

# AN AVERAGING PRINCIPLE FOR INTEGRABLE STOCHASTIC HAMILTONIAN SYSTEMS

XUE-MEI LI

**ABSTRACT.** Consider a stochastic differential equation whose diffusion vector fields are formed from an integrable family of Hamiltonian functions  $H_i, i = 1, \dots, n$ . We investigate the effect of a small transversal perturbation of order  $\epsilon$  to such a system. An averaging principle is shown to hold for this system and the action component of the solution converges, as  $\epsilon \rightarrow 0$ , 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 limiting deterministic system is constant we show that the action component of the solution scaled at  $1/\epsilon^2$  converges to that of a limiting stochastic differentiable equation.

## 1. INTRODUCTION

A time homogeneous stochastic differential equation is related to a partial differential equation and to the geometry and the topology of the state space through its infinitesimal generator, a second order differential operator in Hörmander form. If the second order differential operator has an ‘elliptic’ structure there corresponds an Riemannian metric on the state space and a linear connection on the relevant sub-bundle of the tangent bundle as has been investigated in [ELJL99]. The desire to understand elliptic systems leads to the question of the stability of the geometry when the elliptic system is perturbed. We may ask if the diffusion vector fields of the stochastic differential equation has a certain structure, for example if they form an Hamiltonian system, what properties or structures remain of the solution when a small perturbation is added and turned off slowly?

A deterministic Hamiltonian system on  $\mathbf{R}^2$  is of the form:  $\dot{q}_t = \frac{\partial H}{\partial p}, \dot{p}_t = -\frac{\partial H}{\partial q}$ , where  $H : \mathbf{R}^2 \rightarrow \mathbf{R}$  is a smooth function. The vector field  $(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q})$  is the Hamiltonian vector field associated to  $H$ . The motion of a Hamiltonian system will preserve the energy  $H$ . A small non-Hamiltonian forcing or vector field is added to a Hamiltonian system to model a motion which is not perfectly Hamiltonian. This could be a deterministic non-Hamiltonian vector field, or a Gaussian noise coming from the central limit theorems due to the accumulation of large quantities of small scale perturbations. For example one can consider:  $\dot{q}_t = \frac{\partial H}{\partial p} + \epsilon \dot{W}_t(\omega), \dot{p}_t = -\frac{\partial H}{\partial q} + \epsilon \dot{W}_t(\omega)$  where  $\dot{W}_t(\omega)$  stands for the standard white noise with chance element  $\omega$  and parameter  $\epsilon \geq 0$ . A valid question on perturbations is whether, or to which degree is, any of the characteristics of the dynamics of a Hamiltonian system stable under small perturbation. As the magnitude of the random perturbation may differ at different locations we are naturally lead to consider a multiplicative noise perturbation, e.g.  $\dot{q}_t = \frac{\partial H}{\partial p} + q_t \dot{W}_t(\omega), \dot{p}_t = -\frac{\partial H}{\partial q} + p_t \dot{W}_t(\omega)$ , in which case we have

---

2000 *Mathematics Subject Classification.* Primary 60H10, 58J65; Secondary 58J37.

*Key words and phrases.* Diffusion processes, symplectic manifold, integrable family of Hamiltonians, perturbation, averaging principle.

Research supported by a Royal Society Leverhulme Trust Senior Research Fellowship.

a nontrivial stochastic differential equations, a linear one in this case. Yet no particular structure is assumed of the noise in the above example. If the noise terms in a perturbed Hamiltonian system comes from a Hamiltonian function the system is called a stochastic Hamiltonian system. Stochastic Hamiltonian systems can be formally regarded as a family of time dependent Hamiltonian systems, with parameter  $\omega$ , where the dependence on time is continuous but not differentiable. An example of a linear stochastic Hamiltonian equation is:  $\dot{q}_t = p_t + p_t \dot{W}_t(\omega)$ ,  $\dot{p}_t = -q_t - q_t \dot{W}_t(\omega)$ ; and the following is one with additive noise  $\dot{q}_t = p_t$ ,  $\dot{p}_t = -q_t + W_t(\omega)$ . Note that the Hamiltonian functions  $(p_t^2 + q_t^2)W_t(\omega)$  and  $q_t W_t(\omega)$  are not bounded from below. Another example would be, by taking two independent white noises  $\dot{W}_t^1$  and  $\dot{W}_t^2$ :  $\dot{q}_t = p_t + p_t \dot{W}_t^1(\omega)$ ,  $\dot{p}_t = -q_t - q_t \dot{W}_t^2(\omega)$ . One related question is the rigidity of the stochastic Hamiltonian systems under small non Hamiltonian perturbations. Perturbed stochastic Hamiltonian systems have been studied by various authors in a number of contents, see *e.g.* Albeverio-Hilbert-Kolokotsov [AHK00] for asymptotics of some noisy Hamiltonian systems.

A family of  $n$  Hamiltonians  $\{H_k\}$  on a  $2n$  dimensional symplectic manifold is said to form a (completely) integrable system if they are pointwise Poisson commuting and that the corresponding Hamiltonian vector fields  $X_{H_k}$  are linearly independent at almost all points. Given such an integrable family and a local Hamiltonian vector field  $V$  commuting with the family of vector fields  $X_{H_k}$  in the sense of  $\omega(X_{H_k}, V) = 0$ , we can construct a so called *integrable stochastic symplectic/Hamiltonian system*:

$$(1.1) \quad dx_t = \sum_{k=1}^n X_{H_k}(x_t) \circ dB_t^k + V(x_t)dt,$$

where  $(B_t^k, k = 1, \dots, n)$  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. Note that we have suppressed the chance element  $\omega$  here as is conventional. We call respectively  $X_{H_k}$  the diffusion vector fields and  $V$  the drift vector field for the stochastic differential equation. Strictly speaking we ought to distinguish the use of the ‘stochastic symplectic systems’ from ‘stochastic Hamiltonian systems’ depending whether the vector fields concerned are Hamiltonian or merely locally Hamiltonian. But here we are not too worried about that unless there is a need to emphasise this aspect.

