zero with rate ϵ^{1-q}:

$$\epsilon\left|\int_{t_N}^{\frac{(s+t)}{\epsilon}\wedge\tau^\epsilon} g(y^\epsilon_r)\,\mathrm{d}r\right| \leqslant C\epsilon\Delta t \leqslant Ct\epsilon^{1-q}.$$

For the remaining terms we use the splitting

$$\epsilon\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} g(y^\epsilon_r)\,\mathrm{d}r = \epsilon\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} [g(y^\epsilon_r) - g(F_{r-t_n}(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))]\,\mathrm{d}r$$

$$+\epsilon\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))\,\mathrm{d}r,$$

where ω denotes the chance variable, Θ_t the shift operator on the canonical probability space: $\Theta_t(\omega)(-) = \omega(-+t) - \omega(t)$ and $\{F_t(x,\omega), t \geqslant 0\}$ the solution flow of the unperturbed stochastic differential equation (1.1) with starting point x. Write the summation as the sum of A_1 and A_2 and the second term is

$$A_2(t,\epsilon) \equiv \epsilon\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))\,\mathrm{d}r.$$

The law of large numbers says that for any function f on a compact manifold converging to infinity as t converges to infinity, $\frac{1}{t}\int_s^{s+t} f(x_r)\,\mathrm{d}r$ converges to $\int_M f(z)\,\mathrm{d}z$ when $t \to \infty$ with rate $1/\sqrt{t}$ and the convergence is uniform on compact time intervals in L^p for all $p > 1$. Here $\mathrm{d}z$ is the volume measure. The easiest way to see that this holds is to first assume that $\int f\,\mathrm{d}z$ vanishes and let h be the function solving $\Delta h = 2f$ and apply Itô's formula to $h(x_t)$ on the time interval $[s, s+t]$. Note that the intervals $[t_i, t_{i+1}]$ are either constant intervals or intervals of zero length with the exception of one bad interval of the form $[a, \tau^\epsilon]$ of size at most Δt. Let M be the integer such that $[t_i, t_{i+1}]$ are constant intervals for $i < M$. For the bad interval $[t_M, \tau^\epsilon)$, the following term in A_2:

$$\epsilon\int_{t_M}^{\tau^\epsilon} g(F_{r-t_n}(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))\,\mathrm{d}r$$

is of order $\epsilon\Delta t$. On each constant interval $[t_i, t_{i+1}]$, $i < M$, the corresponding term in A_2 is

$$\epsilon\int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))\,\mathrm{d}r = \epsilon\int_0^{\Delta t} g(F_r(y^\epsilon_{t_n}, \Theta_{t_n}(\omega)))\,\mathrm{d}r.$$

Denote by $\mu_{H^\epsilon(\epsilon t_n)}$, or $\mu_{y^\epsilon_{t_n}}$, the invariant measure on the invariant manifold $M_{H^\epsilon(\epsilon t_n)} \equiv M_{y^\epsilon_{t_n}}$. We apply the law of large numbers to such terms and use the Markov property of the flow to

obtain the following estimates, for all sufficiently small ϵ,

$$\begin{aligned}
&\left[\boldsymbol{E}\sup_{u\leqslant t}\left(A_2(u,\epsilon)-\epsilon\Delta t\sum_{n=0}^{N-1}\int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n),z)\,\mathrm{d}\mu_{H^\epsilon(\epsilon t_n)}(z)\right)^\beta\right]^{\frac{1}{\beta}}\\
&\qquad\leqslant C\epsilon\Delta t+N\sup_n\left[\boldsymbol{E}\left(\epsilon\int_0^{\Delta t} g(F_r(y^\epsilon_{t_n},\Theta_{t_n}(\omega)))\,\mathrm{d}r\right.\right.\\
&\qquad\quad\left.\left.-\epsilon\Delta t\int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n),z)\,\mathrm{d}\mu_{H^\epsilon(\epsilon t_n)}(z)\right)^\beta\right]^{\frac{1}{\beta}}\\
&\qquad\leqslant C\epsilon\Delta t+(\epsilon\Delta t)N\sup_n\left(\boldsymbol{E}\sup_{u\leqslant t}\left|\frac{1}{\Delta t}\int_0^{\Delta t} g(F_r(y^\epsilon_{t_n},\Theta_{t_n}(\omega)))\,\mathrm{d}r\right.\right.\\
&\qquad\quad\left.\left.-\int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n),z)\,\mathrm{d}\mu_{H^\epsilon(\epsilon t_n)}(z)\right|^\beta\right)^{\frac{1}{\beta}}\\
&\qquad\leqslant Ct\epsilon^{1-q}+C\frac{\epsilon\Delta t N}{\sqrt{t/\epsilon^q}}=\leqslant Ct\epsilon^{1-q}+C\frac{1}{\sqrt{t}}\epsilon^{\frac{q}{2}}.
\end{aligned}$$

