Deterministic and stochastic perturbations of area preserving flows on a two-dimensional torus

Dmitry Dolgopyat, Mark Freidlin, Leonid Koralov

Abstract

We study deterministic and stochastic perturbations of incompressible flows on a two-dimensional torus. Even in the case of purely deterministic perturbations, the long-time behavior of such flows can be stochastic. The stochasticity is caused by the instabilities near the saddle points as well as by the ergodic component of the locally Hamiltonian system on the torus.

Key words and phrases: Averaging, Markov Process, Hamiltonian Flow, Gluing Conditions, Diffusion on a Graph.

1 Introduction

Consider a Hamiltonian system with one degree of freedom

$$\dot{x}(t) = v(x(t)), \quad x(0) = x_0 \in \mathbb{R}^2, \tag{1}$$

where $v = \nabla^\perp H = (-H'_{x_2}, H'_{x_1})$ and $H(x)$, $x \in \mathbb{R}^2$, has bounded and continuous second derivatives. Then H is a first integral of (1): $H(x(t)) = H(x_0)$ for all t. Assume, for now, that $\lim_{|x|\to\infty} H(x) = +\infty$. Consider a small deterministic perturbation of (1):

$$\dot{\widetilde{x}}^\varepsilon(t) = v(\widetilde{x}^\varepsilon(t)) + \varepsilon\beta(\widetilde{x}^\varepsilon(t)), \quad \widetilde{x}^\varepsilon(0) = x_0,$$

where the vector field β is assumed to be bounded and continuously differentiable. It is clear that $\widetilde{x}^\varepsilon(t)$ is uniformly close to $x(t)$ on any finite time interval $[0, T]$ if ε is small enough:

$$\lim_{\varepsilon\downarrow 0} \max_{t\in[0,T]} |\widetilde{x}^\varepsilon(t) - x(t)| = 0.$$

Usually, however, one is interested in the behavior of $\widetilde{x}^\varepsilon(t)$ on time intervals that grow when $\varepsilon \downarrow 0$. Then, in general, $\widetilde{x}^\varepsilon(t)$ deviates significantly from $x(t)$. In order to describe such deviations, it is convenient to re-scale time by considering $x^\varepsilon(t) = \widetilde{x}^\varepsilon(t/\varepsilon)$. Then $x^\varepsilon(t)$ satisfies

$$\dot{x}^\varepsilon(t) = \frac{1}{\varepsilon} v(x^\varepsilon(t)) + \beta(x^\varepsilon(t)), \quad x^\varepsilon(0) = x_0. \tag{2}$$

The dynamics described by (2) consists of the fast motion (with speed of order $1/\varepsilon$) along the unperturbed trajectories of (1) together with the slow motion (with speed of order 1) in the direction transversal to the unperturbed trajectories.

Assume, for a moment, that the Hamiltonian H has just one well. Then the slow component of the motion can be described completely by the evolution of $H(x^\varepsilon(t))$:

$$H(x^\varepsilon(t)) - H(x_0) = \int_0^t \langle \beta(x^\varepsilon(s)), \nabla H(x^\varepsilon(s))\rangle ds.$$

Before $H(x^\varepsilon(t))$ changes by δ (a small constant independent of ε), the fast component makes a large number of rotations (of order δ/ε) along the unperturbed trajectory. The classical averaging principle (Chapter 10 of [2]) gives that

$$\lim_{\varepsilon\downarrow 0} H(x^\varepsilon(t)) = y(t)$$

uniformly on each finite time interval, where $y(t)$ is the solution of the averaged equation

$$\dot{y}(t) = \frac{\overline{\beta}(y(t))}{T(y(t))}, \quad y(0) = H(x_0). \tag{3}$$

Here

$$T(h) = \int_{\gamma(h)} \frac{dl}{|\nabla H|}$$

is the period of rotation along the level set $\gamma(h) = \{x \in \mathbb{R}^2 : H(x) = h\}$ and

$$\overline{\beta}(h) = \int_{\gamma(h)} \frac{\langle \beta, H\rangle}{|\nabla H|} dl.$$

Thus the long-time behavior of the perturbed system can be described in terms of the evolution of the slow component according to (3).

The situation becomes more complicated if the Hamiltonian has more than one well: first, since the system (1) has an additional (discrete) first integral and so the slow motion now has two components, and, second, since the limit $\lim_{\varepsilon\downarrow 0} H(x^\varepsilon(t))$ may not exist. In order to describe the slow motion, let us identify all the points that belong to the same connected component of a level set of H. Let h be the identification mapping. It is easy to see that the set $\mathbb{G} = h(\mathbb{R}^2)$ equipped with the natural topology is a graph (see Figure 1). Denote the edges of $\mathbb{G}$ by $I_1, ..., I_m$ and let $k(x)$ be the index of the edge such that $h(x) \in I_{k(x)}$. Thus we get the global coordinate system (k, H) on $\mathbb{G}$ (each interior vertex belongs to several edges, so it can be described by different coordinates). In this coordinate system $h(x) = (k(x), H(x))$, $x \in \mathbb{R}^2$. The integer-valued function $k(x)$, as well as $H(x)$ are first integrals for the unperturbed system (1), and $h(x^\varepsilon(t)) = (k(x^\varepsilon(t)), H(x^\varepsilon(t)))$ is the slow component of system (2). Due to instability of system (1) near the saddle points, the process $h(x^\varepsilon(t))$ is very sensitive to small changes of ε, and the limit $\lim_{\varepsilon\downarrow 0} h(x^\varepsilon(t))$ may not exist for a large class of perturbations.

Indeed, let γ be a separatrix loop of the Hamiltonian with a unique saddle point $O \in \gamma$. Thus γ separates the plane into two bounded domains, U_1 and U_2 (wells of the Hamiltonian), and one unbounded domain $\mathcal{C}$. Suppose that H does not have critical points in $\mathcal{C}$, and $H(x) > H(O)$ for $x \in \mathcal{C}$. Let $\mathrm{div}\beta(x) < 0$ for $x \in \mathbb{R}^2$, and $X_0^\varepsilon = x \in \mathcal{C}$. Put $T^\varepsilon = \inf\{t : X_t^\varepsilon \in \gamma\}$. One can check that $\lim_{\varepsilon\downarrow 0} T^\varepsilon = T^0 < \infty$, and $X^\varepsilon_{T^0+t}$ alternatingly belongs to U_1 or U_2 as $\varepsilon \downarrow 0$ for each $t > 0$. Since the limiting slow motions in different wells are, in general, different, the limit $\lim_{\varepsilon\downarrow 0} X^\varepsilon_{T^0+t}$ does not exist. The limit, in the sense of convergence in distribution of random processes, will exist in certain cases if the initial condition for the process is assumed to have a continuous density, although one can give examples of H and β when the limit does not exist even for continuously distributed initial conditions (see [4]).

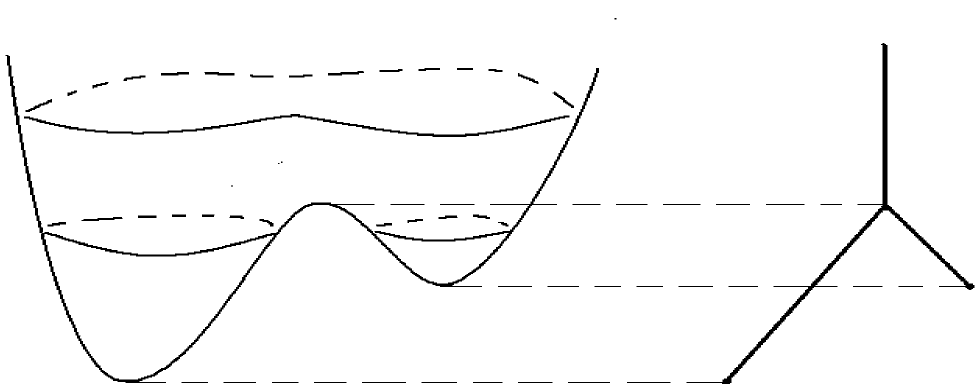

Figure 1: The graph of H and the corresponding graph $\mathbb{G}$.

On the other hand, one can consider perturbations of (1) that contain, besides the vector field $\varepsilon\beta(x)$, a diffusion term which is yet smaller than ε. More precisely, instead of equation (2) let us consider

$$dX_t^{\varkappa,\varepsilon} = \frac{1}{\varepsilon} v(X_t^{\varkappa,\varepsilon})dt + \beta(X_t^{\varkappa,\varepsilon})dt + \varkappa u(X_t^{\varkappa,\varepsilon})dt + \sqrt{\varkappa}\sigma(X_t^{\varkappa,\varepsilon})dW_t\ , \quad X_t^{\varkappa,\varepsilon} \in \mathbb{R}^2, \quad (4)$$

where u is a smooth bounded vector field, σ is a 2×2 smooth bounded matrix such that $\alpha(x) = \sigma(x)\sigma^*(x)$ is positive definite for all x, W_t is a two-dimensional Brownian motion and $\varkappa$ is a small parameter. The slow component $h(X_t^{\varkappa,\varepsilon})$ of the process (4) is a stochastic process on the graph $\mathbb{G}$. One can prove that for fixed $\varkappa$ the process $h(X_t^{\varkappa,\varepsilon})$ converges weakly, as $\varepsilon \downarrow 0$, to a diffusion process $Z_t^\varkappa$ on $\mathbb{G}$. All the diffusion processes on a graph were described in [10]. When $\varkappa \downarrow 0$, the processes $Z_t^\varkappa$ in their turn converge to a stochastic process Z_t on $\mathbb{G}$. The process Z_t is a deterministic motion inside each edge governed by the averaged equation considered above for the one-well case. A trajectory of Z_t can reach an interior vertex O of $\mathbb{G}$ in a finite time and leaves O immediately, going to one of the other two edges that have O as an end point, with probabilities $p_1(O)$ and $p_2(O)$ which can be calculated explicitly. These probabilities as well as the deterministic motion inside the edges are independent of the choice of the matrix σ and vector field u. This means that the convergence of the slow motion of a deterministically pertrubed deterministic system to the stochastic process Z_t is an intrinsic property of the system and the deterministic perturbation. The addition of a small stochastic term is used only as a regularization of the problem. The stochasticity of the limiting slow motion is actually a result of instability of system (1) near the saddle points. These results were obtained by Brin and Freidlin in [4] for the case when all the level sets of H are compact.

