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 Hamiltonian systems on a two-dimensional torus. Even in the case of purely deterministic perturbations, the long-time behavior of such systems can be stochastic, in a certain sense. The stochasticity is caused by the instabilities near the saddle point of the non-perturbed system as well as by the ergodic component of the Hamiltonian system on the torus.

Key words and phrases: Averaging, Markov Process, Diophantine Condition, 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 invariant distribution on the corresponding level set characterizes the fast motion.

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,

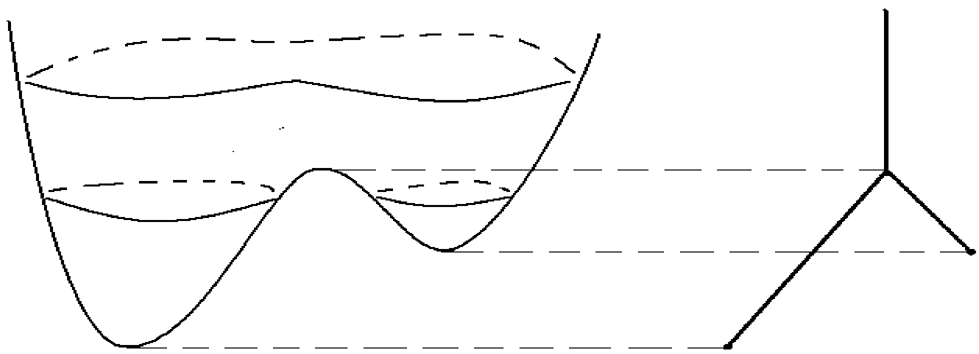

Figure 1: The graph of H and the corresponding graph $\mathbb{G}$.

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.

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

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

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 [6]. 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 [3] for the case when all the level sets of H are compact.

In the current paper we consider a generic incompressible periodic vector field v. In this case some of the level sets of H are unbounded, and the unperturbed flow, when considered on the torus, has an ergodic component. Random perturbations of such flows were considered by Dolgopyat and Koralov in [4] (for generic flows) and by Sowers in [13] (for flows whose stream function is nearly periodic). In this paper, using in particular some of the techniques of [3] and [4], we 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. We shall assume that all the critical points of H are non-degenerate, and that (a, b) satisfy the following Diophantine condition.

Let $\rho = a/b$ be irrational. Without loss of generality we may assume that $0 < a < b$ (the general case can be obtained by interchanging x_1 and x_2, and/or replacing x_i by $-x_i$, if needed). Let $[a_1, a_2 \ldots a_n \ldots]$ be the continued fraction expansion of ρ. We assume that

$$a_n \le n^2 \quad \text{for all sufficiently large } n.$$

It is easy to show that his condition holds for almost all ρ with respect to the Lebesgue measure on $[0, 1]$ (see [10]).

For a and b which are rationally independent, as in our case, it has been shown by Arnold in [1] that the structure of the stream lines of v 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 (see [9])).

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).