On the other hand letting $s_n=\epsilon t_n$ so $\Delta s=\epsilon\Delta t$ and consider $s_0<s_1\cdots<s_N$. As Δs is of order $O(\epsilon^{1-q})$, the following pathwise estimate follows:

$$\begin{aligned}
&\left|\Delta s\sum_{n=0}^{N-1}\int_{M_{H^\epsilon(s_n)}} g(H^\epsilon(s_n),z)\,\mathrm{d}\mu_{H^\epsilon(s_n)}(z)-\int_{s\wedge\tau^\epsilon}^{(s+t)\wedge\tau^\epsilon}\int_{M_{H^\epsilon(s)}} g(H^\epsilon(s_n),z)\,\mathrm{d}\mu_{H^\epsilon(r)}(z)\,\mathrm{d}r\right|\\
&\qquad\leqslant C(g)t\epsilon^{1-q},
\end{aligned}$$

where $C(g)=\max_{U_0}|\mathrm{d}g|$. To summarize:

$$\int_{s\wedge\tau^\epsilon}^{(s+t)\wedge\tau^\epsilon} g(y^\epsilon_{r/\epsilon})\,\mathrm{d}r=\int_{s\wedge\tau^\epsilon}^{(s+t)\wedge\tau^\epsilon} Q^g(H^\epsilon(r))\,\mathrm{d}r+\delta(g,\epsilon,t),\tag{3.5}$$

where

$$\begin{aligned}
|\delta(g,\epsilon,t)|\leqslant{}& C\epsilon^{1-q}+\left|\epsilon\int_{t_N}^{\frac{(t+s)}{\epsilon}\wedge\tau^\epsilon} g(y^\epsilon_r)\,\mathrm{d}r\right|+|A_1(t,\epsilon)|\\
&+\left|A_2(t,\epsilon)-\sum\epsilon\Delta t\int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n),z)\,\mathrm{d}\mu_{H^\epsilon(\epsilon t_n)}(z)\right|\\
&+\left|\sum\Delta s\int_{M_{H^\epsilon(s_n)}} g(H^\epsilon(s_n),z)\,\mathrm{d}\mu_{H^\epsilon(s_n)}(z)\right.\\
&\left.-\int_{s\wedge\tau^\epsilon}^{(s+t)\wedge\tau^\epsilon}\int_{M_{H^\epsilon(s)}} g(H^\epsilon(s),z)\,\mathrm{d}\mu_{H^\epsilon(s)}(z)\,\mathrm{d}s\right|
\end{aligned}$$

and

$$A_1(t,\epsilon)=\epsilon\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}[g(y^\epsilon_r)-g(F_{r-t_n}(y^\epsilon_{t_n},\Theta_{t_n}(\omega)))]\,\mathrm{d}r.$$

By the previous estimates:

$$|\delta(g,\epsilon,t)|\leqslant Ct\epsilon^{1-q}+Ct^{-\frac{1}{2}}\epsilon^{q/2}+|A_1(t,\epsilon)|.$$

To show that $|A_1|$ is reasonably small, we apply lemma 3.1 and Hölder's inequality

$$\begin{aligned}
&\left(\boldsymbol{E}\sup_{s\leqslant t}(A_1(s,\epsilon))^{\beta}\right)^{\frac{1}{\beta}} \\
&\qquad\leqslant Ct\epsilon^{1-q}+\epsilon\left[\boldsymbol{E}\sup_{s\leqslant t}\left(\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}|g(y_r^{\epsilon})-g(F_{r-t_n}(y_{t_n}^{\epsilon},\Theta_{t_n}(\omega)))|\,\mathrm{d}r\right)^{\beta}\right]^{\frac{1}{\beta}} \\
&\qquad\leqslant Ct\epsilon^{1-q}+\epsilon\cdot N^{1-1/\beta}\left(\sum_{n=0}^{N-1}\boldsymbol{E}\left[\sup_{s\leqslant t}\int_{t_n}^{t_{n+1}}|g(y_r^{\epsilon})-g(F_{r-t_n}(y_{t_n}^{\epsilon},\Theta_{t_n}(\omega)))|\,\mathrm{d}r\right]^{\beta}\right)^{\frac{1}{\beta}} \\
&\qquad\leqslant Ct\epsilon^{1-q}+\epsilon N^{1-1/\beta}\cdot N^{\frac{1}{\beta}}C\cdot\epsilon(\Delta t+(\Delta t)^2) \\
&\qquad\leqslant C(t+t^2)\epsilon^{1-q}+Ct\epsilon.
\end{aligned}$$