In the current paper we consider an incompressible periodic vector field v. We assume that v is typical in the sense that all the equilibrium points of v are non-degenerate, there are no saddle connections, and the projections of some of the flow lines on $\mathbb{T}^2$ are not periodic (the case when the projections of all the unbounded flow lines are periodic was covered in [4]). It has been conjectured by M. Freidlin ([9]) that the averaging principle (Theorem 1 below) holds for random perturbations of such flows. This has been proved by Dolgopyat and Koralov in [6] for generic flows (flows with Diophantine rotation numbers) and by Sowers in [16] (for flows whose stream function is nearly periodic). In [7] a general result for arbitrary rotation numbers was obtained, covering in particular the cases considered in [6] and [16]. An assumption was made that the Lebesgue measure on the torus was invariant for the diffusion processes that appear after the small perturbation of the original flow. In the current paper we get rid of this assumption and then study deterministic perturbations of such flows.

Let us start by describing the structure of the stream lines of the unperturbed flow. Since v is periodic, we can write H as

$$H(x_1, x_2) = H_0(x_1, x_2) + ax_1 + bx_2,$$

where H_0 is periodic. Note that a and b are rationally independent (otherwise all the unbounded flow lines would be periodic). It has been shown by Arnold in [1] that in this case the structure of the stream lines of v considered on the torus is as follows. There are finitely many domains U_k, $k = 1, \ldots, n$, bounded by the separatrices of the flow, such that the trajectories of the dynamical system $\dot{X}_t = v(X_t)$ in each U_k behave as in a part of the plane: they are either periodic or tend to a point where the vector field is equal to zero. The trajectories form one ergodic class outside of the domains U_k. More precisely, let $\mathcal{E} = \mathbb{T}^2 \backslash [\bigcup_{k=1}^n U_k]$. Here $[\cdot]$ stands for the closure of a set. Then the dynamical system is ergodic on $\mathcal{E}$ (and is, in fact, mixing for typical rotation numbers (see [12])).

Although H itself is not periodic, we can consider its critical points as points on the torus, since ∇H is periodic. All the maxima and the minima of H are located inside the domains U_k. There may also be saddle points of H inside some of the domains U_k, and the level sets containing such points will be the separatrices of the flow.

Let us introduce the finite graph $\mathbb{G}$ and the mapping $h : \mathbb{T}^2 \to \mathbb{G}$ that correspond to the structure of the stream lines of the flow on the torus. The graph is a tree and h maps the entire ergodic component to one point - to the root of the tree that will be denoted by O. Next, we identify all the points that belong to each of the compact flow lines. This way each connected domain bounded by the separatrices gets mapped into an edge of the graph, while the separatrices and the local maxima and minima of H get mapped into vertices of the graph (see Figure 2). In particular, the root of the graph serves as an end point for n edges (n is the number of domains U_k).

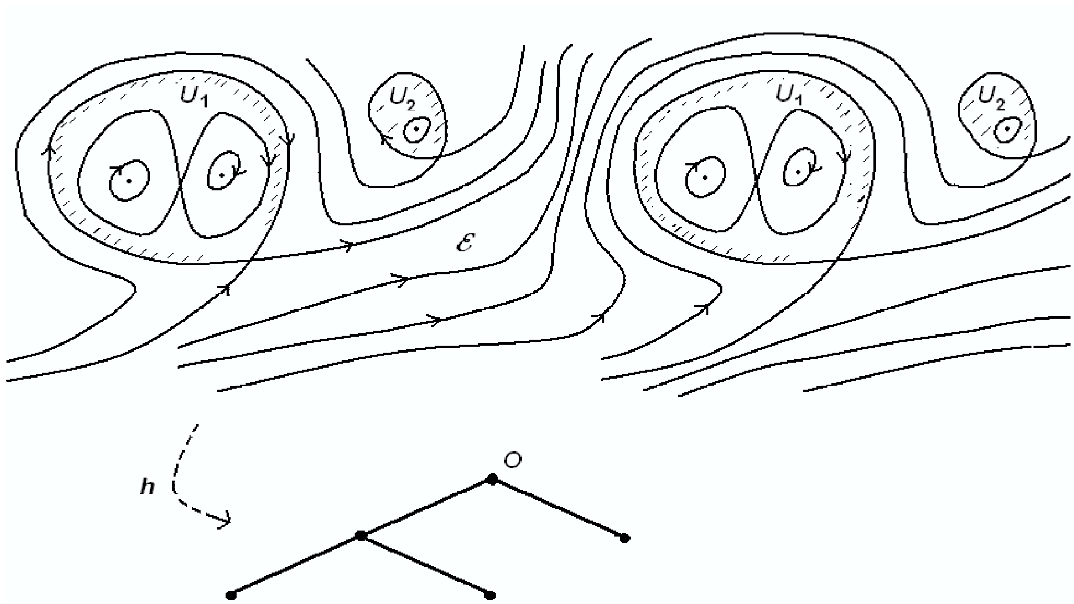

Figure 2: The stream lines of the flow and the corresponding graph

Let $I_1, ..., I_n$ be the edges of the graph. We can introduce coordinates h_k, $1 \leq k \leq n$, on the edges as follows. If V is a connected domain such that $H(V) = I_k$, $x_0 \in \partial V$ is such that $H(x_0) = y_0$, where y_0 is the end point of I_k that is closer to the root, and $x \in V$ is such that $H(x) = y$, then we put $h_k(y) = H(x) - H(x_0)$. Then the value of h_k together with the number of the edge k form a global coordinate system on $\mathbb{G}$ (each interior vertex belongs to several edges, so it can be described by different coordinates).

Now consider the process $X_t^{\varkappa,\varepsilon}$ on $\mathbb{T}^2$ given by the stochastic differential equation

$$dX_t^{\varkappa,\varepsilon} = \frac{1}{\varepsilon} v(X_t^{\varkappa,\varepsilon})dt + \beta(X_t^{\varkappa,\varepsilon})dt + \varkappa u(X_t^{\varkappa,\varepsilon})dt + \sqrt{\varkappa}\sigma(X_t^{\varkappa,\varepsilon})dW_t \,, \quad X_t^{\varkappa,\varepsilon} \in \mathbb{T}^2, \quad (5)$$

which can be viewed as a small stochastic perturbation of (2). Here v is an incompressible periodic vector field, β and u are periodic vector fields, σ is a 2×2 periodic matrix such that $\alpha(x) = \sigma(x)\sigma^*(x)$ is positive definite for all x, W_t is a two-dimensional Brownian motion and $\varkappa > 0$ is a small parameter. We assume that v, β, u and σ are infinitely smooth and have a common period in each of the variables that is equal to one and that the initial distribution of $X_t^{\varkappa,\varepsilon}$ does not depend on ε. We assume that the generator $L^{\varkappa,\varepsilon}$ of the process $X_t^{\varkappa,\varepsilon}$ can be written in the form

$$L^{\varkappa,\varepsilon} f = \frac{1}{\varepsilon}\langle v, \nabla f\rangle + \langle \beta, \nabla f\rangle + \frac{\varkappa}{2}\mathrm{div}(\alpha \nabla f),$$

that is

$$u_i = ((\alpha_{1i})'_{x_1} + (\alpha_{2i})'_{x_2})/2, \quad i = 1, 2. \tag{6}$$

The latter assumption is made only for simplicity of notation, it can be easily avoided by adding a small correction term to β.

Let $Y_t^{\varkappa,\varepsilon} = h(X_t^{\varkappa,\varepsilon})$ be the corresponding process on $\mathbb{G}$. In Section 2 we demonstrate that for fixed $\varkappa > 0$ the process $Y_t^{\varkappa,\varepsilon}$ converges, in the sense of weak convergence of induced measures, as $\varepsilon \downarrow 0$, to a Markov process on the graph. The limiting process will be denoted by $Z_t^{\varkappa}$. In Section 3 we identify the limit of $Z_t^{\varkappa}$ as $\varkappa \downarrow 0$ and show that it does not depend on the random perturbation (choice of the matrix-valued function α). The limiting process, which will be denoted by Z_t, moves deterministically along the edges of the graph. When it reaches a vertex, other than the root, it proceeds with deterministic motion along the "next" edge, which is chosen randomly with probabilities that depend on v and β. If the process reaches the root of the graph, it is delayed there for a random exponentially distributed time, and then moves along the "next" edge, which is chosen randomly.

The parameter of the exponential distribution is independent of the matrix α. This means that stochasticity at O is an intrinsic property of purely deterministic system (2).