As we mentioned earlier a stochastic differential equation can be particularly interesting when a sufficient amount of noise is present, as the infinitesimal generator of the Markov solutions to the stochastic Hamiltonian system relates to geometry and plays a significant role in the analysis of solutions. In the integrable stochastic Hamiltonian system case, the diffusion vector fields span a sub-bundle of the tangent bundle, at least locally. Indeed the purpose of the present article is to investigate the effect of a small non-Hamiltonian, and Hamiltonian, 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 systems and there are corresponding invariant manifolds (level sets). The Markov solution restricts to each compact level set and the restriction has generator

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

Here  $\mathbf{L}_V$  indicates Lie differentiation in the direction of  $V$ . If the integrable stochastic Hamiltonian system is perturbed by a vector field  $\epsilon K$  for  $K$  a vector field not taking values

in the span of  $\{X_{H_k}, k = 1, 2 \dots n\}$  the solution to the resulting equation

$$\begin{aligned} dy_t^\epsilon &= \sum_{i=1}^n X_{H_i}(y_t^\epsilon) \circ dB_t^i + V(y_t^\epsilon)dt + \epsilon K(y_t^\epsilon)dt, \\ y_t^\epsilon &= y_0 \end{aligned} \tag{1.2}$$

will not conserve the energies. Here  $\epsilon$  is a parameter. On the other hand letting  $\epsilon \rightarrow 0$ , 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 generator given by  $\frac{1}{\epsilon}\mathcal{L}_0 + \mathbf{L}_K$ . Note that the motion splits into two parts with the fast component approximately an elliptic diffusion on the invariant torus and the slow motion governed by the transversal part of the vector field  $\epsilon K$ . 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  $\epsilon^{-1}$  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 is no diffusion terms, is ergodic and the evolution of the action component of  $y^\epsilon(t/\epsilon)$  will not depend on the angular variable in the limit as  $\epsilon \rightarrow 0$  and can be described by a system of  $n$ -dimensional ordinary differential equations whose vector fields are given by  $\omega(K, X_{H_i})$ ,  $i = 1, \dots, n$ . Here  $\omega$  is the symplectic 2-form. On the other hand 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  $y^\epsilon(t\epsilon^{-2})$  will involve a martingale term in the limit and the asymptotic law of  $y^\epsilon(t\epsilon^{-2})$  cross the level sets is shown to be given by a stochastic differential equation. However it remains open to find an estimate for the rate of the convergence of the law of  $y^\epsilon(t\epsilon^{-2})$ .

This work is in the framework of Arnold on averaging principle of 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 not bounded from below nor differentiable in time). Averaging of stochastic systems has been pioneered by Khasminskii [Kha64], Papanicolaou, Stroock and Varadhan [PSV77]. The structure of the main averaging results are very close to that described in the excellent survey of Papanicolaou [Pap78]. Investigations described in this article relate particularly to the body of work of Freidlin et al on random perturbations of Hamiltonian systems, see *e.g.* Borodin-Freidlin [BF95], Freidlin-Weber [FWeb01, FWeb04], Freidlin-Wentzell [FW93]. In their work, they allow generic random perturbations and the systems have less restrictive structure than ours and there are other differences. For example in Eizenberg-Freidlin [EF93] and Borodin-Freidlin [BF95], the diffusion part of the motion belongs to the slow component. In Freidlin-Weber [FWeb01, FWeb04] the fast component is allowed to be a diffusion, but the perturbation is assumed to be elliptic. We refer to Abraham-Marsden [AM78], Arnold [Arn89], Hofer-Zehnder [HZ94] and McDuff-Salamon [MS95] as references for Hamiltonian systems on symplectic manifolds, to Givon-Kupferman-Stuart [GKS04] for some physical models behind these problems and for recent progress in the direction of deterministic averaging, and to Freidlin-Wentzell [FW98], Sowers [Sow02], Korolov [Kor04], Khasminskii-Krylov [KK01], and Khasminskii-Yin [KY04] for related work on random perturbations of Hamiltonian systems as well as Arnold-Imkeller-Namachchivaya [AIN04] 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 stochastic framework we would like to refer to Bismut's work [Bis81]. However in this article we do not investigate the stochastic mechanics related to the SDEs.

## 2. PRELIMINARIES

**2.1. Hamiltonian and Symplectic Vector fields.** We will need the following terminology. 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  $\omega$  which is nondegenerate in the sense that for each  $x \in M$ ,  $\omega(v, w) = 0$  for all  $w \in T_x M$  implies  $v = 0$ . 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 will not be explicitly expressed unless indicated otherwise. Equivalently  $M$  admits a set of coordinates mapping such that the coordinate changing maps are symplectic on  $\mathbf{R}^{2n}$  with the standard symplectic form  $\omega_0 = \sum dp_i \wedge dq_i$ . 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  $\omega$ . 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 *local 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  $H : M \rightarrow \mathbf{R}$  we can associate a *Hamiltonian vector field* (also called *symplectic gradient vector field*) given by:

$$\iota_{X_H} \omega = dH.$$

The canonical symplectic structure on  $\mathbf{R}^{2n}$  with coordinates  $(q_1, \dots, q_n, p_1, \dots, p_n)$  is  $\omega = \sum_{i=1}^n dq_i \wedge dp_i$ . Darboux's theorem asserts that any symplectic manifold is locally  $\mathbf{R}^{2n}$  with its canonical symplectic structure. If the first de Rham cohomology  $\mathbb{H}^1(M; \mathbf{R})$  vanishes, as in the case of  $\mathbf{R}^{2n}$ , every Hamiltonian vector field is given by a Hamiltonian function. There are locally Hamiltonian vector fields which are not given by a Hamiltonian function. For example take the two torus  $T^2$  with coordinates  $x$  and  $y$  periodic in  $x$  and  $y$ . The canonical symplectic structure on  $\mathbf{R}^2$  induces the symplectic structure on  $T^2$ :  $\omega = dx \wedge dy$ . A vector field  $X(x, y) = a(x, y) \frac{\partial}{\partial x} + b(x, y) \frac{\partial}{\partial y}$  with  $\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} \equiv 0$  is clearly locally Hamiltonian as  $\omega(X, -) = a(x, y)dy - b(x, y)dx$ . However it is not given by a Hamiltonian function on  $T^2$  if  $(a(x, y), b(x, y)) \neq (0, 0)$  for any  $(x, y)$ .