Consequently,

$$\left(\boldsymbol{E}\sup_{s\leqslant t}|\delta(g,\epsilon,s)|^{\beta}\right)^{\frac{1}{\beta}}\leqslant C(t+t^2)\epsilon^{1-q}+Ct^{-\frac{1}{2}}\epsilon^{q/2}+Ct\epsilon$$

and finally take $q=2/3$ to obtain

$$\left\|\sup_{s\leqslant t}\delta(g,\epsilon,s)\right\|_{L_\beta}\leqslant C(t+t^2)\epsilon^{\frac{1}{3}}+C\epsilon^{\frac{1}{3}}t^{-\frac{1}{2}}. \tag{3.6}$$

□

Theorem 3.3. *Consider the stochastic differential equation (1.2) satisfying condition R. Let T^{ϵ} be the first time that the solution $y_{t/\epsilon}$ starting from y_0 exits U_0. Set*

$$H_i^{\epsilon}(t)=H_i(y_{t/\epsilon}^{\epsilon}).$$

(1) Let $\bar{H}(t)\equiv\bar{H}_t\equiv(\bar{H}_1(t),\ldots,\bar{H}_n(t))$ be the solution to the following system of deterministic equations.

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{H}_i(t)=\int_{M_{\bar{H}(t)}}\omega(X_{H_i},K)(\bar{H}(t),z)\,\mathrm{d}\mu_{\bar{H}_t}(z), \tag{3.7}$$

with initial condition $\bar{H}(0)=H(y_0)$. Let T^0 be the first time that $\bar{H}(t)$ exits from U_0. Then for all $t<T_0$, $\beta>1$, there exists a constant $C_2>0$ such that

$$\left(\boldsymbol{E}\left(\sup_{s\leqslant t}\|H^{\epsilon}(s\wedge T^{\epsilon})-\bar{H}(s\wedge T^{\epsilon})\|^{\beta}\right)\right)^{\frac{1}{\beta}}\leqslant C_2\epsilon^{1/3},$$

(2) Let $r>0$ be such that $U\equiv\{x:\|H(x)-H(y_0)\|\leqslant r\}\subset U_0$ and define

$$T_\delta=\inf_t\{|\bar{H}_t-H(y_0)|\geqslant r-\delta\}.$$

Then for any $\beta>1$, $\delta>0$ and a constant C depending on T_δ,

$$P(T^{\epsilon}<T_\delta)\leqslant C(T_\delta)\delta^{-\beta}\epsilon^{\beta/3}.$$

Remark 3.4. To see that (3.7) is a genuine system of ordinary differential equations, take the canonical transformation map $x_a:M_a\to T^n$. The pushed forward measure $x_*(\mu_a)$ is the Lebesque measure μ on the torus and (3.7) becomes:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{H}_i(t)=\int_{T^n}\omega(X_{H_i},K)(x_{\bar{H}_t}^{-1}(\bar{H}_t,z))\,\mathrm{d}\mu(z).$$

Proof. By Itô's formula, for $t < T_0 \wedge T^\epsilon$.

$$H_i^\epsilon(t) = H_i(y_0) + \int_0^t \omega(X_{H_i}, K)(y^\epsilon_{\frac{s}{\epsilon}})\,\mathrm{d}s.$$

For i fixed, write

$$g_i = \omega(X_{H_i}, K). \tag{3.8}$$

We only need to estimate

$$|H_i^\epsilon(t) - \bar{H}_i(t)| = \left| \int_0^t g_i(y^\epsilon_{s/\epsilon})\,\mathrm{d}s - \bar{H}_i(t) \right|. \tag{3.9}$$

Using the notation of the previous lemma then equation (3.7) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{H}_i(t) = Q^{g_i}(\bar{H}_t),$$
$$\bar{H}_0 = H(y_0).$$