2 Averaging principle for random perturbations

2.1 Formulation of the result

We assume for brevity that each of the domains U_k, $k = 1, ..., n$, contains a single critical point M_k of H (a maximum or a minimum of H). The general case can be easily considered using the results of this paper and [4]. Let A_k, $k = 1, ..., n$, be the saddle points of H, such that A_k is on the boundary of U_k. We denote the boundary of U_k by γ_k.

For now we are assuming that $\varkappa$ is fixed and ε tends to zero. Therefore, we can temporarily omit the dependence of the process on $\varkappa$ from the notations. Let X_t^ε solve the stochastic differential equation

$$dX_t^\varepsilon = \frac{1}{\varepsilon} v(X_t^\varepsilon)dt + \beta(X_t^\varepsilon)dt + u(X_t^\varepsilon)dt + \sigma(X_t^\varepsilon)dW_t\ , \quad X_t^\varepsilon \in \mathbb{T}^2. \tag{7}$$

We assume that the initial distribution of X_t^ε does not depend on ε.

The phase space of the limiting process will be a graph $\mathbb{G}$, which consists of n edges I_k, $k = 1, ..., n$, (segments labeled by k), where each segment is either $[H(M_k) - H(A_k), 0]$ (if M_k is a minimum) or $[0, H(M_k) - H(A_k)]$ (if M_k is a maximum). All the edges share a common vertex (the root) that will be denoted by O. Thus a point in $\mathbb{G} \setminus O$ can be determined by specifying k (the number of the edge) and the coordinate on the edge. We define the mapping $h : \mathbb{T}^2 \to \mathbb{G}$ as follows

$$h(x) = \begin{cases} O & \text{if } x \in [\mathcal{E}] \\ (k, H(x) - H(A)) & \text{if } x \in U_k, \end{cases}$$

where $[\mathcal{E}]$ is the closure of $\mathcal{E}$. We shall use the notation h_k for the coordinate on I_k. For a function f defined on $\mathbb{G}$ we will often write $f(h_k)$ instead of $f(k, h_k)$ when it is clear that the argument belongs to the k-th edge of the graph.

We denote the set $\{x \in [U_k] : H(x) - H(A) = h_k\}$ by $\gamma_k(h_k)$. Thus $\gamma_k = \gamma_k(0) = \partial U_k$. Let $L_k f(h_k) = a_k(h_k) f'' + b_k(h_k) f'$ be the differential operator on the interior of I_k with the coefficients

$$a_k(h_k) = \frac{1}{2} \left(\int_{\gamma_k(h_k)} \frac{1}{|\nabla H|} dl \right)^{-1} \int_{\gamma_k(h_k)} \frac{\langle \alpha \nabla H, \nabla H \rangle}{|\nabla H|} dl \quad \text{and} \tag{8}$$

$$b_k(h_k) = \frac{1}{2} \left(\int_{\gamma_k(h_k)} \frac{1}{|\nabla H|} dl \right)^{-1} \int_{\gamma_k(h_k)} \frac{2\langle \beta + u, \nabla H \rangle + \alpha \cdot H''}{|\nabla H|} dl, \tag{9}$$

where $\alpha \cdot H''(x) = \sum_{1 \le i,j \le 2} \alpha_{ij}(x) H''_{x_i x_j}(x)$. Let

$$p_k = \pm \frac{1}{2} (\text{Area}(\mathcal{E}))^{-1} \int_{\gamma_k} \frac{\langle \alpha \nabla H, \nabla H \rangle}{|\nabla H|} dl = \pm \frac{1}{2} (\text{Area}(\mathcal{E}))^{-1} \left| \int_{U_k} \text{div}(\alpha \nabla H)(x) dx \right|, \tag{10}$$

where the sign $+$ is taken if A_k is a local minimum for H restricted to U_k, and $-$ is taken otherwise.

Consider the process Y_t on $\mathbb{G}$ which is defined via its generator $\mathcal{L}$ as follows. The domain of $\mathcal{L}$ denoted by $D(\mathcal{L})$ consists of those functions $f \in C(\mathbb{G})$ which

(a) Are twice continuously differentiable in the interior of each of the edges;

(b) Have the limits $\lim_{h_k \to 0} L_k f(h_k)$ and $\lim_{h_k \to (H(M_k) - H(A_k))} L_k f(h_k)$ at the endpoints of each of the edges. Moreover, the value of the limit $q = \lim_{h_k \to 0} L_k f(h_k)$ is the same for all edges;

(c) Have the limits $\lim_{h_k \to 0} f'(h_k)$, and

$$\sum_{k=1}^{n} p_k \lim_{h_k \to 0} f'(h_k) = q. \tag{11}$$

For functions f which satisfy the above three properties, we define $\mathcal{L}f = L_k f$ in the interior of each edge, and as the limit of $L_k f$ at the endpoints of I_k.

It well-known (see [10], [15]) that there exists a strong Markov process on $\mathbb{G}$ with continuous trajectories, with the generator $\mathcal{L}$. The measure on $C([0, \infty), \mathbb{G})$ induced by the process is uniquely defined by the operator and the initial distribution of the process.

The rest of this section is devoted to the proof of the following theorem.

Theorem 1. *The measure on $C([0,\infty),\mathbb{G})$ induced by the process $Y_t^\varepsilon = h(X_t^\varepsilon)$ converges weakly to the measure induced by the process with the generator $\mathcal{L}$ with the initial distribution $h(X_0^\varepsilon)$.*

In the case $\beta = 0$ the limiting operator has a clear intuitive meaning. First let us consider the motion inside an edge. Let the support of $f \in D(\mathcal{L})$ belong to the interior of one edge I_k. Applying the Ito formula to $f(h_k(X_t^\varepsilon))$, we see that

$$f(h_k(X_t^\varepsilon)) - \int_0^t \left(a(X_s^\varepsilon) f''(h_k(X_s^\varepsilon)) + b(X_s^\varepsilon) f'(h_k(X_s^\varepsilon))\right) ds$$

is a martingale, where

$$a = \frac{1}{2}\langle \alpha \nabla H, \nabla H\rangle, \quad b = \langle u, \nabla H\rangle + \frac{1}{2}\alpha H''.$$

When ε is small, the trajectories of the diffusion process converge to the motion along the streamlines of H, so the integrals over time are well approximated by the averaged values over the streamlines.

Next we explain the gluing conditions. Note that for each ε the Lebesgue measure is invariant for process on $\mathbb{T}^2$, so its projection μ to $\mathbb{G}$ should be invariant for the limiting process. In other words, for each $f \in \mathcal{D}(\mathcal{L})$ we should have

$$\int_{\mathbb{G}} (\mathcal{L}f) d\mu = 0. \tag{12}$$

The projection has the following form

$$d\mu = \sum_{k=1}^n g_k(h_k) dh_k + \lambda(\mathcal{E})\delta_O,$$

where

$$g_k(h_k) = \int_{\gamma_k(h_k)} \frac{1}{|\nabla H|} dl.$$

Integrating by parts, we get

$$\int_{\mathbb{G}} (\mathcal{L}f) d\mu = \sum_{k=1}^n \int_{I_k} (\mathcal{L}^* g_k) dh_k + \lambda(\mathcal{E})\mathcal{L}f(0) + \sum_{k=1}^n s_k \left(((a_k g_k)' - b_k g_k) f - a_k g_k f'\right)(0),$$

where $s_k = 1$ if A_k is a local minimum for H restricted to U_k, and $s_k = -1$ otherwise.

Note that from (8), (9) and the Stokes formula it follows that $(g_k a_k)' = g_k b_k$. Therefore (12) reduces to

$$\lambda(\mathcal{E})\mathcal{L}f(0) = \sum_{k=1}^n s_k (a_k g_k f')(0),$$

which explains the choice of coefficients p_k in (11).

If $\beta \neq 0$, then the form of $\mathcal{L}$ inside the edges can be found by the same reasoning as above, but the meaning of the gluing conditions is less clear. The main result of this section is that the gluing conditions remain the same as in the incompressible case. Roughly speaking, the reason is the following. In [7] we showed that the orbit can not stay in $\mathcal{E}$ for a long time. Therefore, the Girsanov Theorem shows that the behavior of the process with $\beta \neq 0$ in a neighborhood of $\mathcal{E}$ should be similar to the behavior of the process with the same coefficients u and σ, but with $\beta = 0$, and so the gluing conditions for the two processes should be the same.

Let us now give a rigorous argument. We need the following lemma.

Lemma 2.1. *For any function $f \in D(\mathcal{L})$ and any $T > 0$ we have*

$$\mathbb{E}_x[f(h(X^\varepsilon_T)) - f(h(X^\varepsilon_0)) - \int_0^T \mathcal{L}f(h(X^\varepsilon_s))ds] \to 0 \quad \text{as } \varepsilon \to 0, \tag{13}$$

uniformly in $x \in \mathbb{T}^2$.

An analogous lemma was used in the monograph of Freidlin and Wentzell ([11], Chapter 8) to justify the convergence of the process Y^ε_t to the limiting process on the graph. The main idea, roughly speaking, is to use the tightness of the family Y^ε_t, and then to show that the limiting process (along any subsequence), is a solution of the martingale problem, corresponding to the operator $\mathcal{L}$.

The main difference between our case and that of [11] is the presence of an ergodic component. However, all the arguments used to prove the main theorem based on (13) remain the same. Thus, upon referring to Lemma 3.1 of [11], it is enough to prove our Lemma 2.1 above.