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  $\{F_1, F_2\} = dF_1(X_{F_2}) = \omega(X_{F_1}, X_{F_2})$ . 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  $\mathbf{R}^{2n}$ .** The Hamiltonian vector field given by an Hamiltonian function  $H$  is given by  $X_H = JdH$  where  $J$  is the canonical complex structure:

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

where  $\mathbf{1}$  denotes the  $n$  by  $n$  identity matrix. The corresponding Hamiltonian system thus takes the familiar form

$$\begin{aligned}\dot{q}_i &= \frac{\partial H}{\partial p_i}, & 1 \leq i \leq n \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}, & 1 \leq i \leq n.\end{aligned}$$

For simplicity write  $p = (p_1, \dots, p_n)$  and  $q = (q_1, \dots, q_n)$ . An important class of examples of Hamiltonian functions on  $\mathbf{R}^{2n}$  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  $\mathbf{R}^{2n}$  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

$$\begin{aligned}dq_i(t) &= \frac{\partial K}{\partial p_i} dt + \sum_{k=1}^n \frac{\partial H_k}{\partial p_i} \circ dB_t^k \\ dp_i(t) &= -\frac{\partial K}{\partial q_i} dt + \sum_{k=1}^n \frac{\partial H_k}{\partial q_i} \circ dB_t^k.\end{aligned}$$

where

$$\begin{aligned}H_1 &= \frac{1}{2} \sum_{i=1}^n a_i^2 q_i^2 + \frac{1}{2} \sum_{i=1}^n p_i^2 \\ H_k &= \frac{1}{2} a_k q_k^2 + \frac{1}{2} p_k^2, & 2 \leq k \leq n,\end{aligned}$$

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

**2.3. The invariant manifolds and integrable symplectic Hamiltonian systems.** As before 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  $a = (a_1, \dots, a_n) \in \mathbf{R}^n$  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  $a = (H_1(x), \dots, H_n(x))$ , which we shall write also as  $M_{H(x)}$ .

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

$$dH_i(x_t) = \sum_k dH_i(X_{H_k}(x_t)) \circ dB_t^k + dH_i(V(x_t))dt = 0, \quad 1 \leq i \leq n.$$

For simplicity we shall assume throughout the paper the following:

- The invariant manifolds are compact,

which is the case if the map  $x \in M \mapsto (H_1(x), \dots, H_n(x)) \in \mathbf{R}^n$  is proper. Note that the  $n$  vector fields  $\{X_{H_i(x)}\}$  are tangent to  $M_{H(x)}$  and the symplectic form  $\omega$  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  $\mu_a$  the unique invariant probability measure on  $M_a$ ; it can be considered as the uniform probability measure on the torus as shall be seen below.

**2.4. The invariant measure and the divergence operator for semi-elliptic stochastic symplectic systems.** Let  $\{A^0, A^1, \dots, A^n\}$  be smooth symplectic vector fields with  $[A^i, A^j] = 0$  for all  $i, j$ . Assume that  $\{A^1, \dots, A^n\}$  spans a sub-bundle  $E$  of the tangent bundle  $TM$  of rank  $n$ . Consider the following stochastic differential equation:

$$(2.1) \quad dx_t = \sum_{i=1}^n A^i(x_t) \circ dB_t^i + A^0(x_t) dt.$$

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

For each  $x \in M$ , define a linear map  $A(x) : \mathbf{R}^{2n} \rightarrow T_x M$  by

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

where  $\{e^i\}$  is an orthonormal basis of  $\mathbf{R}^{2n}$ . The linear map is onto  $E_x$  with kernel  $\{0\} \times \mathbf{R}^n$  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  $\mathbf{R}^n \times \{0\}$  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  $\nabla_E f = \sum_i df(A^i) A^i$ . The symplectic structure  $\omega$  restricts to  $E$  defines a complex structure on  $M$  as following: we first give the tangent bundle  $TM$  any Riemannian metric which agrees with the one constructed on  $E$  using the linear map  $A$ . Define  $J_x : T_x M \rightarrow T_x M$  by

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

To see that this identity defines  $J_x$  uniquely suppose that for  $u \in T_x M$  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  $\mathbf{R}^{2n}$ .

Next take  $A^i = X_{H_i}$  in (2.1),  $i = 1, 2, \dots, n$ , 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  $X_{H_i}$ 's. Under our assumption that

$$H : x \rightarrow (H_1(x), \dots, H_n(x))$$

is a proper map, then for almost every point  $a_0$  in  $\mathbf{R}^n$  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  $\Sigma_H$  the set of values in  $\mathbf{R}^n$

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  $\omega^n$ , 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  $\mathbf{R}^n$ , see e.g. [AKN93]. More precisely if  $V$  is an open set of  $\mathbf{R}^n$  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  $(\sum_i dI^i \wedge d\theta^i)^n$  on the product space naturally splits to give us a probability measure, the Haar measure  $d\theta^1 \wedge \dots \wedge d\theta^n$  on  $T^n$ . We take the corresponding one on  $M_a$  and denote it by  $\mu_a$ . Let  $U$  be a section of  $E$ . Define the divergences  $\text{div}_E^a U$  to be the functions such that

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

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

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

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  $\mu_a$ , which varies smoothly with  $a$  in sufficiently small neighbourhoods of a regular value.

*Proof.* The measure  $\omega^n$  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 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  $\mu_a$  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.  $\square$

### 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  $H$  is proper and the set of critical points of the function  $H = (H_1, \dots, H_n) : M \rightarrow \mathbf{R}^n$  has measure zero and so the corresponding Hamiltonian vector fields are complete and that the solutions to (1.1) are globally defined. Note that the vector fields  $\{X_{H_i}\}$  forms an integrable distribution and so through each point of the manifold there is an integrable  $n$  dimensional manifold.