Applying (3.3) to the functions g_i we have for any $t < T^\epsilon$,

$$|H_i^\epsilon(t \wedge T^\epsilon) - \bar{H}_i(t \wedge T^\epsilon)| \leqslant \int_0^{t\wedge T^\epsilon} |Q^{g_i}(H^\epsilon(s)) - Q^{g_i}(\bar{H}(s))|\,\mathrm{d}s + \delta(g_i, \epsilon, t)$$
$$\leqslant C(g, \phi) \int_0^t \|H^\epsilon(s)) - \bar{H}(s)\|, \mathrm{d}s + \delta(g_i, \epsilon, t).$$

By Gronwall's inequality,

$$\left(\boldsymbol{E}\left(\sup_{s \leqslant t\wedge T^\epsilon} \|H^\epsilon(s) - \bar{H}(s)\|^\beta \right) \right)^{\frac{1}{\beta}} \leqslant C(t)\epsilon^{\frac{1}{3}},$$

concluding part (1) of theorem 3.3.

Part (2) of the theorem easily follows. By definition T_δ is the first time that

$$\sqrt{\sum_i |\bar{H}_i(s) - H_i(y_0)|^2} \geqslant r - \delta$$

then

$$P(T^\epsilon < T_\delta) \leqslant P\left(\sup_{s \leqslant T_\delta \wedge T^\epsilon} \|\bar{H}_s - H^\epsilon(s)\| > \delta \right) \leqslant \delta^{-\beta} \boldsymbol{E}\left(\sup_{s \leqslant T_\delta \wedge T^\epsilon} \|\bar{H}_i(s) - H_i^\epsilon(s)\|^\beta \right)$$
$$\leqslant C\delta^{-\beta}\epsilon^{\frac{\beta}{3}}.$$

□

4. Perturbation by a Hamiltonian vector field

If the perturbation K to the stochastic Hamiltonian system (1.1) is a Hamiltonian vector field, i.e. $L_X\omega = 0$, then $\int_{M_c} \omega(X_{H_i}, K)\,\mathrm{d}\mu_c$ vanishes if c is not a bifurcation value. In fact let (U_0, ϕ) be an action-angle coordinate around M_c. We can write $K = X_k$ for some smooth function k,

$$\int_{M_c} \omega(X_{H_i}, K)(z)\,\mathrm{d}\mu_c(z) = \int_{T^n} d(k \circ \phi) \left(-\sum_{k=1}^n \frac{\partial(H_k \circ \phi)}{\partial I^k} \frac{\partial}{\partial \theta_k} \right) \mathrm{d}\theta$$
$$= -\sum_{\beta=1}^n \omega^i_\beta(I) \int_{T^n} \left(\frac{\partial}{\partial \theta_\beta} \right)(k \circ \phi)\ \mathrm{d}\theta = 0,$$

where $\mathrm{d}\theta$ is the standard measure on the n-torus. The ordinary differential equation (3.7) governing $\lim_{\epsilon\to 0} H_i(y^\epsilon_{t/\epsilon})$ thus has a constant solution. In this case we may consider the second order scaling $y^\epsilon_{t/\epsilon^2}$ and the accumulation of the perturbation over a large time interval of order ϵ^{-2}. The proof is inspired by a proof in Hairer–Pavliotis [14] and this also benefited from the papers by Khasminski, Papanicolau–Stroock–Varadhan and Freidlin.

Let

$$\mathcal{L}_0(I) = \frac{1}{2}\sum \boldsymbol{L}_{X_{H_i}}\boldsymbol{L}_{X_{H_i}} + \boldsymbol{L}_V$$

be the restriction of the elliptic operator on the invariant manifold with value I. If f on M_I has $\int f\,\mathrm{d}\mu = 0$ then the Poisson equation

$$\mathcal{L}_0 h = f \tag{4.1}$$

is solvable. We denote by $\mathcal{L}^{-1} f$ the solution to the Poisson equation satisfying $\int \mathcal{L}^{-1} f\,\mathrm{d}\mu = 0$.

Since $\mathcal{L}_0$ is elliptic on each level set manifold M_a and $\{H_i, k\}$ is centred there, the Poisson equation has a unique solution h_i. Note that the functions $\boldsymbol{L}_K\{H_i, k\}$ and $\boldsymbol{L}_{X_{H_i}} h_i$ are well defined.

Note that if $K = X_k$ then the matrix with the (i, j)th entry given by

$$-\mathrm{d}H_i(K)\mathcal{L}_0^{-1}(\mathrm{d}H_j(K)))$$

is positive definite.