Lemma 2.1 was proved in [7] for the case when $\beta = 0$. One of the important ingredients in the proof is an estimate on the time it takes for the process to exit the ergodic component. The estimate uses the results of [8] and [17] (which are also closely related to [3] and [5]) which allow one to relate the time it takes the process to exit $\mathcal{E}$ to the spectral properties of the operators related to the generator of the process in $\mathcal{E}$. In the current paper we reduce the general case to the one with $\beta = 0$.

2.2 Proof of Lemma 2.1.

Before we proceed with the rigorous arguments, let us briefly discuss the main idea for the proof of Lemma 2.1. Together with X^ε_t, we will consider an auxiliary process $\widetilde{X}^\varepsilon_t$ obtained from X^ε_t by setting $\beta = 0$ in the right hand side of (7). For $\rho > 0$, the measure on $C([0, \rho])$ induced by X^ε_t is absolutely continuous with respect to the measure induced by $\widetilde{X}^\varepsilon_t$ starting at the same point, and the density of the first measure with respect to the second one is close to one if ρ is small, as follows from the Girsanov theorem. We can split the interval $[0, T]$ into subintervals of length ρ and consider the contribution to the expectation in (13) from each of the small intervals separately. We further split each of the small intervals into random subintervals as follows. Introduce the curves $\overline{\gamma}_k$ inside U_k that are asymptotically close to γ_k when $\varepsilon \downarrow 0$, thus separating the time axis into the intervals (between hitting γ_k and γ) that the process spends in U_k on the way to the ergodic component and the intervals (between hitting γ and hitting one of the intervals $\overline{\gamma}_k$) that the process spends in a neighborhood of $\mathcal{E}$. The contribution from the intervals of the first type is treated using the classical averaging theory. The contribution from the intervals of the second type is compared to the contribution from the same intervals for the auxiliary process, for which the result is already available, using the fact that the measures induced by the two processes are similar.

The proof of Lemma 2.1 will rely on several other lemmas. Below we shall introduce a number of processes, stopping times, and sets, which will depend on ε. However, we shall not always incorporate this dependence on ε into notation, so one must be careful to distinguish between the objects which do not depend on ε and those which do.

Fix an arbitrary $\alpha \in (1/4, 1/2)$. Let $\overline{\gamma}_k = \gamma_k(\varepsilon^\alpha)$ and $\overline{\gamma} = \bigcup_{k=1}^n \overline{\gamma}_k$. Let $\gamma = \bigcup_{k=1}^n \gamma_k$ be the boundary of $U = \bigcup_{k=1}^n U_k$. Let σ be the first time when the process X_t^ε reaches γ and τ be the first time when the process reaches $\overline{\gamma}$.

We inductively define the following two sequences of stopping times. Let $\sigma_0 = \sigma$. For $n \geq 0$ let τ_n be the first time following σ_n when the process reaches $\overline{\gamma}$. For $n \geq 1$ let σ_n be the first time following τ_{n-1} when the process reaches γ.

Lemma 2.2. *We have the following limit*

$$\lim_{\varepsilon \downarrow 0} \sup_{x \in \mathrm{Cl}(\mathcal{E})} \mathbb{E}_x \sigma = 0.$$

Proof. The same lemma was proved in [7] under the additional assumption that $\beta = 0$. Consider an auxiliary process $\widetilde{X}_t^\varepsilon$ obtained from X_t^ε by setting $\beta = 0$ in the right hand side of (7). Since the result holds when $\beta = 0$, for each $\delta > 0$ we have

$$\lim_{\varepsilon \downarrow 0} \sup_{x \in \mathrm{Cl}(\mathcal{E})} \mathbb{P}_x(\widetilde{\sigma} > \delta) = 0, \tag{14}$$

where $\widetilde{\sigma}$ is the first time when the process $\widetilde{X}_t^\varepsilon$ reaches γ. Let μ_x^ε be the measure on $C([0,\delta], \mathbb{T}^2)$ induced by X_t^ε starting at x and $\widetilde{\mu}_x^\varepsilon$ the measure induced by $\widetilde{X}_t^\varepsilon$ starting at x. By the Girsanov theorem, μ_x^ε is absolutely continuous with respect to $\widetilde{\mu}_x^\varepsilon$ with a density p_x^ε, and

$$\inf_{x \in \mathrm{Cl}(\mathcal{E})} \widetilde{\mu}_x^\varepsilon(p_x^\varepsilon \geq 3/4) \geq 3/4$$

provided that δ is sufficiently small. Therefore, by (14),

$$\sup_{x \in \mathrm{Cl}(\mathcal{E})} \mathbb{P}_x(\sigma > \delta) \leq 1/2,$$

provided that ε is sufficiently small. Since δ was arbitrary, the lemma follows from the Markov property of the process X_t^ε. □

Lemma 2.3. *For each function $f \in D(\mathcal{L})$ we have*

$$\sup_{x\in\mathbb{T}^2}\sup_{\sigma'\le\sigma}|\mathbb{E}_x[f(h(X^\varepsilon_{\sigma'}))-f(h(X^\varepsilon_0))-\int_0^{\sigma'}\mathcal{L}f(h(X^\varepsilon_s))ds]|\to 0 \quad \text{as } \varepsilon\to 0, \tag{15}$$

where the first supremum is taken over all stopping times $\sigma' \le \sigma$.

Proof. If the supremum is restricted to the set $\mathbb{T}^2\setminus\mathrm{Cl}(\mathcal{E})$, then the statement follows from the averaging principle inside a periodic component (see [13] for the case when there are no saddle points inside U_k and [11] in the general case). The statement with the supremum taken over $\mathrm{Cl}(\mathcal{E})$ immediately follows from Lemma 2.2 if one takes into account that $f(h(x)) = \mathrm{const}$ for $x \in \mathrm{Cl}(\mathcal{E})$. □

Lemma 2.4. *For each function $f \in D(\mathcal{L})$ we have*

$$\sup_{x\in\mathbb{T}^2}|\mathbb{E}_x[f(h(X^\varepsilon_{\tau}))-f(h(X^\varepsilon_0))-\int_0^{\tau}\mathcal{L}f(h(X^\varepsilon_s))ds]|\to 0 \quad \text{as } \varepsilon\to 0. \tag{16}$$

Proof. Let $\overline{U}\subseteq U$ be the union of the domains bounded by $\overline{\gamma}$. Note that

$$\sup_{x\in\overline{U}}|\mathbb{E}_x[f(h(X^\varepsilon_{\tau}))-f(h(X^\varepsilon_0))-\int_0^{\tau}\mathcal{L}f(h(X^\varepsilon_s))ds]|\to 0 \quad \text{as } \varepsilon\to 0,$$

as follows from the classical averaging principle. Since $\mathcal{L}f$ is bounded and f is nearly constant on $\mathbb{T}^2\setminus\overline{U}$ (which is a small neighborhood of $\mathcal{E}$), it is sufficient to show that

$$\sup_{x\in\mathbb{T}^2\setminus\overline{U}}\mathbb{E}_x\tau\to 0 \quad \text{as } \varepsilon\to 0. \tag{17}$$

This has been done in [7] for the case when $\beta = 0$. The general case follows from the Markov property of the process and the Girsanov theorem the same way as in the proof of Lemma 2.2. □

Lemma 2.5. *For each function $f \in \mathcal{D}$ we have the following asymptotic estimate*

$$\sup_{x\in\overline{\gamma}}\sup_{\sigma'\le\sigma}|\mathbb{E}_x[f(h(X^\varepsilon_{\sigma'}))-f(h(X^\varepsilon_0))-\int_0^{\sigma'}\mathcal{L}f(h(X^\varepsilon_s))ds]| = o(\varepsilon^\alpha) \quad \text{as } \varepsilon\to 0, \tag{18}$$

where the first supremum is taken over all stopping times $\sigma' \le \sigma$.

This lemma is similar to the averaging principle inside the periodic component – the difference is that now the initial point is not fixed, but is located at a distance of order ε^α from the boundary of the periodic component. This guarantees that the expectation of the exit time from the periodic component is of order $O(\varepsilon^\alpha)$, which allows for the $o(\varepsilon^\alpha)$ estimate of the left hand side of (18). The needed modifications to the averaging principle are not difficult (see, for example, Lemma 4.4 of [14] where a similar statement was proved).

We'll need to control the number of excursions between $\overline{\gamma}$ and γ before time T. For this purpose we formulate the following lemma, whose proof is similar to that of Lemma 2.5 in [6].

Lemma 2.6. *There is a constant $r > 0$, such that for all sufficiently small ε we have*

$$\sup_{x\in\overline{\gamma}} \mathbb{E}_x e^{-\sigma} \le 1 - r\varepsilon^{\alpha}.$$

Using the Markov property of the process and Lemma 2.6, for $n \ge 1$ we get the estimate

$$\sup_{x\in\mathbb{T}^2} \mathbb{E}_x e^{-\sigma_n} \le \sup_{x\in\overline{\gamma}} \mathbb{E}_x e^{-\sigma_{n-1}} \le (\sup_{x\in\overline{\gamma}} \mathbb{E}_x e^{-\sigma})^{n-1} \le (1 - r\varepsilon^{\alpha})^{n-1}. \tag{19}$$

The first inequality here follows from the definition of σ_n. Note that for $n \ge 0$

$$\mathbb{P}_x(\tau_n < T) \le \mathbb{P}_x(\sigma_n < T) \le \mathbb{P}_x(e^{-\sigma_n} > e^{-T}) \le e^T(1 - r\varepsilon^{\alpha})^{n-1}.$$

The last inequality here is due to (19) and the Chebyshev inequality if $n \ge 1$ and is obvious for $n = 0$. Taking the sum in n, we obtain

$$\sum_{n=0}^{\infty} \mathbb{P}_x(\tau_n < T) \le \sum_{n=0}^{\infty} \mathbb{P}_x(\sigma_n < T) \le \sum_{n=0}^{\infty} e^T(1 - r\varepsilon^{\alpha})^{n-1} \le K\varepsilon^{-\alpha}, \tag{20}$$

where the constant K depends on T.