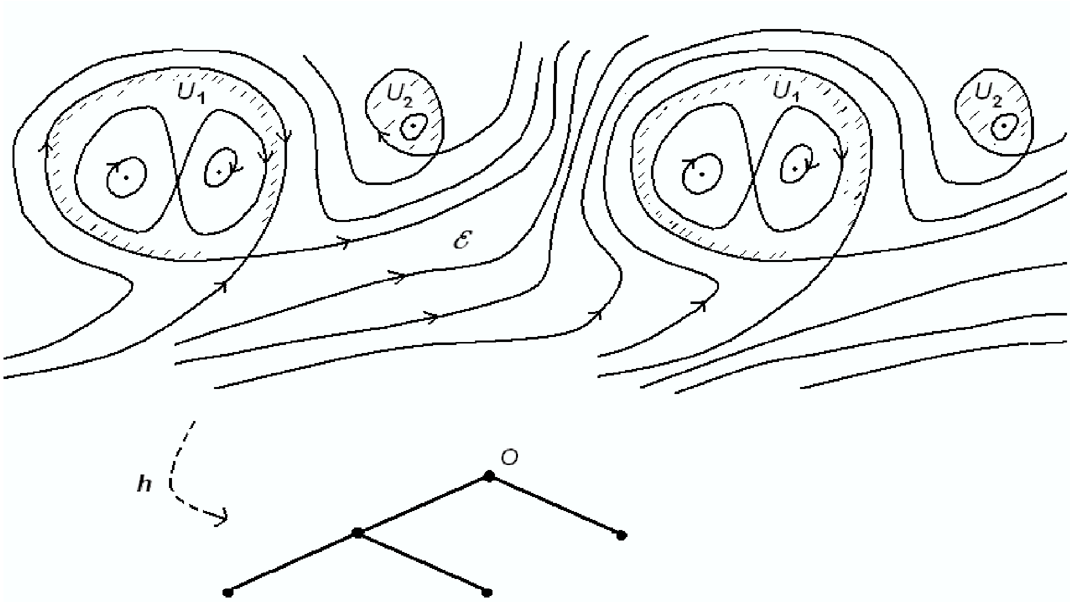

Figure 2: The stream lines of the flow and the corresponding graph

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

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

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 Sections 2 - 4 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 5 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 The case of one periodic component

We assume for brevity that H has no saddle points inside the domains U_k. The general case can be easily considered using the results of this paper and [3].

In this section we shall consider the case when there is just one periodic component U, which contains only one critical point of H (a maximum or a minimum). An example of a phase portrait of such a vector field v (considered on the plane) is given in Figure 3. The general case is discussed in Section 4.

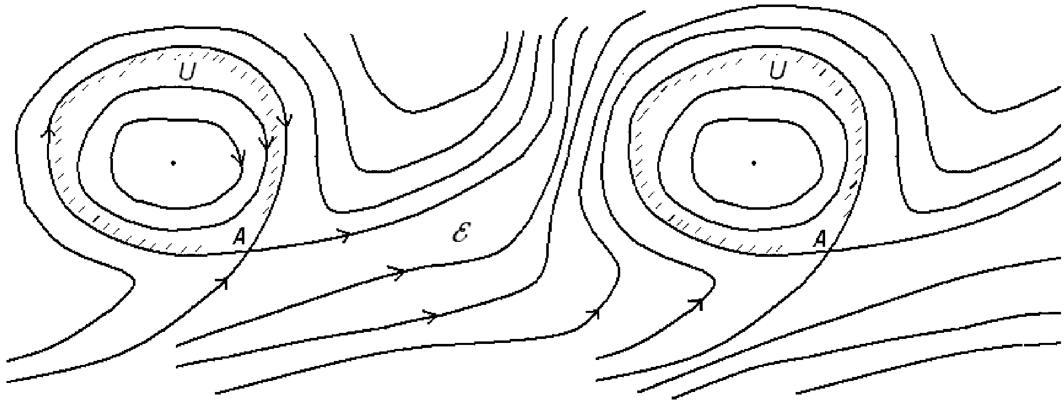

Figure 3

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. We also consider two auxiliary processes $\widehat{X}_t^\varepsilon$ and $\widetilde{X}_t^\varepsilon$ defined by

$$d\widehat{X}_t^\varepsilon = \frac{1}{\varepsilon} v(\widehat{X}_t^\varepsilon)dt + \sigma(\widehat{X}_t^\varepsilon)dW_t \ , \quad \widehat{X}_t^\varepsilon \in \mathbb{T}^2, \tag{8}$$

and

$$d\widetilde{X}_t^\varepsilon = \frac{1}{\varepsilon} v(\widetilde{X}_t^\varepsilon)dt + u(\widetilde{X}_t^\varepsilon)dt + \sigma(\widetilde{X}_t^\varepsilon)dW_t \ , \quad \widetilde{X}_t^\varepsilon \in \mathbb{T}^2. \tag{9}$$

To be specific, assume that the critical point of H inside U is a maximum. We shall denote the saddle point of H on the torus by A and the maximum by M. Consider the following mapping of the torus onto the segment $I = [0, H(M) - H(A)]$ of the real line

$$h(x) = \begin{cases} 0 & \text{if } x \in \mathcal{E} \\ H(x) - H(A) & \text{otherwise.} \end{cases}$$