Theorem 4.1. *Assume condition R and suppose that K is a smooth local Hamiltonian vector field so that $K = X_k$ for some smooth function k in the chart U_0. Define the matrices (a_{ij}) by*

$$a_{ij}(a) = -\int_{M_a} \omega(K, X_{H_j})\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))(a, z)\,\mathrm{d}\mu_a(z), \qquad a \in D \subset \boldsymbol{R}^n$$

and let (σ_i^j) be its square root. Set

$$b_j(a) = \frac{1}{2}\int_{M_a} \boldsymbol{L}_K \mathcal{L}_0^{-1}(\omega(X_{H_j}, K))(a, z)\,\mathrm{d}\mu_a(z).$$

Let z_t be the solution to the following stochastic differential equation

$$\mathrm{d}z_t^j = \sum_i \sigma_i^j(z_t) \circ \mathrm{d}B_t^i + b_j(z_t)\,\mathrm{d}t.$$

Then the law of the stochastic process $H(y^\epsilon_{t/\epsilon^2})$ stopped at S^ϵ, the first time that the process $y^\epsilon_{t/\epsilon^2}$ exits from U_0, converges to that of $H(z_{t\wedge S^\epsilon})$.

Remark. The limiting measure is clearly well defined as a_{ij} and b_j are invariant with different choices of the inverse to $\mathcal{L}_0$.

Proof. In the following calculation we restrict ourselves on the event $\{t < S^\epsilon\}$, equivalently consider the relevant processes stopped at S^ϵ. Set

$$\hat{y}_t^\epsilon = y^\epsilon_{\frac{t}{\epsilon^2}\wedge S^\epsilon},$$

$$\hat{H}^\epsilon(t) \equiv (\hat{H}_1^\epsilon(t), \ldots, \hat{H}_n^\epsilon(t)) = (H_1(\hat{y}_t^\epsilon), \ldots, H_n(\hat{y}_t^\epsilon)).$$

Then

$$\hat{H}_i^\epsilon(t) = H_i(y_0) - \epsilon\int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \omega(K, X_{H_i})(y_s^\epsilon)\,\mathrm{d}s.$$

We first show that the family of the laws μ^ϵ, distribution of $\hat{H}^\epsilon(t \wedge S^\epsilon)$, is tight. It follows, by Prohorov's theorem, that $\{\mu^\epsilon\}$ is relatively compact in the space of probability measures with the topology of weak convergence and hence has a weak limit $\bar{\mu}$. To see the tightness of the family μ^ϵ, we show that for any $a, \eta > 0$ there is a $\delta > 0$ with

$$P\left(\sup_{|s-t|<\delta} \|\hat{H}^\epsilon(t) - \hat{H}^\epsilon(s)\|^2 \geqslant a\right) \leqslant \eta.$$

For this, observe that

$$\|\hat{H}^\epsilon(t) - \hat{H}^\epsilon(s)\|^2 = \sum_{i=1}^n \left| -\epsilon \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \omega(K, X_{H_i})(y_r^\epsilon)\,\mathrm{d}r \right|^2.$$

Set h_i to be the solution to the Poisson equation:

$$\mathcal{L}_0 h_i = \omega(K, X_{H_i})$$

with $\int_{M_a} h_0 = 0$ for any $a \in \boldsymbol{R}^n$. Then

$$\begin{aligned}\|\hat{H}^\epsilon(t) - \hat{H}^\epsilon(s)\|^2 = \sum_{i=1}^n \Bigg| &\epsilon \sum_{j=1}^n \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_{X_{H_j}} h_i(y_r^\epsilon)\,\mathrm{d}B_r^j \\ &+ \epsilon^2 \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_K h_i(y_r^\epsilon)\,\mathrm{d}r - \epsilon h_i(\hat{y}_t^\epsilon) + \epsilon h_i(\hat{y}_s^\epsilon)\Bigg|^2.\end{aligned}$$

Applying lemma 3.2 with ϵ replaced by ϵ^2, one sees that the drift term has a nice bound in $|t-s|$:

$$\epsilon^2 \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_K h_i(y_r^\epsilon)\,\mathrm{d}r = \int_{s\wedge\tau^\epsilon}^{t\wedge\tau^\epsilon} \int_{\hat{H}^\epsilon(r)} \boldsymbol{L}_K h_i(z)\,\mathrm{d}\mu(z)\,\mathrm{d}r + \delta(\boldsymbol{L}_K h_i, \epsilon, t-s).$$