Proof of Lemma 2.1. Let $f \in \mathcal{D}$, $T > 0$, and $\eta > 0$ be fixed. We would like to show that the absolute value of the left hand side of (13) is less than η for all sufficiently small positive ε. Using the stopping times τ_n and σ_n we can rewrite the expectation in the left hand side of (13) as follows

$$\begin{gathered}\mathbb{E}_x[f(h(X^{\varepsilon}_T)) - f(h(X^{\varepsilon}_0)) - \int_0^T \mathcal{L}f(h(X^{\varepsilon}_s))ds] = \\ \mathbb{E}_x[f(h(X^{\varepsilon}_{T\wedge\sigma})) - f(h(X^{\varepsilon}_0)) - \int_0^{T\wedge\sigma} \mathcal{L}f(h(X^{\varepsilon}_s))ds] + \\ \sum_{n=0}^{\infty} \mathbb{E}_x(\chi_{\{\sigma_n<T\}}[f(h(X^{\varepsilon}_{\tau_n\wedge T})) - f(h(X^{\varepsilon}_{\sigma_n})) - \int_{\sigma_n}^{\tau_n\wedge T} \mathcal{L}f(h(X^{\varepsilon}_s))ds]) + \\ \sum_{n=0}^{\infty} \mathbb{E}_x(\chi_{\{\tau_n<T\}}[f(h(X^{\varepsilon}_{\sigma_{n+1}\wedge T})) - f(h(X^{\varepsilon}_{\tau_n})) - \int_{\tau_n}^{\sigma_{n+1}\wedge T} \mathcal{L}f(h(X^{\varepsilon}_s))ds]),\end{gathered} \tag{21}$$

provided that the sums in the right hand side converge absolutely (which follows from the arguments below). Due to (15), the absolute value of the first term on the right hand side of this equality can be made smaller than $\eta/4$ for all sufficiently small ε. Therefore, it remains to estimate the two infinite sums.

Let us start with the second sum. By (18), we can find ε_0, such that for all $\varepsilon < \varepsilon_0$ we have

$$\sup_{x\in\overline{\gamma}}\sup_{\sigma'\le\sigma}|\mathbb{E}_x[f(h(X^\varepsilon_{\sigma'}))-f(h(X^\varepsilon_0))-\int_0^{\sigma'}\mathcal{L}f(h(X^\varepsilon_s))ds]|\le\frac{\eta\varepsilon^\alpha}{4K}.$$

Therefore, by (20) and due to the Markov property of the process, for $\varepsilon < \varepsilon_0$ we have

$$|\sum_{n=0}^\infty \mathbb{E}_x(\chi_{\{\tau_n<T\}}[f(h(X^\varepsilon_{\sigma_{n+1}\wedge T}))-f(h(X^\varepsilon_{\tau_n}))-\int_{\tau_n}^{\sigma_{n+1}\wedge T}\mathcal{L}f(h(X^\varepsilon_s))ds])|\le$$

$$\sup_{x\in\overline{\gamma}}\sup_{\sigma'\le\sigma}|\mathbb{E}_x[f(h(X^\varepsilon_{\sigma'}))-f(h(X^\varepsilon_0))-\int_0^{\sigma'}\mathcal{L}f(h(X^\varepsilon_s))ds]|\sum_{n=0}^\infty\mathbb{E}_x\chi_{\{\tau_n<T\}}\le\frac{\eta}{4}\,. \qquad (22)$$

It remains to estimate the first sum in the right hand side of (21). Fix $\rho > 0$, to be specified later. We introduce the stopping times $\overline{\sigma}_n$, where $\overline{\sigma}_0 = \sigma_0$ and $\overline{\sigma}_n$, $0 \le n \le [T/\rho]$, is the first of the stopping times σ_k which exceeds $\overline{\sigma}_{n-1}$ by at least ρ. We wish to replace the sum by

$$\sum_{n=0}^{[T/\rho]}\mathbb{E}_x(\chi_{\{\overline{\sigma}_n<T\}}\mathbb{E}_{X^\varepsilon_{\overline{\sigma}_n}}\sum_{k=0}^\infty\chi_{\{\sigma_k<\rho\}}[f(h(X^\varepsilon_{\tau_k}))-f(h(X^\varepsilon_{\sigma_k}))-\int_{\sigma_k}^{\tau_k}\mathcal{L}f(h(X^\varepsilon_s))ds]). \qquad (23)$$

Indeed, by Lemma 2.4 and the Markov property of the process, we can replace $\tau_n \wedge T$ by τ_n everywhere in first sum in the right hand side of (21), and the difference will be smaller than $\eta/4$ if ε is sufficiently small. The difference between the resulting expression and the one in (23) is estimated using the Markov property by

$$\sup_{x\in\gamma}\sup_{\alpha\le\rho}|\mathbb{E}_x\sum_{k=0}^\infty\chi_{\{\sigma_k<\alpha\}}[f(h(X^\varepsilon_{\tau_k}))-f(h(X^\varepsilon_{\sigma_k}))-\int_{\sigma_k}^{\tau_k}\mathcal{L}f(h(X^\varepsilon_s))ds]|, \qquad (24)$$

where the second supremum is taken over stopping times $\alpha \le \rho$. In order to estimate the expressions in (23) and (24), we will need the following lemma, whose proof is provided below.

Lemma 2.7. *For each $f \in \mathcal{D}$ and $\delta > 0$ there is $\rho > 0$ such that*

$$\sup_{x\in\gamma}\sup_{\alpha\le\rho}|\mathbb{E}_x\sum_{n=0}^\infty\chi_{\{\sigma_n<\alpha\}}[f(h(X^\varepsilon_{\tau_n}))-f(h(X^\varepsilon_{\sigma_n}))-\int_{\sigma_n}^{\tau_n}\mathcal{L}f(h(X^\varepsilon_s))ds]|\le\delta\rho \qquad (25)$$

for all sufficiently small ε, where the second supremum is taken over stopping times $\alpha \le \rho$.

If we choose $\delta = \eta/(4(T+1))$ and take $\rho \in (0,1)$ such that (25) holds, then the absolute value of the expression in (23) and the expression in (24) are estimated by $\eta/4$. This shows that the right hand side of (21) is estimated by η, as required. $\square$

Proof of Lemma 2.7. We'll divide the proof into several steps.

(a) Consider the process $\widetilde{X}_t^\varepsilon$ obtained from X_t^ε by setting $\beta = 0$ in the right hand side of (7). Let us show that for each $\rho > 0$,

$$\lim_{\varepsilon \downarrow 0} \sup_{x \in \gamma} \sup_{\alpha \leq \rho} |\mathbb{E}_x \sum_{n=0}^{\infty} \chi_{\{\sigma_n < \alpha\}} [f(h(\widetilde{X}^\varepsilon_{\tau_n})) - f(h(\widetilde{X}^\varepsilon_{\sigma_n})) - \int_{\sigma_n}^{\tau_n} \mathcal{L} f(h(\widetilde{X}^\varepsilon_s)) ds]| = 0. \tag{26}$$

Indeed, Lemma 2.1 holds for the process $\widetilde{X}_t^\varepsilon$, as shown in [7]. Moreover, the same proof shows that (13) remains valid if T is replaced by a stopping time $\alpha \leq T$ and the convergence is uniform in α if T is fixed. Therefore,

$$\lim_{\varepsilon \downarrow 0} \sup_{x \in \gamma} \sup_{\alpha \leq \rho} |\mathbb{E}_x [f(h(\widetilde{X}^\varepsilon_{\alpha})) - f(h(\widetilde{X}^\varepsilon_0)) - \int_0^{\alpha} \mathcal{L} f(h(\widetilde{X}^\varepsilon_s)) ds]| = 0. \tag{27}$$

The expectation in (27) can be represented as a sum of terms corresponding to the intervals $[\sigma_n, \tau_n]$ and intervals $[\tau_n, \sigma_{n+1}]$, as above. The contribution from the intervals of the second type tends to zero, as in (22). Therefore the contribution from the intervals of the first type also tends to zero. Notice that the expectation in (26) is equal to the contribution from the intervals of the first type (up to a term found in (16)), thus proving (26).

(b) Given $\rho > 0$, let μ_x^ε be the measure on $\mathcal{C} = C([0, 2\rho], \mathbb{R}^2)$ induced by the process X_t^ε starting at x and $\widetilde{\mu}_x^\varepsilon$ be the measure on $\mathcal{C}$ induced by the process $\widetilde{X}_t^\varepsilon$ starting at x. Let p_x^ε be the density of μ_x^ε with respect to $\widetilde{\mu}_x^\varepsilon$. By the Girsanov theorem, for each $c > 0$ there is $\rho_0 > 0$ such that for $\rho \leq \rho_0$ we have

$$\widetilde{\mu}_x^\varepsilon (1 - c\delta \leq p_x^\varepsilon \leq 1 + c\delta) \geq 1 - \rho^2 \tag{28}$$

for all sufficiently small ε and all $x \in \gamma$. Let $\mathcal{C}' \subseteq \mathcal{C}$ be the event where $p_x^\varepsilon \notin [1 - c\delta, 1 + c\delta]$ and $\Omega' \subseteq \Omega$ be the event that $(X_t^\varepsilon, \, t \in [0, 2\rho]) \in \mathcal{C}'$.