We denote the set $\{x \in [U] : H(x) - H(A) = h\}$ by $\gamma(h)$. Let $\gamma = \gamma(0) = \partial U$. Let $Lf(h) = a(h)f'' + b(h)f'$ be the differential operator on the interior of I with the coefficients

$$a(h) = \frac{1}{2}\left(\int_{\gamma(h)} \frac{1}{|\nabla H|} dl\right)^{-1} \int_{\gamma(h)} \frac{\langle \alpha \nabla H, \nabla H\rangle}{|\nabla H|} dl \quad \text{and} \tag{10}$$

$$b(h) = \frac{1}{2}\left(\int_{\gamma(h)} \frac{1}{|\nabla H|} dl\right)^{-1} \int_{\gamma(h)} \frac{2\langle \beta + u, \nabla H\rangle + \alpha \cdot H''}{|\nabla H|} dl, \tag{11}$$

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

$$\widetilde{b}(h) = \frac{1}{2}\left(\int_{\gamma(h)} \frac{1}{|\nabla H|} dl\right)^{-1} \int_{\gamma(h)} \frac{2\langle u, \nabla H\rangle + \alpha \cdot H''}{|\nabla H|} dl \tag{12}$$

and

$$p = \frac{1}{2}(\mathrm{Area}(\mathcal{E}))^{-1} \int_\gamma \frac{\langle \alpha \nabla H, \nabla H\rangle}{|\nabla H|} dl = \frac{1}{2}(\mathrm{Area}(\mathcal{E}))^{-1} \left| \int_U \mathrm{div}(\alpha \nabla H)(x) dx \right|. \tag{13}$$

Consider the process Y_t on the closed segment I which is defined via its generator $\mathcal{L}$ as follows. The domain of the generator consists of those functions $f \in C(I)$ which

(a) are twice continuously differentiable in the interior of I,

(b) have limits $\lim_{h\to 0} Lf(h)$ and $\lim_{h \to (H(M)-H(A))} Lf(h)$ at the endpoints of I,

(c) have the limit $\lim_{h\to 0} f'(h)$, and $p \lim_{h\to 0} f'(h) = \lim_{h\to 0} Lf(h)$.

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

It well-known (see [12], for example) that there exists a strong Markov process on I with continuous trajectories, with the generator $\mathcal{L}$. The measure on $C([0,\infty), I)$ 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 measures on on $C([0,\infty), I)$ induced by the processes $Y_t^\varepsilon = h(X_t^\varepsilon)$ with $X_0^\varepsilon = x \in \mathbb{T}^2$ converges weakly, as $\varepsilon \downarrow 0$, to the measure induced by the process with the generator $\mathcal{L}$ with the initial distribution concentrated at $h(x)$.*

Let Ψ be the subset of $C(I)$, which consists of all bounded functions, which are continuously differentiable on $[0, H(M) - H(A))$ (the derivative at $h = 0$ is one-sided). Note that this is a measure defining set, that is the equality $\int_I u d\mu_1 = \int_I u d\mu_2$ for all $u \in \Psi$ implies that $\mu_1 = \mu_2$. Let $\mathcal{D}$ be the subset of $D(\mathcal{L})$, which consists of all the functions f for which $\mathcal{L}f \in \Psi$.

We formulate the following lemma.

Lemma 2.1. *For any function $f \in \mathcal{D}$, any initial point $x \in \mathbb{T}^2$, and any $T > 0$ we have*

$$\mathbb{E}_x[f(h(X_T^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^T \mathcal{L}f(h(X_s^\varepsilon))ds] \to 0 \quad \text{as } \varepsilon \to 0, \tag{14}$$

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

An analogous lemma was used in the monograph of Freidlin and Wentzell [8] to justify the convergence of the process Y_t^ε to the limiting process on the graph. The main idea, roughly speaking, is to use the tightness of the family Y_t^ε, and then to show that the limiting process (along any subsequence), is a solution of the martingale problem, corresponding to the operator $\mathcal{L}$. Here, as in [8], it is used that for every $u \in \Psi$ and $\lambda > 0$ the equation $\lambda f - \mathcal{L}f = u$ has a solution $f \in \mathcal{D}$.

The main difference between our case and that of [8] is the presence of an ergodic component. However, all the arguments used to prove the main theorem based on (14) remain the same. Thus, as follows from Lemma 3.1 of [8], in order to prove Theorem 1 it is enough to prove our Lemma 2.1 above. We mainly pay attention to the differences in the proof that arise due to the presence of the ergodic component.

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 distinguish between the objects which do not depend on ε and those which do.

Let $\overline{\tau}$ be the first time when the process X_t^ε reaches the set $\gamma(\varepsilon^{\frac{1}{2}})$. We shall need the following estimate on the expectation of $\overline{\tau}$, which is proved in Section 3. This is a key lemma and some of the lemmas below rely on this result.