Take an action-angle coordinate:  $\phi^{-1} : U_0 \rightarrow D \times T^n$ . In this coordinate,  $x = \phi(I, \theta)$ ,  $I \in D$ ,  $\theta \in T^n$ , and  $(\phi^{-1})_* \omega = dI \wedge d\theta$  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  $\dot{I}_i^k =$

$$-\frac{\partial \tilde{H}_i}{\partial \theta_k} = 0 \text{ and}$$

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

with  $\omega_k^i$  smooth functions. In fact  $X_{\tilde{H}_i} = (\phi^{-1})_*(X_{H_i}) = -\sum_{k=1}^n \frac{\partial(H_i \circ \phi)}{\partial I_k} \frac{\partial}{\partial \theta_k}$ . 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  $H_k = \bar{q}_k$ ,  $k = 2, \dots, n$ , through the action angle coordinates change  $(q, p) \mapsto (\bar{q}, \bar{p})$ :

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

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

Let  $K$  be a smooth vector field, transversal in the sense that  $\omega(X_{H_i}, K)$  are not all identically zero. Denote by  $y_t^\epsilon$  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 shall 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:  $\phi^{-1} : U_0 \rightarrow D \times T^n$ , where  $D$  is an open set of  $\mathbf{R}^n$ .

We adopt the notation that if  $f$  is a function on  $U_0$ ,  $\tilde{f}$  shall be the representation of  $f$  in  $D \times T^n$ .

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

(1)

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

where  $C_1 = C_1(V, K, H_i, f)$  depends on the upper bounds of the functions  $|df|$ ,  $|\frac{\partial^2 \tilde{H}_k}{\partial I_i \partial I_j}|$ ,  $|d\tilde{V}|$ ,  $|\tilde{K}|$  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 shall write the flows in action-angle coordinates,  $x_t = \phi(I_t, \theta_t)$  and  $y_t^\epsilon = \phi(I_t^\epsilon, \theta_t^\epsilon)$ . Set  $\tilde{f} = f \circ \phi$ . Then

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

using the fact that  $\frac{\partial \tilde{f}}{\partial I}$  and  $\frac{\partial \tilde{f}}{\partial \theta}$  are bounded on  $T^n \times D$  as  $D$  is relatively compact. In the local chart,  $\frac{\partial \tilde{V}}{\partial \theta_i} = 0$  and we can write  $V(I, \theta) = V_j(I) \frac{\partial H_j}{\partial I}(I, \theta) = \omega_0^j(I) \frac{\partial}{\partial \theta_j}$  for some smooth functions  $\omega_0^j$  on  $D$ . The perturbation vector field can be written as  $(K^\theta, K^I)$  were  $K_\theta = (K_\theta^1, \dots, K_\theta^n)$  and  $K_I = (K_I^1, \dots, K_I^n)$  be 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  $T^n \times D$ :

$$\begin{aligned} dI_t^{i,\epsilon} &= \epsilon K_I^i(I_t^\epsilon, \theta_t^\epsilon) dt, \\ d\theta_t^{i,\epsilon} &= \sum_{k=1}^n \omega_k^i(I_t^\epsilon) \circ dB_t^k + \omega_0^i(I_t^\epsilon) dt + \epsilon K_\theta^i(I_t^\epsilon, \theta_t^\epsilon) dt, \end{aligned}$$

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

$$\sup_{s \leq t \wedge \tau^\epsilon} |I_s^\epsilon - I_s| = \epsilon \sup_{s \leq t \wedge \tau^\epsilon} \left| \int_0^s K^I(I_s^\epsilon, \theta_s^\epsilon) ds \right| \leq \epsilon t \sup_{D \times T^n} |K^I|,$$

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

$$\begin{aligned} \theta_s^{i,\epsilon} - \theta_s^i &= \sum_{k=1}^n \int_0^s \left( \omega_k^i(I_r^\epsilon) - \omega_k^i(I_r) \right) \circ dB_r^k \\ &\quad + \int_0^s \left( \omega_0^i(I_r^\epsilon) - \omega_0^i(I_r) \right) dr + \epsilon \int_0^s K_\theta^i(I_r^\epsilon, \theta_r^\epsilon) dr. \end{aligned}$$

As

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

Summing up over  $i$ , we have

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

by  $L_p$  inequalities for martingales. Combining the estimates we obtain

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

for some constant  $C_3$ .

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

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 changes 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  $t/\epsilon$  and we obtain a new dynamical system in the limit: as  $\epsilon$  goes to zero the motion along the torus moves very fast compared to 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 shall be shown below. Recall that  $H(x) = (H_1(x), \dots, H_n(x))$ .

We shall first prove a lemma:

**Lemma 3.2.** *Assume condition R holds. Let  $g$  be a real valued function on  $M$ , which is considered in the action angle co-ordinates as a function from  $D \times T^n$  to  $\mathbf{R}$ . Define  $Q^g : D \subset \mathbf{R}^n \rightarrow \mathbf{R}$  by its local representative :*

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

Suppose that  $g$  is  $C^1$  on  $U_0$ . Set

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

Then

$$(3.3) \quad \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} g(y_{r/\epsilon}^\epsilon) dr = \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} Q^g(H_\epsilon(r)) dr + \delta(g, \epsilon, t)$$

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

$$(3.4) \quad \left( \mathbf{E} \sup_{s \leq t} |\delta(g, \epsilon, s)|^\beta \right)^{\frac{1}{\beta}} \leq C(g)(t^2 + t)\epsilon^{1/2} + \epsilon\sqrt{t},$$