(c) Note that by the Markov property of the process and Lemma 2.4, we can replace the stopping times τ_n in (25) by $\tau_n' = \min(\tau_n, 2\rho)$.

(d) For $0 < \rho < 1$, we can take the same sum as in (20), but starting with $n = [\varepsilon^{-\alpha} \ln(C/\rho)]$ instead of $n = 0$. We then obtain that for each $\delta > 0$ there is a sufficiently large $C > 0$ that does not depend on ρ such that

$$\sum_{n=[\varepsilon^{-\alpha} \ln(C/\rho)]}^{\infty} \mathbb{P}_x(\sigma_n < \rho) \leq \sum_{n=[\varepsilon^{-\alpha} \ln(C/\rho)]}^{\infty} e^{\rho} (1 - r\varepsilon^{\alpha})^{n-1} \leq \delta \rho \varepsilon^{-\alpha}.$$

Therefore, if $\delta > 0$, Ω' is the event constructed above and ρ is sufficiently small (ρ may depend on δ now), then

$$\sum_{n=0}^{\infty} \mathbb{P}_x(\Omega' \cap \{\sigma_n < \alpha\}) \leq \sum_{n=0}^{\infty} \mathbb{P}_x(\Omega' \cap \{\sigma_n < \rho\}) \leq$$

$$\sum_{n=0}^{[\varepsilon^{-\alpha}\ln(C/\rho)]-1} \mathbb{P}_x(\Omega') + \sum_{n=[\varepsilon^{-\alpha}\ln(C/\rho)]}^{\infty} \mathbb{P}_x(\sigma_n < \rho) \tag{29}$$
$$\leq \varepsilon^{-\alpha}\ln(C/\rho)\rho^2 + \delta\rho\varepsilon^{-\alpha} \leq 2\delta\rho\varepsilon^{-\alpha}.$$

(e) Let us show that we can replace $\chi_{\{\sigma_n<\alpha\}}$ in (25) by $\chi_{\{\sigma_n<\alpha\}\setminus\Omega'}$. Indeed,

$$|\mathbb{E}_x \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}\cap\Omega'}[\int_{\sigma_n}^{\tau'_n} \mathcal{L}f(h(X_s^\varepsilon))ds]| \leq 2\rho \sup|\mathcal{L}f|\mathbb{P}_x(\Omega').$$

For arbitrary $\delta > 0$, this can be made smaller than $\delta\rho$ for all sufficiently small ε by taking a sufficiently small ρ. Also,

$$|\mathbb{E}_x \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}\cap\Omega'}[f(h(X_{\tau'_n}^\varepsilon)) - f(h(X_{\sigma_n}^\varepsilon))]| \leq 2|f'(0)|\varepsilon^\alpha \sum_{n=0}^{\infty} \mathbb{P}_x(\Omega' \cap \{\sigma_n < \alpha\}),$$

which can be made smaller than $\delta\rho$ for all sufficiently small ε by taking a sufficiently small ρ due to (29).

We have thus demonstrated that the expectation in the left hand side of (25) can be approximated (with the accuracy of $\delta\rho$ with arbitrarily small δ) by

$$\mathbb{E}_x \left(\chi_{\Omega\setminus\Omega'} \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}}[f(h(X_{\tau'_n}^\varepsilon)) - f(h(X_{\sigma_n}^\varepsilon)) - \int_{\sigma_n}^{\tau'_n} \mathcal{L}f(h(X_s^\varepsilon))ds] \right). \tag{30}$$

(f) We claim that there is a constant K such that

$$\sup_{x\in\gamma}\sup_{\alpha\leq\rho} \mathbb{E}_x| \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}}[f(h(\widetilde{X}_{\tau'_n}^\varepsilon)) - f(h(\widetilde{X}_{\sigma_n}^\varepsilon)) - \int_{\sigma_n}^{\tau'_n} \mathcal{L}f(h(\widetilde{X}_s^\varepsilon))ds]| \leq K\rho \tag{31}$$

for all sufficiently small ε. Indeed,

$$\mathbb{E}_x| \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}}[\int_{\sigma_n}^{\tau'_n} \mathcal{L}f(h(\widetilde{X}_s^\varepsilon))ds]| \leq 2\rho\sup|\mathcal{L}f|,$$

while

$$\mathbb{E}_x| \sum_{n=0}^{\infty} \chi_{\{\sigma_n<\alpha\}}[f(h(\widetilde{X}_{\tau'_n}^\varepsilon)) - f(h(\widetilde{X}_{\sigma_n}^\varepsilon))]| \leq 2|f'(0)|\varepsilon^\alpha \sum_{n=0}^{\infty} \mathbb{P}_x(\sigma_n < \alpha) \leq \overline{K}\rho,$$

for some $\overline{K}$, where the last inequality is due to (20).

(g) Let F be the functional on $\mathcal{C} = C([0,2\rho],\mathbb{R}^2)$ corresponding to the sum in (30). Thus the expectation in (30) can be written as

$$\int_{\mathcal{C}\setminus\mathcal{C}'} F d\mu_x^\varepsilon = \int_{\mathcal{C}\setminus\mathcal{C}'} F p_x^\varepsilon d\widetilde{\mu}_x^\varepsilon = \int_{\mathcal{C}\setminus\mathcal{C}'} F d\widetilde{\mu}_x^\varepsilon + \int_{\mathcal{C}\setminus\mathcal{C}'} F(p_x^\varepsilon - 1)d\widetilde{\mu}_x^\varepsilon. \tag{32}$$

The first integral on the right hand side can be made smaller than $\delta\rho$ for all sufficiently small ε. Indeed, the arguments in steps (c), (d) and (e) can be applied to the process $\widetilde{X}_t^\varepsilon$, and therefore due to (26) the expression in (30) with X_t^ε replaced by $\widetilde{X}_t^\varepsilon$ can be made smaller than $\delta\rho$.

Finally, the second integral on the right hand side of (32) can be estimates as follows

$$|\int_{\mathcal{C}\setminus\mathcal{C}'} F(p_x^\varepsilon - 1) d\widetilde{\mu}_x^\varepsilon| \leq c\delta \int_{\mathcal{C}\setminus\mathcal{C}'} |F| d\widetilde{\mu}_x^\varepsilon \leq cK\delta\rho,$$

where the first inequality follows from the definition of $\mathcal{C}'$ and the second one from (31). It remains to take a sufficiently small constant c in (28). □

3 Averaging principle for deterministic perturbations

Recall that the process $X_t^{\varkappa,\varepsilon}$ is defined in (5), which is different from (7) in that now the terms $u(X_t^\varepsilon)dt + \sigma(X_t^\varepsilon)dW_t$ in the right hand side are replaced by $\varkappa u(X_t^\varepsilon)dt + \sqrt{\varkappa}\sigma(X_t^\varepsilon)dW_t$, where $\varkappa > 0$ is a small parameter.

Let $Y_t^{\varkappa,\varepsilon} = h(X_t^{\varkappa,\varepsilon})$ be the corresponding process on the graph $\mathbb{G}$. In Section 2 we demonstrated that the distribution of $Y_t^{\varkappa,\varepsilon}$ converges, as $\varepsilon \downarrow 0$, to the distribution of a limiting process, which will be denoted by $Z_t^\varkappa$. In this section we show that the distribution of $Z_t^\varkappa$, in turn, converges to the distribution of a limiting Markov process on $\mathbb{G}$ when $\varkappa \downarrow 0$.

We need additional notations in order to describe the limiting distribution of $Z_t^\varkappa$. Let

$$\overline{\varphi}_k = \int_{\gamma_k} \frac{\langle \alpha \nabla H, \nabla H\rangle}{|\nabla H|} dl,$$

$$\overline{\psi}_k = 2\int_{\gamma_k} \frac{\langle \beta, \nabla H\rangle}{|\nabla H|} dl.$$

Let us recall that $Z_0^\varkappa$ is distributed as $h(X_0^{\varkappa,\varepsilon})$ (we assume that $X_0^{\varkappa,\varepsilon}$ does not depend on ε). Denote the generator of $Z_t^\varkappa$ by $\mathcal{L}^\varkappa$. Recall that $\mathcal{L}^\varkappa$ can be described as follows.