Lemma 2.2. *For any $\varkappa > 0$ there is $\varepsilon_0 > 0$, such that $\mathbb{E}_x \overline{\tau} \le \varepsilon^{\frac{1}{2}-\varkappa}$ for $\varepsilon \le \varepsilon_0$ for all $x \in [\mathcal{E}]$.*

Let us now pick constants $\varkappa$ and α, such that $0 < \varkappa < \frac{1}{4} < \alpha < \frac{1}{2}$. Let $\overline{\gamma} = \gamma(\varepsilon^\alpha)$. Recall that $\gamma = \gamma(0)$ is the boundary of U. Let τ be the first time when the process X_t^ε reaches $\overline{\gamma}$, and σ be the first time when the process reaches γ. We inductively define the following two sequences of stopping times. Let $\sigma_0 = 0$ and $\tau_1 = \tau$. For $n \ge 1$ let σ_n be the first time following τ_n when the process reaches γ. For $n \ge 2$ let τ_n be the first time following σ_{n-1} when the process reaches $\overline{\gamma}$.

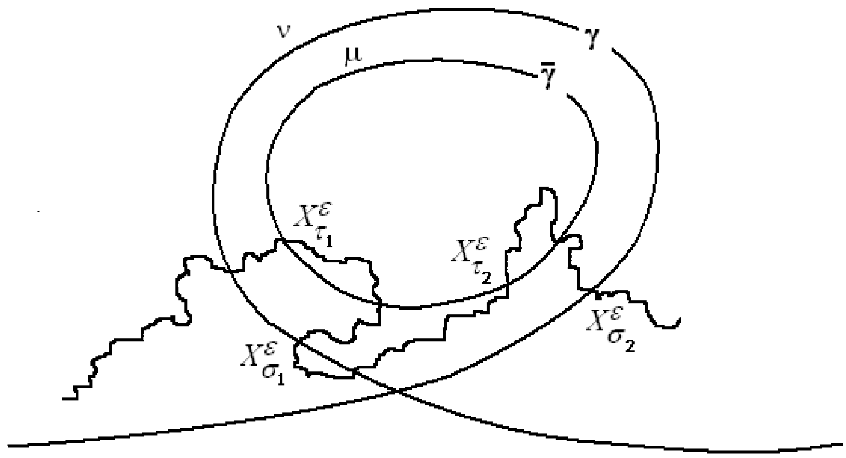

Figure 4

We can consider the following discrete time Markov chains $\xi_n^1 = X^\varepsilon_{\tau_n}$ and $\xi_n^2 = X^\varepsilon_{\sigma_n}$ with the state spaces $\overline{\gamma}$ and γ, respectively. Let $P_1(x, dy)$ and $P_2(x, dy)$ be transition operators for the Markov chains ξ_n^1 and ξ_n^2, respectively. They are uniformly exponentially mixing in the following sense.

Lemma 2.3. *There exist constants $0 < c < 1$, $\varepsilon_0 > 0$, $n_0 > 0$, and probability measures μ and ν (which depend on ε) on $\overline{\gamma}$ and γ, respectively, such that for $\varepsilon < \varepsilon_0$ and $n \ge n_0$ we have*

$$\sup_{x \in \overline{\gamma}} (\mathrm{Var}(P_1^n(x, dy) - \mu(dy))) \le c^n, \qquad \sup_{x \in \gamma} (\mathrm{Var}(P_2^n(x, dy) - \nu(dy))) \le c^n, \tag{15}$$

where Var *is the total variation of the signed measure.*

This lemma was proved in [4] in the case when $\beta = u = 0$ and σ is an identity matrix. The proof goes through without major modifications in the general case, so we do not repeat it here.

Let us now examine the transitions times between $\overline{\gamma}$ and γ, particularly for the case when we start with the invariant measures. The following lemma is proved in Section 3.