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

*Proof.* For  $q \in (0, 1)$ , let  $\Delta t = (t + s)/\epsilon^q \wedge T^\epsilon - s/\epsilon^q \wedge T^\epsilon$  and set  $N \equiv N(\epsilon) = [\epsilon^{q-1}] + 1$  where  $[\epsilon^{q-1}]$  is the integer part of  $\epsilon^{q-1}$  and all terms may depend on the sample paths of  $\omega$ . Take  $t_n = (s/\epsilon) \wedge T^\epsilon + n\Delta t$ ,  $1 \leq n \leq N$  so that

$$\frac{s}{\epsilon} \wedge T^\epsilon = t_0 < t_1 < \dots < t_N = \frac{s+t}{\epsilon} \wedge T^\epsilon$$

is a partition of  $[\frac{s}{\epsilon} \wedge T^\epsilon, \frac{s+t}{\epsilon} \wedge T^\epsilon]$ . We shall first make some pathwise estimates. Let  $\tau^\epsilon$  be the first time that  $y_t^\epsilon$  exit from  $U_0$ , then for any  $C^1$  function  $g$  on  $M$ ,

$$\begin{aligned} \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} g(y_{s/\epsilon}^\epsilon) ds &= \epsilon \int_{\frac{s}{\epsilon} \wedge T^\epsilon}^{\frac{s+t}{\epsilon} \wedge T^\epsilon} g(y_r^\epsilon) dr \\ &= \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(y_r^\epsilon) dr + \epsilon \int_{t_N}^{\frac{s+t}{\epsilon} \wedge T^\epsilon} g(y_r^\epsilon) dr \end{aligned}$$

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

$$\epsilon \left| \int_{t_N}^{\frac{s+t}{\epsilon} \wedge T^\epsilon} g(y_r^\epsilon) dr \right| \leq C\epsilon\Delta t \leq Ct\epsilon^{1-q}.$$

For the remaining terms we use the splitting

$$\begin{aligned} \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(y_r^\epsilon) dr &= \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} [g(y_r^\epsilon) - g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega)))] dr \\ &\quad + \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega))) dr. \end{aligned}$$

where  $\omega$  denotes the chance variable,  $\Theta_t$  the shift operator on the canonical probability space:  $\Theta_t(\omega)(-) = \omega(-+t) - \omega(t)$ , and  $\{F_t(x, \omega), t \geq 0\}$  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 first term we can apply a law of large numbers to  $A_1$ :

$$\begin{aligned} A_1(t, \epsilon) &\equiv \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega))) dr \\ &= \epsilon \Delta t \sum_{n=0}^{N-1} \frac{1}{\Delta t} \int_0^{\Delta t} g(F_r(y_{t_n}^\epsilon, \Theta_{t_n}(\omega))) dr. \end{aligned}$$

The invariant measure on the invariant manifold  $M_{H^\epsilon(\epsilon t_n)} \equiv M_{y_{t_n}^\epsilon}$  shall be denoted as  $\mu_{H^\epsilon(\epsilon t_n)}$  or as  $\mu_{y_{t_n}^\epsilon}$ . The law of the large numbers on  $T^n$  says that for any function  $f$  on a compact manifold and any stopping time  $S$ ,  $\frac{1}{(t+s) \wedge S - s \wedge S} \int_{s \wedge S}^{(s+t) \wedge S} f(x_r) dr$  converges to  $\int_M f(z) dz$  when  $t \rightarrow \infty$  with rate  $\frac{1}{\sqrt{t}}$  and the convergence is uniform on compact time intervals in  $L^p$  for all  $p > 1$ . Here  $dz$  is the volume measure. The easiest way to see this holds is to first assume that  $\int f dz$  vanishes and let  $h$  be the function solving  $\Delta h = 2f$  and apply Itô's formula to  $h(x_t)$  on the time interval  $[s \wedge S, (s+t) \wedge S]$ . By the Markov property of the flow we apply the law of large numbers and have the following estimates for all sufficiently small  $\epsilon$ ,

$$\begin{aligned} &E \sup_{u \leq t} \left( A_1(u, \epsilon) - \epsilon \Delta t \sum_{n=0}^{N-1} \int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n), z) d\mu_{H^\epsilon(\epsilon t_n)}(z) \right)^\beta \\ &\leq N^{\beta-1} \sum_{n=0}^{N-1} E \sup_{u \leq t} \epsilon \Delta t \\ &\quad \left| \frac{1}{\Delta t} \int_0^{\Delta t} g(F_r(y_{t_n}^\epsilon, \Theta_{t_n}(\omega))) dr - \int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n), z) d\mu_{H^\epsilon(\epsilon t_n)}(z) \right|^\beta \\ &\leq C(\epsilon \epsilon^{q-1} \frac{t}{\epsilon^q})^\beta (\frac{1}{\sqrt{t/\epsilon^q}})^\beta = C(\epsilon^{q/2} \sqrt{t})^\beta. \end{aligned}$$

On the other hand letting  $s_n = \epsilon t_n$  and so  $0 = s_0 < s_1 < \dots < s_N = t$  is a partition of  $[s \wedge \tau^\epsilon, (s+t) \wedge \tau^\epsilon]$  with  $\Delta s = \epsilon \Delta t \rightarrow 0$ , and we have the following pathwise estimate:

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

where  $C(g) = \max_{U_0} |dg|$ . Summarize we have:

$$(3.5) \quad \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} g(y_{r/\epsilon}^\epsilon) dr = \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} Q^g(H^\epsilon(r)) dr + \delta(g, \epsilon, t)$$

where

$$\begin{aligned}
\delta(g, \epsilon, t) &= \epsilon \int_{t_N}^{\frac{(t+s)}{\epsilon} \wedge T^\epsilon} g(y_r^\epsilon) dr + A_2(t, \epsilon) \\
&\quad + A_1(t, \epsilon) - \epsilon \Delta t \int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n), z) d\mu_{H^\epsilon(\epsilon t_n)}(z) \\
&\quad + \Delta s \int_{M_{H^\epsilon(s_n)}} g(H^\epsilon(s_n), z) d\mu_{H^\epsilon(s_n)}(z) \\
&\quad - \int_{s \wedge \tau^\epsilon}^{(s+t) \wedge \tau^\epsilon} \int_{M_{H^\epsilon(s)}} g(H^\epsilon(s), z) d\mu_{H^\epsilon(s)}(z) ds.
\end{aligned}$$

and

$$A_2(t, \epsilon) = \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} [g(y_r^\epsilon) - g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega)))] dr.$$

By the previous estimates:

$$\begin{aligned}
|\delta(g, \epsilon, t)| &\leq Ct\epsilon^{1-q} + C(g)\epsilon t^q + |A_2(t, \epsilon)| \\
&\quad + \left| A_1(t, \epsilon) - \epsilon \Delta t \sum_{n=0}^{N-1} \int_{M_{H^\epsilon(\epsilon t_n)}} g(H^\epsilon(\epsilon t_n), z) d\mu_{H^\epsilon(\epsilon t_n)}(z) \right|.
\end{aligned}$$