Let $L_k^\varkappa f(h_k) = a_k^\varkappa(h_k) f''(h_k) + b_k^\varkappa(h_k) f'(h_k)$ be the differential operator on the interior of I_k with the coefficients

$$a_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \int_{\gamma_k(h_k)} \frac{\langle \varkappa\alpha\nabla H, \nabla H\rangle}{|\nabla H|} dl \quad \text{and}$$

$$b_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \int_{\gamma_k(h_k)} \frac{2\langle \beta + \varkappa u, \nabla H\rangle + \varkappa\alpha \cdot H''}{|\nabla H|} dl,$$

where

$$T_k(h_k) = \int_{\gamma_k(h_k)} \frac{1}{|\nabla H|} dl$$

is the period of the unperturbed system. Note that

$$a_k^{\varkappa}(h_k) = \frac{1}{2}(T_k(h_k))^{-1}\varkappa\overline{\varphi}_k(1+o(1)), \quad |h_k| \downarrow 0,$$

$$b_k^{\varkappa}(h_k) = \frac{1}{2}(T_k(h_k))^{-1}(\overline{\psi}_k(1+o(1)) + \varkappa O(\ln(|h_k|))), \quad |h_k| \downarrow 0,$$

and therefore

$$\frac{b_k^{\varkappa}(h_k)}{a_k^{\varkappa}(h_k)} = \frac{\overline{\psi}_k}{\varkappa\overline{\varphi}_k}(1+o(1)) + O(\ln(|h_k|))), \quad |h_k| \downarrow 0. \tag{33}$$

The domain of $\mathcal{L}^{\varkappa}$ consists of those functions $f \in C(\mathbb{G})$ which

(a) Are twice continuously differentiable in the interior of each of the edges;

(b) Have the limits $\lim_{h_k \to 0} L_k^{\varkappa} f(h_k)$ and $\lim_{h_k \to (H(M_k)-H(A_k))} L_k^{\varkappa} f(h_k)$ at the endpoints of each of the edges. Moreover, the value of the limit $q^{\varkappa} = \lim_{h_k \to 0} L_k^{\varkappa} f(h_k)$ is the same for all edges;

(c) Have the limits $\lim_{h_k \to 0} f'(h_k)$, and

$$\varkappa \sum_{k=1}^{n} p_k \lim_{h_k \to 0} f'(h_k) = q^{\varkappa}, \tag{34}$$

where p_k are given by (10).

For functions f which satisfy the above three properties, we define $\mathcal{L}^{\varkappa} f = L_k^{\varkappa} f$ in the interior of each edge, and as the limit of $L_k^{\varkappa} f$ at the endpoints of I_k.

We assume that $\overline{\psi}_k \neq 0$. Let s_k, $1 \leq k \leq n$, take values zero and one. We set $s_k = 1$ if $\overline{\psi}_k > 0$ and M_k is a local maximum of H as well as if $\overline{\psi}_k < 0$ and M_k is a local minimum of H. Otherwise, we set $s_k = 0$. Let

$$r_k = \left| \frac{s_k p_k \overline{\psi}_k}{\overline{\varphi}_k} \right| = \frac{s_k |\overline{\psi}_k|}{2\text{Area}(\mathcal{E})}, \quad 1 \leq k \leq n.$$

Note that r_k do not depend on α. Let Z_t be the family of processes (that depend on the initial point) on the state space $\mathbb{G}$ whose distribution is determined by the following conditions:

(a) Z_t is a strong Markov family with continuous trajectories;

(b) If $Z_0 = O$, where O is the root of $\mathbb{G}$, then the process spends a random time τ in O. There is a random variable ξ that is independent of τ, takes values in the set $\{1, ..., n\}$, and is such that $Z_t \in I_\xi$ for $t > \tau$.

If $s_k = 0$ for $1 \leq k \leq n$, then $\tau = \infty$. If $s_k = 1$ for some k, then τ is exponentially distributed with the parameter

$$\mu = \sum_{k=1}^{n} r_k. \tag{35}$$

If $s_k = 1$ for some k, then

$$\mathbb{P}(Z_t \in I_k \text{ for } t > \tau) = r_k \left(\sum_{i=1}^{n} r_i \right)^{-1}.$$

(c) If $Z_0 \in \text{int}(I_k)$, then $dZ_t/dt = \overline{b}_k(Z_t)$ for $t < \sigma$, where $\sigma = \inf(t : Z_t = O)$ and

$$\overline{b}_k(h_k) = b_k^0(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \int_{\gamma_k(h_k)} \frac{2\langle \beta, \nabla H \rangle}{|\nabla H|} dl.$$

Thus Z_t moves deterministically along the edge I_k of the graph with the speed $\overline{b}_k(h_k)$. If the process reaches O in finite time (in which case $s_k = 0$), then it either stays at O (if $s_m = 0$, $1 \le m \le n$) or spends exponential time in O and then continues with deterministic motion away from O along a randomly selected edge (if $s_m = 1$ for some m).

Theorem 2. *The measure on* $C([0, \infty), \mathbb{G})$ *induced by the process* $Z_t^\varkappa$ *converges weakly to the measure induced by the process* Z_t *with the initial distribution* $h(X_0^\varepsilon)$.

We would like to underline again that the process Z_t is defined by the deterministic system (2). The stochastic perturbations are used just for regularization purposes.

The motion of $Z_t^\varkappa$ inside each edge can be understood by standard perturbation theory. Namely, let

$$0 \le \delta < \min_{1 \le k \le n} |H(A_k) - H(M)|, \quad \sigma^\varkappa(\delta) = \inf(t : |Z_t^\varkappa| = \delta), \quad \sigma(\delta) = \inf(t : |Z_t| = \delta).$$

Using the fact that for small $\varkappa$ we have a small perturbation of the deterministic system $dZ_t/dt = \overline{b}_k(Z_t)$, one can easily obtain the following statements:

For the processes $Z_t^\varkappa$ and Z_t starting on the edge I_k with $|Z_0^\varkappa| = |Z_0| = \delta$,

$$\text{if } \sigma(0) < \infty, \text{ then } \lim_{\varkappa \downarrow 0} (\sigma^\varkappa(0) - \sigma(0)) = 0 \text{ in probability}$$

and for each $T < \infty$,

$$\lim_{\varkappa \downarrow 0} \Big(\max_{t \le \min(T, \sigma^\varkappa(0))} |Z_t^\varkappa - Z_t| \Big) = 0 \text{ in probability.}$$

From here it easily follows that for the process $Z_t^\varkappa$ starting at O

$$\lim_{\varkappa \downarrow 0} \sigma^\varkappa(\delta) = \infty \quad \text{in probability}$$

if $r_k = 0$ for all k. It remains to describe the behavior of the process $Z_t^\varkappa$starting at O till the time it exits a small neighborhood of O in the case when $r_k \neq 0$ for some k. Thus Theorem 2 will follow from the two lemmas below.

Lemma 3.1. *If $Z_0^\varkappa = O$ and $r_k \neq 0$ for some k, then*

$$\lim_{\varkappa\downarrow 0} \mathbb{P}(Z^\varkappa_{\sigma^\varkappa(\delta)} \in I_k) = r_k \left(\sum_{i=1}^n r_i\right)^{-1}.$$

Proof. Let $\mathbb{G}^\delta = \{(i,h_i) \in \mathbb{G} : |h_i| \leq \delta\}$. Let $f_{k,\varkappa}(h)$, $h \in \mathbb{G}^\delta$, be the probability that the process $Z_t^\varkappa$ starting at h exits $\mathbb{G}^\delta$ through the point that belongs to I_k. Thus $f_{k,\varkappa}$ is a continuous function on $\mathbb{G}^\delta$, is twice continuously differentiable for $|h_i| \in (0,\delta)$ and is such that

(a) $L_i^\varkappa f_{k,\varkappa}(h_i) = 0$ for $|h_i| \in (0,\delta)$, $1 \leq i \leq n$;

(b) The limits $\lim_{h_i \to 0} f'_{k,\varkappa}(h_i)$ exist and (34) holds with $f_{k,\varkappa}$ instead of f and $q^\varkappa = 0$.

(c) $f_{k,\varkappa}(h_k) = 1$ for $|h_k| = \delta$; $f_{k,\varkappa}(h_i) = 0$ for $|h_i| = \delta$ if $i \neq k$.

Note that we are interested in the limit $\lim_{\varkappa\downarrow 0} f_{k,\varkappa}(O)$. Assuming that $f_{k,\varkappa}(O)$ is known, we can use the differential relation (a) to find $f_{i,\varkappa}(h_i)$, $0 \leq |h_i| \leq \delta$, $1 \leq i \leq n$. Namely,

$$f_{i,\varkappa}(h_i) = f_{k,\varkappa}(O) + c_{i,\varkappa} \int_0^{h_i} \exp\left(-\int_0^s \frac{b_i^\varkappa(u)}{a_i^\varkappa(u)} du\right) ds.$$

The constants $c_{i,\varkappa}$ can be found from the boundary condition (c) and are equal to

$$c_{k,\varkappa} = \frac{1 - f_{k,\varkappa}(O)}{I_k(\varkappa)}; \quad c_{i,\varkappa} = \frac{-f_{k,\varkappa}(O)}{I_i(\varkappa)}, \quad i \neq k,$$

where

$$I_i(\varkappa) = \int_0^\delta \exp\left(-\int_0^s \frac{b_i^\varkappa(u)}{a_i^\varkappa(u)} du\right) ds. \tag{36}$$

From (b) we find that $\sum_{i=1}^n p_i c_{i,\varkappa} = 0$, and therefore

$$f_{k,\varkappa}(O) = \frac{p_k I_k^{-1}(\varkappa)}{\sum_{i=1}^n p_i I_i^{-1}(\varkappa)}. \tag{37}$$

From (33) it easily follows that

$$I_i(\varkappa) = (\varkappa + o(\varkappa)) \frac{\overline{\varphi}_i}{\overline{\psi}_i} \quad \text{when} \quad \varkappa \downarrow 0 \quad \text{if} \quad s_i = 1; \tag{38}$$

$$I_i(\varkappa) \to \infty \quad \text{when} \quad \varkappa \downarrow 0 \quad \text{if} \quad s_i = 0. \tag{39}$$

Substituting this into (37), we obtain the desired result. □

The next lemma shows that the distribution of the time spent by the process in a small neighborhood of O is asymptotically exponential with parameter μ and that this time is asymptotically independent of which edge it chooses upon exiting from O.