Lemma 2.4. *We have the following asymptotic relations for the transition times*

$$\mathbb{E}_\mu \sigma = k_1 \varepsilon^\alpha (1 + o(1)), \quad \text{as} \quad \varepsilon \to 0, \tag{16}$$

$$\sup_{x \in \gamma} \mathbb{E}_x \tau = o(\varepsilon^{\alpha - \varkappa}), \quad \text{as} \quad \varepsilon \to 0, \tag{17}$$

where

$$k_1 = \int_I a^{-1}(t) \exp\left(\int_0^t \frac{b(h)}{a(h)} dh \right) dt. \tag{18}$$

Besides, if $\beta = 0$, then

$$\mathbb{E}_\nu \tau = k_2 \varepsilon^\alpha (1 + o(1)), \quad \text{as} \quad \varepsilon \to 0, \tag{19}$$

where

$$k_2 = \frac{\mathrm{Area}(\mathcal{E})}{\mathrm{Area}(U)} \int_I a^{-1}(t) \exp\left(\int_0^t \frac{\widetilde{b}(h)}{a(h)} dh \right) dt = 2(\mathrm{Area}(\mathcal{E})) \left(\left| \int_U \mathrm{div}(\alpha \nabla H)(x) dx \right| \right)^{-1}, \tag{20}$$

The functions a, b and $\widetilde{b}$ in the expressions for k_1 and k_2 are defined by (10), (11) and (12), respectively.

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 [4].

Lemma 2.5. *There is a constant $r > 0$, such that for all sufficiently small ε we have*

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

Using the Markov property of the process and Lemma 2.5, we get the estimate

$$\sup_{x \in \mathbb{T}^2} \mathbb{E}_x e^{-\sigma_n} \leq \sup_{x \in \overline{\gamma}} \mathbb{E}_x e^{-\sigma_n} \leq (\sup_{x \in \overline{\gamma}} \mathbb{E}_x e^{-\sigma})^n \leq (1 - r\varepsilon^\alpha)^n. \tag{21}$$

The first inequality here follows from the definition of σ_n. The next lemma allows us to estimate integrals of the type (14) over intervals $[0, \tau]$ and $[0, \sigma]$. Let $V^\varepsilon = \{x \in U : 0 \leq h(x) \leq \varepsilon^\alpha\}$ (the region between γ and $\overline{\gamma}$). Let $\mathcal{E}^\varepsilon = \mathcal{E} \cup V^\varepsilon$.

Lemma 2.6. *For any function $f \in \mathcal{D}$ we have the following asymptotic estimates*

$$\sup_{x \in \overline{\gamma}} |\mathbb{E}_x[f(h(X_\sigma^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\sigma \mathcal{L}f(h(X_s^\varepsilon))ds]| = o(\varepsilon^\alpha) \quad \text{as} \quad \varepsilon \to 0, \tag{22}$$

$$\sup_{x \in \mathbb{T}^2} |\mathbb{E}_x[f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L}f(h(X_s^\varepsilon))ds]| \to 0 \quad \text{as} \quad \varepsilon \to 0, \tag{23}$$

$$\sup_{x\in\mathcal{E}^\varepsilon} |\mathbb{E}_x[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]| = o(\varepsilon^{\alpha-\varkappa}) \quad \text{as} \ \ \varepsilon \to 0, \tag{24}$$

$$\sup_{x\in\mathcal{E}^\varepsilon} \mathbb{E}_x\tau = o(\varepsilon^{\alpha-\varkappa}) \quad \text{as} \ \ \varepsilon \to 0. \tag{25}$$

Besides, if $\beta = 0$, then

$$\mathbb{E}_\nu[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds] = o(\varepsilon^\alpha) \quad \text{as} \ \ \varepsilon \to 0. \tag{26}$$

The proof of this lemma is essentially the same as that of the corresponding lemma in [4]. It relies on Lemma 2.4 and on the estimate for the expectation of the time it takes for the process to exit the ergodic component (Lemma 2.2). The proof of Lemma 2.2, however, requires an extra step compared to the proof of the corresponding lemma in [4]. We prove Lemma 2.2 in Section 3.

Here we only prove (26). This argument will explain the choice of the constant p in (13).