This gives us a comfortable estimate since $\delta(\boldsymbol{L}_K h_i, \epsilon, t-s)$ is of the order $\sqrt{\epsilon}(t-s)$. Similarly the quadratic variation of each of the martingale terms also converges with the same rate of convergence:

$$\boldsymbol{E}\left\langle \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_{X_{H_j}} h_i(y_r^\epsilon)\,\mathrm{d}B_r^j \right\rangle = \epsilon^2 \int_{s\wedge\tau^\epsilon}^{t\wedge\epsilon} \int \boldsymbol{E}|\boldsymbol{L}_{X_{H_j}} h_i(y_r^\epsilon)|^2\,\mathrm{d}r.$$

Applying Burkerholder–Gundy inequality to obtain an estimate on the L_β norm of

$$\sup_{|s-t|<\delta} \epsilon \left| \sum_{j=1}^n \int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_{X_{H_j}} h_i(y_r^\epsilon)\,\mathrm{d}B_r^j \right|,$$

which is a constant multiple of $|s-t|$ plus an error term of the order $\sqrt{\epsilon}(t-s)$.

Finally it is clear that

$$\sup_{|s-t|<\delta} |\epsilon h_i(\hat{y}_t^\epsilon) - \epsilon h_i(\hat{y}_s^\epsilon)|^2 \leqslant C\epsilon \to 0.$$

To identify the limiting measure let h be the solution to the Poisson equation,

$$h = \frac{1}{2}\sum_{i=1}^n \partial_i F(H)\mathcal{L}_0^{-1}(\omega(K, X_{H_i})),$$

where $\mathcal{L}_0^{-1}$ is considered to act on the angle variable only and $\int_{M_a} h = 0$ for each a. For any smooth function F on $\boldsymbol{R}^n$, we have

$$F(\hat{H}^\epsilon(t))) - F(\hat{H}^\epsilon(0)) = -\epsilon \sum_{i=1}^n \int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \partial_i F(H(y_s^\epsilon))\omega(K, X_{H_i})(y_s^\epsilon)\,\mathrm{d}s$$

$$= \epsilon \sum_{j=1}^n \int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_{X_{H_j}} h(y_s^\epsilon)\,\mathrm{d}B_s^j + \epsilon^2 \int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_K h(y_s^\epsilon)\,\mathrm{d}s + \epsilon(h(y_0) - h(\hat{y}_t^\epsilon)).$$

The first term on the right-hand side is a martingale and the last term converges to zero as $\epsilon \to 0$. We first identify $\boldsymbol{L}_K h$ in terms of the function F. By assumption the functions $\omega(K, X_{H_i})$ are centred and $\mathcal{L}_0^{-1}$ has no effect on functions of H and so

$$\boldsymbol{L}_K h = \frac{1}{2}\boldsymbol{L}_K \mathcal{L}_0^{-1}\Big(\sum_{i=1}^n \partial_i F(H)\omega(K, X_{H_i})\Big) = \frac{1}{2}\boldsymbol{L}_K\Big(\sum_{i=1}^n (\partial_i F)(H)\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))\Big)$$

$$= -\frac{1}{2}\sum_{i=1}^n \partial_j\partial_i F(H)\omega(K, X_{H_j})\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))$$

$$+ \frac{1}{2}\Big(\sum_{i=1}^n \partial_i F(H)\boldsymbol{L}_K \mathcal{L}_0^{-1}(\omega(X_{H_i}, K))\Big).$$

Set

$$\bar{\mathcal{L}} = -\frac{1}{2}\sum_{i,j} \omega(K, X_{H_j})\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))\partial_i\partial_j + \frac{1}{2}\sum_{i=1}^n \boldsymbol{L}_K \mathcal{L}_0^{-1}(\omega(K, X_{H_i}))\partial_i,$$

to see

$$F(\hat{H}^\epsilon(t))) - F(\hat{H}^\epsilon(0)) = \epsilon \sum_{j=1}^n \int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \boldsymbol{L}_{X_{H_j}} h(y_s^\epsilon)\,\mathrm{d}B_s^j$$

$$+ \epsilon^2 \int_0^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \bar{\mathcal{L}} F \circ H(y_s^\epsilon)\,\mathrm{d}s + \epsilon(h(y_0) - h(\hat{y}_t^\epsilon)).$$