To show that  $|A_2|$  is reasonably small, we apply Lemma 3.1 and Hölder's inequality

$$\begin{aligned}
&\left( \mathbf{E} \sup_{s \leq t} (A_2(s, \epsilon))^\beta \right)^{\frac{1}{\beta}} \\
&= \epsilon \left[ \mathbf{E} \sup_{s < t \wedge T^\epsilon} \left( \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} |g(y_r^\epsilon) - g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega)))| dr \right)^\beta \right]^{\frac{1}{\beta}} \\
&\leq \epsilon \cdot N^{1-1/\beta} \left( \sum_{n=0}^{N-1} \mathbf{E} \left[ \sup_{s \leq t} \int_{t_n}^{t_{n+1}} |g(y_r^\epsilon) - g(F_{r-t_n}(y_{t_n}^\epsilon, \Theta_{t_n}(\omega)))| dr \right]^\beta \right)^{\frac{1}{\beta}} \\
&\leq \epsilon N^{1-1/\beta} \cdot N^{\frac{1}{\beta}} C \cdot \epsilon (\Delta t + (\Delta t)^2) \\
&\leq Ct^2 \epsilon^{1-q} + Ct \epsilon^{2-q}.
\end{aligned}$$

Consequently,

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

and finally

$$(3.6) \quad \left\| \sup_{s \leq t} \delta(g, \epsilon, s) \right\|_{L_\beta} \leq C(t + t^2) \epsilon^{\frac{1}{2}} + \epsilon \sqrt{t}.$$

□

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

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

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

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

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 some constant  $C_2$ ,  $t < T_0$ ,  $\beta > 1$ ,

$$\left( \mathbf{E} \left( \sup_{s \leq t} \|H^\epsilon(s \wedge T^\epsilon) - \bar{H}(s \wedge T^\epsilon)\|^\beta \right) \right)^{\frac{1}{\beta}} \leq C_2 \epsilon^{1/2},$$

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

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

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

$$P(T^\epsilon < T_\delta) \leq C(T_\delta) \delta^{-\beta} \epsilon^{\beta/2}.$$

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

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

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

$$H_i^\epsilon(t) = H_i(y_0) + \int_0^t \omega(X_{H_i}, K)(y_{s/\epsilon}^\epsilon) ds.$$

For  $i$  fixed, write

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

We only need to estimate

$$(3.9) \quad |H_i^\epsilon(t) - \bar{H}_i(t)| = \left| \int_0^t g_i(y_{s/\epsilon}^\epsilon) ds - \bar{H}_i(t) \right|$$

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

$$\begin{aligned} \frac{d}{dt} \bar{H}_i(t) &= Q^{g_i}(\bar{H}_t) \\ \bar{H}_0 &= H(y_0). \end{aligned}$$

Apply (3.5) to the functions  $g_i$  we have for any  $t < T^\epsilon$ ,

$$\begin{aligned} |H_i^\epsilon(t \wedge T^\epsilon) - \bar{H}_i(t \wedge T^\epsilon)| &\leq \int_0^{t \wedge T^\epsilon} |Q^{g_i}(H^\epsilon(s)) - Q^{g_i}(\bar{H}(s))| ds + \delta(g_i, \epsilon, t) \\ &\leq C(g, \phi) \int_0^t \|H^\epsilon(s) - \bar{H}(s)\| ds + \delta(g_i, \epsilon, t). \end{aligned}$$

By Gronwall's inequality,

$$\left( \mathbf{E} \left( \sup_{s \leq t \wedge T^\epsilon} \|H^\epsilon(s) - \bar{H}(s)\|^\beta \right) \right)^{\frac{1}{\beta}} \leq C_1 \epsilon^{1/2} (e^{-C_2 t} + C_3 t^2),$$

concluding part (1) of Theorem 3.3.

Part (2) of the theorem easily follows. By definition  $T_\delta$  is the first time that

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

then

$$\begin{aligned} P(T^\epsilon < T_\delta) &\leq P\left(\sup_{s \leq T_\delta \wedge T^\epsilon} \|\bar{H}_s - H^\epsilon(s)\| > \delta\right) \\ &\leq \delta^{-\beta} \mathbf{E}\left(\sup_{s \leq T_\delta \wedge T^\epsilon} \|\bar{H}_i(s) - H_i^\epsilon(s)\|^\beta\right) \\ &\leq C_1^\beta (e^{-C_2 t} + C_3 t^2)^\beta \delta^{-\beta} \epsilon^{\frac{\beta}{2}}. \end{aligned}$$

□

#### 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  $\int_{M_c} \omega(X_{H_i}, K) d\mu_c$  vanishes if  $c$  is not a bifurcation value. In fact let  $(U_0, \phi)$  be an action angle coordinate around  $M_c$ . We can write  $K = X_k$  for some smooth function  $k$ ,

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

where  $d\theta$  is the standard measure on the  $n$ -torus. The ordinary differential equation (3.7) governing  $\lim_{\epsilon \rightarrow 0} H_i(y_{t/\epsilon}^\epsilon)$  has thus a constant solution. In this case we may consider the second order scaling  $y_{t/\epsilon^2}^\epsilon$  and the accumulation of the perturbation over a large time interval of order  $\epsilon^{-2}$ . The proof is inspired by a proof in Hairer-Pavliotis [HP04] and this also benefited from the articles by Kahmsinski, Papanicolau-Stroock-Varadhan and Friedlin.