Lemma 3.2. *Let $\lambda \geq 0$, $Z_0^\varkappa = O$ and $r_k \neq 0$ for some k. Let A_m denote the event that $Z^\kappa_{\sigma^\varkappa(\delta)} \in I_m$. Then*

$$\mathbb{E}\left(\chi_{A_m} \exp(-\lambda \sigma^\varkappa(\delta))\right) = \frac{r_m}{\mu+\lambda}(1+\xi_m(\lambda,\delta,\varkappa)) + \frac{\lambda \eta_m(\lambda,\delta,\varkappa)}{\mu+\lambda}, \tag{40}$$

where μ is defined in (35), $\lim_{\varkappa\downarrow 0} \xi_m(\lambda,\delta,\varkappa)) = 0$ uniformly in $\lambda \geq 0, \delta < \delta_0$ for some positive δ_0 and $\lim_{\delta \downarrow 0} \eta_m(\lambda,\delta,\varkappa) = 0$ uniformly in $\lambda \geq 0, \varkappa < \varkappa_0$ for some positive $\varkappa_0$.

In particular

$$\mathbb{E} \exp(-\lambda \sigma^\varkappa(\delta)) = \frac{\mu}{\mu+\lambda}(1+\xi(\lambda,\delta,\varkappa)) + \frac{\lambda \eta(\lambda,\delta,\varkappa)}{\mu+\lambda}, \tag{41}$$

where ξ and η have the same properties as ξ_m and η_m.

Proof. Let us prove (40). Let $f_\varkappa(h)$, $h \in \mathbb{G}^\delta$, be equal to the expectation in the left hand side of (40), where the stopping time $\sigma^\varkappa(\delta)$ is that of the process starting at h instead of O. Then $f_\varkappa$ is a continuous function on $\mathbb{G}^\delta$, is twice continuously differentiable for $|h_k| \in (0,\delta)$ and is such that

(a) $L_k^\varkappa f_\varkappa(h_k) - \lambda f_\varkappa(h_k) = 0$ for $|h_k| \in (0,\delta)$, $1 \leq k \leq n$;

(b) The limits $\lim_{h_k \to 0} f_\varkappa'(h_k)$ exist and (34) holds with $f_\varkappa$ instead of f and $q^\varkappa$ replaced by $\lambda f_\varkappa(O)$.

(c) $f_\varkappa(h_m) = 1$ for $|h_m| = \delta$, and $f_\varkappa(h_k) = 1$ for $|h_k| = \delta$, $k \neq m$.

Note that we are interested in the asymptotics of $f_\varkappa(O)$ as $\varkappa \downarrow 0$. Let us temporarily treat $\lambda f_\varkappa$ as a known function, which we denote by $g_\varkappa$. Note that $g_\varkappa$ is continuous and $|g_\varkappa|$ is bounded by λ. Then $f_\varkappa(O) = g_\varkappa(O)/\lambda$. From this and the differential relation (a) we can find $f'_\varkappa(h_k)$, $0 \leq |h_k| \leq \delta$, $1 \leq k \leq n$. Namely,

$$f'_\varkappa(h_k) = \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds + c_{k,\varkappa} \right) \exp\left(- \int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)} ds \right), \tag{42}$$

where $c_{k,\varkappa}$ are constants. From (b) it follows that

$$\sum_{k=1}^n p_k c_{k,\varkappa} = g_\varkappa(O)/\varkappa. \tag{43}$$

Upon integrating (42) from 0 to δ and using (c), we obtain

$$f_\varkappa(O) = \boldsymbol{\delta}_{km} - \int_0^\delta \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds + c_{k,\varkappa} \right) \exp\left(- \int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)} ds \right) dh_k =$$

$$\boldsymbol{\delta}_{km} - \int_0^\delta \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds \right) \exp\left(- \int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)} ds \right) dh_k - c_{k,\varkappa} I_k(\varkappa) =$$

$$\boldsymbol{\delta}_{km} - J_k(\delta,\lambda,\varkappa) - c_{k,\varkappa} I_k(\varkappa),$$

where

$$J_k(\delta,\lambda,\varkappa) = \int_0^\delta \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds \right) \exp\left(- \int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)} ds \right) dh_k, \quad (44)$$

$I_k(\varkappa)$ was defined in (36) and $\boldsymbol{\delta}_{km}$ equals 1 if $k = m$ and 0 otherwise. Let us multiply both sides of this equality by $p_k / I_k(\varkappa)$ and take the sum in k. Upon using (43), we obtain

$$f_\varkappa(O) \sum_{k=1}^n \frac{p_k}{I_k(\varkappa)} = \frac{p_m}{I_m(\varkappa)} - \sum_{k=1}^n \frac{p_k J_k(\delta,\lambda,\varkappa)}{I_k(\varkappa)} - \frac{\lambda f_\varkappa(O)}{\varkappa}.$$

This is a linear equation on $f_\varkappa(O)$. Solving it, we obtain

$$f_\varkappa(O) = \left(\sum_{k=1}^n p_k I_k(\varkappa) + \frac{\lambda}{\varkappa} \right)^{-1} \frac{p_m}{I_m(\varkappa)} + \left(\sum_{k=1}^n \frac{p_k}{I_k(\varkappa)} + \frac{\lambda}{\varkappa} \right)^{-1} \sum_{k=1}^n \frac{p_k J_k(\delta,\lambda,\varkappa)}{I_k(\varkappa)}.$$

By (38) and (39), the first term on the right hand side converges, as $\varkappa \downarrow 0$, to $r_m/(\mu+\lambda)$. It remains to show that

$$\lim_{\delta \downarrow 0} \frac{\varkappa J_k(\delta,\lambda,\varkappa)}{I_k(\varkappa)} = 0$$

for each k. When k is such that $s_k = 0$, we use the fact that by (33)

$$\int_0^{h_k} \frac{g_\varkappa(s)}{\lambda} \frac{\varkappa}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds$$

converges to zero when $\delta \downarrow 0$, while the second factor inside the integral in (44) is the same as the integrand in the definition of $I_k(\varkappa)$. When $s_k = 1$, we rewrite J_k as follows

$$J_k(\delta,\lambda,\varkappa) = \int_0^\delta \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(- \int_s^{h_k} \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)} du \right) ds \right)$$

and again use (33) to show that $J_k(\delta,\lambda,\varkappa)$ tends to zero when $\delta \downarrow 0$. This proves (40). Now (41) follows from (40) by summing over m. $\square$

Finally, as it was already mentioned, the case where some of the periodic components contains saddles could be treated using the analysis of [4]. Namely the limit process is still Markov. Upon reaching a vertex corresponding to a saddle point the process instantly chooses one of the edges where the averaged field points inside the edge and the probability to choose the edge k is proportional to $|\bar{\psi}_k|$.

Acknowledgements: While working on this article, D. Dolgopyat was supported by NSF grant DMS-0854982, M. Freidlin was supported by NSF grants DMS-0803287 and DMS-0854982 and L. Koralov was supported by NSF grants DMS-0706974 and DMS-0854982.

References

[1] Arnold V. I. *Topological and ergodic properties of closed* 1*-forms with incommensurable periods,* Funct. Anal. Appl. **25** (1991), no. 2, 81–90.

[2] Arnold V. I. *Mathematical Methods of Classical Mechanics*, Springer (1989).

[3] Berestycki H., Hamel F., Nadirashvili N., *Elliptic eigenvalue problems with large drift and applications to nonlinear propagation phenomena*, Comm. Math. Physics 253 (2005) 451-480.

[4] Brin M., Freidlin M. *On stochastic behavior of perturbed Hamiltonian systems*, Ergodic Theory and Dynamical Systems (2000), **20**, 55-76.

[5] Constantin P., Kisilev A., Ryzhik L., Zlatos A., *Diffusion and mixing in fluid flow*, Annals of Math. 168 (2008), 643-674.

[6] Dolgopyat D., Koralov L. *Averaging of Hamiltonian Flows with an Ergodic Component* The Annals of Probability, 2008, Vol 36, No 6, 1999-2049.

[7] Dolgopyat D., Koralov L. *Averaging of incompressible flows on two-dimensional surfaces* preprint.

[8] Franke B., *Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence free drift*, Mathematiche Zeitschrift, 246, 373-403 (2004).

[9] Freidlin M. I. *Markov Processes and Differential Equations: Asymptotic Problems* Lectures in Mathematics, ETH Zurich. Birkhauser, Basel, 1996.

[10] Freidlin M. I., Wentzell A. D. *Diffusion processes on graphs and averaging principle*, Annals of Probability, 21, 4, pp 2215-2245.

[11] Freidlin M. I., Wentzell A. D. *Random Perturbations of Dynamical Systems,* 2-nd edition, Springer, 1998.

[12] Khanin K. M., Sinai Ya. G. *Mixing of some classes of special flows over rotations of the circle,* Funct. Anal. Appl. **26** (1992), no. 3, 155–169.

[13] Hasminskii R. Z. *On the principle of averaging the It's stochastic differential equations*. (Russian) Kybernetika (Prague) 4 1968 260–279.

[14] Koralov L. *Random Perturbations of 2-dimensional Hamiltonian Flows*, Probab. Theory and Related Fields. **129**, (2004), 37-62.

[15] Mandl P. *Analytical Treatment of One-Dimensional Markov Provesses*, Springer-Verlag, 1968.

[16] Sowers R., *Random Perturbations of Two-Dimensional Pseudoperiodic Flows*, Illinois Journal of Mathematics 50 (2006), pp. 853–959.

[17] Zlatos A., *Diffusion in fluid flow: Dissipation enhancement by flows in 2D*, to appear in Comm. Partial Differential Equations.