Mimicking Papanicolaou–Stroock–Varadhan, we define $\mathcal{F}_s^H \equiv \sigma\{\hat{H}_{r\wedge S^\epsilon}^\epsilon : r \leqslant s\}$ and so $\{\mathcal{F}_s^H : s \geqslant 0\}$ is the filtration generated by $\hat{H}_{\cdot\wedge S^\epsilon}^\epsilon$. We need the following estimates:

$$\epsilon^2 \int_{a\wedge T^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon} \bar{\mathcal{L}} F(H(y_s^\epsilon))\,\mathrm{d}s = \int_{a\wedge T^\epsilon}^{t\wedge T^\epsilon} \Big(\int_{M_{\hat{H}^\epsilon(s)}} \bar{\mathcal{L}} F \circ H(z)\,\mathrm{d}\mu_{\hat{H}^\epsilon(s)}(z)\Big)\,\mathrm{d}s + \delta(\bar{\mathcal{L}} F, \epsilon^2, t-a)$$

$$= \int_{a\wedge T^\epsilon}^{t\wedge T^\epsilon} \mathcal{L} F \circ \hat{H}^\epsilon(s)\,\mathrm{d}s + \delta(\bar{\mathcal{L}} F \circ H, \epsilon^2, t-a),$$

where in the action-angle local coordinate,

$$\mathcal{L} F(a) = -\frac{1}{2}\sum_{i,j=1}^n \partial_j\partial_i F(a)\int_{M_a} \omega(K, X_{H_j})\mathcal{L}_0^{-1}(\omega(K, X_{H_i}))(a, z)\,\mathrm{d}\mu_I(z)$$

$$+ \frac{1}{2}\sum_{i=1}^n \partial_i F(a)\int_{M_a} \boldsymbol{L}_K \mathcal{L}_0^{-1}(\omega(X_{H_i}), K)(a, z)\,\mathrm{d}\mu(z).$$

Then for any $\mathcal{F}_s^H$- measurable L^2 random function G, using again lemma 3.2,

$$\boldsymbol{E}G\left[F(\hat{H}^\epsilon(t\wedge S^\epsilon))) - F(\hat{H}^\epsilon(s\wedge S^\epsilon))) - \int_{s\wedge S^\epsilon}^{t\wedge S^\epsilon}\int_{M_{\hat{H}^\epsilon(r)}}(\bar{\mathcal{L}}F)(z)\,\mathrm{d}\mu_{\hat{H}^\epsilon(r)}(z)\,\mathrm{d}r\right]$$

$$= \boldsymbol{E}G\left[\epsilon\sum_{j=1}^{n}\int_{\frac{s}{\epsilon^2}\wedge S^\epsilon}^{\frac{t}{\epsilon^2}\wedge S^\epsilon}\boldsymbol{L}_{X_{H_j}}h(y_s^\epsilon)\,\mathrm{d}B_s^j\right]$$

$$+\boldsymbol{E}G[\delta(\bar{\mathcal{L}}F,\epsilon^2,t-s)+\epsilon(h(y^\epsilon_{\frac{s}{\epsilon^2}\wedge S^\epsilon})-h(y^\epsilon_{\frac{t}{\epsilon^2}\wedge S^\epsilon}))]$$

$$= \boldsymbol{E}G[\delta(\bar{\mathcal{L}}F,\epsilon^2,t-s)+\epsilon(h(y^\epsilon_{\frac{s}{\epsilon^2}\wedge S^\epsilon})-h(y^\epsilon_{\frac{t}{\epsilon^2}\wedge S^\epsilon}))] \to 0.$$

Consequently

$$\boldsymbol{E}\left\{F(\hat{H}^\epsilon(t\wedge S^\epsilon))) - F(\hat{H}^\epsilon(s\wedge S^\epsilon))) - \int_{s\wedge S^\epsilon}^{t\wedge S^\epsilon}\int_{M_{\hat{H}^\epsilon(r)}}(\bar{\mathcal{L}}F)(z)\,\mathrm{d}\mu_{\hat{H}^\epsilon(r)}(z)\,\mathrm{d}r|\mathcal{F}_s^H\right\} \to 0,$$

and so any weak limit of the law $\hat{H}^\epsilon_\cdot$ is the solution to the martingale problem for the second order differential operator $\mathcal{L}$. □

Acknowledgments

This research benefited from a Royal Society Leverhulme Trust Senior Research Fellowship. This work has also benefitted from inspiring discussions with Martin Hairer. The author would also like to thank R Hudson and the referees for critical reading and comments.