Let

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

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

$$(4.1) \quad \mathcal{L}_0 h = f.$$

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

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

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

$$-dH_i(K) \mathcal{L}_0^{-1} \left( dH_j(K) \right)$$

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

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

and let  $(\sigma_i^j)$  be its square root. Set

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

Let  $z_t$  be the solution to the following stochastic differential equation

$$dz_t^j = \sum_i \sigma_i^j(z_t) \circ dB_t^i + b_j(z_t) dt.$$

The law of the stochastic process  $H(y_{\frac{t}{\epsilon^2}})$  stopped at  $S^\epsilon$ , the first time that the process  $y_{\frac{t}{\epsilon^2}}$  exits from  $U_0$ , converging 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 shall restrict ourselves on the event  $\{t < S^\epsilon\}$ , equivalently consider the relevant processes stopped at  $S^\epsilon$ . Set

$$\begin{aligned} \hat{y}_t^\epsilon &= y_{\frac{t}{\epsilon^2} \wedge S^\epsilon}^\epsilon, \\ \hat{H}^\epsilon(t) &\equiv (\hat{H}_1^\epsilon(t), \dots, \hat{H}_n^\epsilon(t)) = (H_1(\hat{y}_t^\epsilon), \dots, H_n(\hat{y}_t^\epsilon)). \end{aligned}$$

Then

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

It is easy to see that the family of the laws  $\mu^\epsilon$ , distribution of  $\hat{H}^\epsilon(t \wedge S^\epsilon)$  is tight, i.e. relatively compact and has a weak limit  $\bar{\mu}$ . To see the tightness of the family  $\mu^\epsilon$ , we show that for any  $a, \eta > 0$  there is a  $\delta > 0$  with

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

For this, observe that

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

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  $a \in \mathbf{R}^n$ . Then

$$\begin{aligned} &\|\hat{H}^\epsilon(t) - \hat{H}^\epsilon(s)\|^2 \\ &= \sum_{i=1}^n \left| \epsilon \sum_{j=1}^n \int_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_{X_{H_j}} h_i(y_r^\epsilon) dB_r^j + \epsilon^2 \int_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_K h_i(y_r^\epsilon) dr - \epsilon h_i(\hat{y}_t^\epsilon) + \epsilon h_i(\hat{y}_s^\epsilon) \right|^2. \end{aligned}$$

Applying Lemma 3.2 with  $\epsilon$  replaced by  $\epsilon^2$ , one see that the drift term has a nice bound in  $|t - s|$ :

$$\begin{aligned} & \epsilon^2 \int_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_K h_i(y_r^\epsilon) dr \\ &= \int_{s \wedge \tau^\epsilon}^{t \wedge \tau^\epsilon} \int_{\hat{H}^\epsilon(r)} \mathbf{L}_K h_i(z) d\mu(z) dr + \delta(\mathbf{L}_K h_i, \epsilon, t - s), \end{aligned}$$

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

$$\mathbf{E} \left\langle \int_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_{X_{H_j}} h_i(y_r^\epsilon) dB_r^j \right\rangle = \epsilon^2 \int_{s \wedge \tau^\epsilon}^{t \wedge \tau^\epsilon} \int \mathbf{E} |\mathbf{L}_{X_{H_j}} h_i(y_r^\epsilon)|^2 dr.$$

Applying Burkholder-Gundy inequality to obtain an estimate on the  $L_\beta$  norm of

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

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

Finally it is clear that

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

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

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

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  $\mathbf{R}^n$ , we have

$$\begin{aligned} & F(\hat{H}^\epsilon(t)) - F(\hat{H}^\epsilon(0)) \\ &= -\epsilon \sum_{i=1}^n \int_0^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \partial_i F(H(y_s^\epsilon)) \omega(K, X_{H_i})(y_s^\epsilon) ds \\ &= \epsilon \sum_{j=1}^n \int_0^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_{X_{H_j}} h(y_s^\epsilon) dB_s^j + \epsilon^2 \int_0^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_K h(y_s^\epsilon) ds + \epsilon \left( h(y_0) - h(\hat{y}_t^\epsilon) \right). \end{aligned}$$

The first term on the right hand side is a martingale and the last term converges to zero as  $\epsilon \rightarrow 0$ . We shall first identify  $\mathbf{L}_K h$  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

$$\begin{aligned} \mathbf{L}_K h &= \frac{1}{2} \mathbf{L}_K \mathcal{L}_0^{-1} \left( \sum_{i=1}^n \partial_i F(H) \omega(K, X_{H_i}) \right) \\ &= \frac{1}{2} \mathbf{L}_K \left( \sum_{i=1}^n (\partial_i F)(H) \mathcal{L}_0^{-1}(\omega(K, X_{H_i})) \right) \\ &= -\frac{1}{2} \sum_{i=1}^n \partial_j \partial_i F(H) \omega(K, X_{H_j}) \mathcal{L}_0^{-1}(\omega(K, X_{H_i})) \\ &\quad + \frac{1}{2} \left( \sum_{i=1}^n \partial_i F(H) \mathbf{L}_K \mathcal{L}_0^{-1}(\omega(X_{H_i}, K)) \right). \end{aligned}$$

Set

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

to see

$$\begin{aligned} &F(\hat{H}^\epsilon(t)) - F(\hat{H}^\epsilon(0)) \\ &= \epsilon \sum_{j=1}^n \int_0^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_{X_{H_j}} h(y_s^\epsilon) dB_s^j + \epsilon^2 \int_0^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \bar{\mathcal{L}} F \circ H(y_s^\epsilon) ds + \epsilon (h(y_0) - h(\hat{y}_t^\epsilon)). \end{aligned}$$

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

$$\begin{aligned} &\epsilon^2 \int_{a \wedge T^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \bar{\mathcal{L}} F(H(y_s^\epsilon)) ds \\ &= \int_{a \wedge T^\epsilon}^{t \wedge T^\epsilon} \left( \int_{M_{\hat{H}^\epsilon(s)}} \bar{\mathcal{L}} F \circ H(z) d\mu_{\hat{H}^\epsilon(s)}(z) \right) ds + \delta(\bar{\mathcal{L}} F, \epsilon^2, t - a) \\ &= \int_{a \wedge T^\epsilon}^{t \wedge T^\epsilon} \mathcal{L} F \circ \hat{H}^\epsilon(s) ds + \delta(\bar{\mathcal{L}} F \circ H, \epsilon^2, t - a), \end{aligned}$$

where in the action-angle local coordinate,

$$\begin{aligned} \mathcal{L} F(a) &= -\frac{1}{2} \sum_{i,j=1}^n \partial_j \partial_i F(a) \int_{M_a} \omega(K, X_{H_j}) \mathcal{L}_0^{-1}(\omega(K, X_{H_i}))(a, z) d\mu_I(z) \\ &\quad + \frac{1}{2} \sum_{i=1}^n \partial_i F(a) \int_{M_a} \mathbf{L}_K \mathcal{L}_0^{-1}(\omega(X_{H_i}, K))(a, z) d\mu(z). \end{aligned}$$