References

[1] Arnold L, Imkeller P and Sri Namachchivaya N 2004 The asymptotic stability of a noisy non-linear oscillator *J. Sound Vib.* **269** 1003–29

[2] Arnold V I, Kozlov V V and Neĭshtadt A I 1993 Mathematical aspects of classical and celestial mechanics *Dynamical Systems, III* (*Encyclopaedia of Mathematical Sciences* vol 3) (Berlin: Springer)

[3] Abraham R and Marsden J E 1978 *Foundations of Mechanics* 2nd edn (New York: Benjamin/Cummings)

[4] Arnold V I 1989 *Mathematical Methods of Classical Mechanics* (*Graduate Texts in Mathematics* vol 60) 2nd edn (Berlin: Springer) Translators K Vogtmann and A Weinstein (Russian transl.)

[5] Bismut J-M 1981 *Mécanique Aléatoire* (*Lecture Notes in Mathematics* vol 866) (Berlin: Springer) (with an English summary)

[6] Borodin A and Freidlin M 1995 Fast oscillating random perturbations of dynamical systems with conservation laws *Ann. Inst. H. Poincaré Probab. Statist.* **31** 485–525

[7] Dolgopyat D 2004 Evolution of additive invariants in stochastic averaging *Stochastics Dyn.* **4** 265–75

[8] Eizenberg A and Freidlin M I 1993 Averaging principle for perturbed random evolution equations and corresponding Dirichlet problems *Probab. Theory Relat. Fields* **94** 335–74

[9] Freidlin M I and Wentzell A D 1993 Diffusion processes on graphs and the averaging principle *Ann. Probab.* **21** 2215–45

[10] Freidlin M I and Wentzell A D 1998 *Random Perturbations of Dynamical Systems Grundlehren der Mathematischen Wissenschaften* (*Fundamental Principles of Mathematical Sciences* vol 260) 2nd edn (New York: Springer) (translated from the 1979 Russian original by Joseph Szücs)

[11] Freidlin M and Weber M 2001 On random perturbations of Hamiltonian systems with many degrees of freedom *Stochastic Process. Appl.* **94** 199–239

[12] Freidlin M and Weber M 2004 Random perturbations of dynamical systems and diffusion processes with conservation laws *Probab. Theory Relat. Fields* **128** 441–66

[13] Givon D, Kupferman R and Stuart A 2004 Extracting macroscopic dynamics: model problems and algorithms *Nonlinearity* **17** R55–127

[14] Hairer M and Pavliotis G A 2004 Periodic homogenization for hypoelliptic diffusions *J. Statist. Phys.* **117** 261–79

[15] Hofer H and Zehnder E 1994 *Symplectic Invariants and Hamiltonian Dynamics* (*Birkhäuser Advanced Texts: Basler Lehrbücher*) (*Birkhäuser Advanced Texts: Basel Textbooks*) (Basel: Birkhäuser)

[16] Khas'minskii R Z 1964/1965 The behavior of a conservative system under the action of slight friction and slight random noise *J. Appl. Math. Mech.* **28** 1126–30

[17] Khasminskii R and Krylov N 2001 On averaging principle for diffusion processes with null-recurrent fast component *Stochastic Process. Appl.* **93** 229–40

[18] Koralov L 2004 Random perturbations of 2-dimensional Hamiltonian flows *Probab. Theory Relat. Fields* **129** 37–62

[19] Khasminskii R Z and Yin G 2004 On averaging principles: an asymptotic expansion approach *SIAM J. Math. Anal.* **35** 1534–60 (electronic)

[20] McDuff D and Salamon D 1995 *Introduction to Symplectic Topology* (*Oxford Mathematical Monographs*) (Oxford Science Publications) (New York: The Clarendon Press/Oxford University Press)

[21] Papanicolaou G C 1978 Asymptotic analysis of stochastic equations *Studies in Probability Theory* (*MAA Studies in Mathematics* vol 18) (Washington, DC: Mathematical Association of America) pp 111–79

[22] Papanicolaou G C, Stroock D and Varadhan S R S 1977 Martingale approach to some limit theorems *Papers from the Duke Turbulence Conference (Duke University, Durham, NC, 1976)* (*Duke University Mathematical Series* vol III) (Durham, NC: Duke University) Paper No 6, pages ii+120pp

[23] Sowers R B 2002 Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere) *Trans. Am. Math. Soc.* **354** 853–900 (electronic)

[24] Srinamachchivaya N and Sowers R B 2002 Rigorous stochastic averaging at a center with additive noise *Meccanica* **37** 85–114