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

$$\begin{aligned}
& \mathbf{E}G \left[ F(\hat{H}^\epsilon(t \wedge S^\epsilon)) - F(\hat{H}^\epsilon(s \wedge S^\epsilon)) - \int_{s \wedge S^\epsilon}^{t \wedge S^\epsilon} \int_{M_{\hat{H}^\epsilon(r)}} (\bar{\mathcal{L}}F)(z) d\mu_{\hat{H}^\epsilon(r)}(z) dr \right] \\
&= \mathbf{E}G \left[ \epsilon \sum_{j=1}^n \int_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^{\frac{t}{\epsilon^2} \wedge S^\epsilon} \mathbf{L}_{X_{H_j}} h(y_s^\epsilon) dB_s^j \right] \\
&\quad + \mathbf{E}G \left[ \delta(\bar{\mathcal{L}}F, \epsilon^2, t-s) + \epsilon \left( h(y_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^\epsilon) - h(y_{\frac{t}{\epsilon^2} \wedge S^\epsilon}^\epsilon) \right) \right] \\
&= \mathbf{E}G \left[ \delta(\bar{\mathcal{L}}F, \epsilon^2, t-s) + \epsilon \left( h(y_{\frac{s}{\epsilon^2} \wedge S^\epsilon}^\epsilon) - h(y_{\frac{t}{\epsilon^2} \wedge S^\epsilon}^\epsilon) \right) \right] \rightarrow 0.
\end{aligned}$$

Consequently

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

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

**Acknowledgement.** This research has benefitted from inspiring discussions with Martin Hairer.

#### REFERENCES

- [AHK00] Sergio Albeverio, Astrid Hilbert, and Vassily Kolokoltsov. Sur le comportement asymptotique du noyau associé à une diffusion dégénérée. *C. R. Math. Acad. Sci. Soc. R. Can.*, 22(4):151–159, 2000.
- [AIN04] L. Arnold, P. Imkeller, and N. Sri Namachchivaya. The asymptotic stability of a noisy non-linear oscillator. *J. Sound Vibration*, 269(3-5):1003–1029, 2004.
- [AKN93] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. Mathematical aspects of classical and celestial mechanics. In *Dynamical systems, III*, volume 3 of *Encyclopaedia Math. Sci.* Springer, Berlin, 1993.
- [AM78] R. Abraham and J.E. Marsden. *Foundations of Mechanics, second edition*. The Benjamin/Cummings publishing Company, 1978.
- [Arn89] V. I. Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1989. Translated from the Russian by K. Vogtmann and A. Weinstein.
- [Bis81] Jean-Michel Bismut. *Mécanique aléatoire*, volume 866 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1981. With an English summary.
- [BF95] A. Borodin and M. Freidlin, M. Fast oscillating random perturbations of dynamical systems with conservation laws. (English. English, French summary) *Ann. Inst. H. Poincaré Probab. Statist.* 31 (1995), no. 3, 485–525.
- [EF93] A. Eizenberg and M. I. Freidlin. Averaging principle for perturbed random evolution equations and corresponding Dirichlet problems. *Probab. Theory Related Fields*, 94(3):335–374, 1993.
- [ELJL99] K. D. Elworthy, Y. Le Jan, and Xue-Mei Li. *On the geometry of diffusion operators and stochastic flows*, volume 1720 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1999.
- [FW93] M. I. Freidlin and A. D. Wentzell. Diffusion processes on graphs and the averaging principle. *Ann. Probab.*, 21(4):2215–2245, 1993.
- [FW98] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szűcs.
- [FWeb01] M. Freidlin and M. Weber On random perturbations of Hamiltonian systems with many degrees of freedom. *Stochastic Process. Appl.* 94 (2001), no. 2, 199–239.
- [FWeb04] M. Freidlin and M. Weber Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab. Theory Related Fields* 128 (2004), no. 3, 441–466.

- [GKS04] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.
- [HP04] M. Hairer and G. A. Pavliotis. Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.*, 117(1-2):261–279, 2004.
- [HZ94] Helmut Hofer and Eduard Zehnder. *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, 1994.
- [Kha64] R. Z. Khas'minskii. The behavior of a conservative system under the action of slight friction and slight random noise. *J. Appl. Math. Mech.*, 28:1126–1130 (1965), 1964.
- [KK01] R. Khasminskii and N. Krylov. On averaging principle for diffusion processes with null-recurrent fast component. *Stochastic Process. Appl.*, 93(2):229–240, 2001.
- [Kor04] L. Koralov. Random perturbations of 2-dimensional Hamiltonian flows. *Probab. Theory Related Fields*, 129(1):37–62, 2004.
- [KY04] R. Z. Khasminskii and G. Yin. On averaging principles: an asymptotic expansion approach. *SIAM J. Math. Anal.*, 35(6):1534–1560 (electronic), 2004.
- [MS95] D. McDuff and D. Salamon. *Introduction to symplectic topology*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1995. Oxford Science Publications.
- [Pap78] G. C. Papanicolaou. Asymptotic analysis of stochastic equations. In *Studies in probability theory*, volume 18 of *MAA Stud. Math.*, pages 111–179. Math. Assoc. America, Washington, D.C., 1978.
- [PSV77] G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan. Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference (Duke Univ., Durham, N.C., 1976)*, Paper No. 6, pages ii+120 pp. Duke Univ. Math. Ser., Vol. III. Duke Univ., Durham, N.C., 1977.
- [Sow02] Richard B. Sowers. Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere). *Trans. Amer. Math. Soc.*, 354(3):853–900 (electronic), 2002.

MATHEMATICAL SCIENCES, LOUGHBOROUGH UNIVERSITY, LOUGHBOROUGH, LE11 3TU, U.K.

E-mail address: xue-mei.li@lboro.ac.uk