Let us denote the one-sided derivative of $f(h)$ at $h = 0$ by $f'(0)$. Then

$$f(h) = f(0) + f'(0)h + o(h) \ \ \text{as} \ \ h \to 0, \quad \text{and} \quad \mathcal{L}f(h) = pf'(0) + o(1) \ \ \text{as} \ \ h \to 0,$$

where p is the same as in the definition of the operator $\mathcal{L}$. Therefore, we can estimate the left hand side of (26) as follows

$$\mathbb{E}_\nu[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds] =$$

$$f'(0)\varepsilon^\alpha + o(\varepsilon^\alpha) - pf'(0)\mathbb{E}_\nu\tau + o(1)\mathbb{E}_\nu\tau = o(\varepsilon^\alpha) \quad \text{as} \ \ \varepsilon \to 0.$$

Here we used the facts that $0 \leq h(X^\varepsilon_s) \leq \varepsilon^\alpha$ for $0 \leq s \leq \tau$, that $\mathbb{E}_\nu\tau = k_2\varepsilon^\alpha(1 + o(1))$ as $\varepsilon \to 0$, where k_2 is the same as in Lemma 2.4, and that $k_2 = 1/p$.

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

$$\sup_{x\in\gamma} \mathbb{E}_x \sum_{n=0}^\infty \chi_{\{\sigma_n \leq \widetilde{\sigma}\}} |\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]| \leq \delta\rho \tag{27}$$

for all sufficiently small ε, where $\widetilde{\sigma}$ be the first of the stopping times σ_n such that $\widetilde{\sigma} > \rho$.

Proof. We'll divide the proof into several steps.

(a) Given $\rho > 0$, let μ^ε_x be the measure on $C([0, 2\rho], \mathbb{R}^2)$ induced by the process X^ε_t starting at x and $\widetilde{\mu}^\varepsilon_x$ be the measure on $C([0, 2\rho], \mathbb{R}^2)$ induced by the process $\widetilde{X}^\varepsilon_t$ starting at x. Let p^ε_x be the density of $\widetilde{\mu}^\varepsilon_x$ with respect to μ^ε_x. By the Girsanov theorem, there is $\rho_0 > 0$ such that for $\rho \leq \rho_0$ we have $\mu^\varepsilon_x(p^\varepsilon_x \leq 2) \geq 1 - \rho^2$ for all sufficiently small ε and all $x \in \gamma$. Let $\overline{\Omega} \subseteq C([0, 2\rho], \mathbb{R}^2)$ be the event where $p^\varepsilon_x > 2$ and $\Omega' \subseteq \Omega$ be the event that $(X^\varepsilon_t, t \in [0, 2\rho]) \in \overline{\Omega}$.

(b) For $T > 0$, let $\widetilde{\tau}$ be the first of the stopping times τ_n, which is greater than or equal to T, that is

$$\widetilde{\tau} = \min_{n:\tau_n \geq T} \tau_n. \tag{28}$$

Note that

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

The last inequality here is due to (21) and the Chebyshev inequality. Taking the sum in n, we obtain

$$\sum_{n=1}^{\infty} \mathbb{P}_x(\sigma_n < \widetilde{\tau}) \leq \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_n < \widetilde{\tau}) \leq \sum_{n=1}^{\infty} e^T(1 - r\varepsilon^\alpha)^{n-1} \leq K\varepsilon^{-\alpha}, \tag{29}$$

where the constant K depends on T. For $0 < \rho < 1$, we can take the same sum as in (29), but starting with $n = \varepsilon^{-\alpha}\ln(c/\rho)$ instead of $n = 1$. We then obtain that for each $\delta > 0$ there is a sufficiently large $c > 0$ that does not depend on ρ such that

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

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

$$\sum_{n=1}^{\infty} \mathbb{P}_x(\Omega' \cap \{\sigma_n < \rho\}) \leq \sum_{n=1}^{\infty} \mathbb{P}_x(\Omega' \cap \{\tau_n < \rho\}) \leq \varepsilon^{-\alpha}\ln(c/\rho)\rho^2 + \delta\rho\varepsilon^{-\alpha} \leq 2\delta\rho\varepsilon^{-\alpha}. \tag{30}$$

(c) It is sufficient to prove (27) with $\widetilde{\sigma}$ in the left hand side replaced by ρ, i.e., with the left hand side replaced by

$$\sup_{x\in\gamma} \mathbb{E}_x \sum_{n=0}^{\infty} \chi_{\{\sigma_n \leq \rho\}} |\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]|. \tag{31}$$

Indeed, the difference between this and the left hand side of (27) does not exceed

$$\sup_{x\in\gamma} |\mathbb{E}_x[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]|,$$

which is estimated using (24).

(d) Let us show that we can replace τ by $\tau' = \tau \wedge \rho$ in (31). Indeed, both $\mathbb{E}_x(\tau - \tau')$ and $\mathbb{P}_x(\tau \neq \tau')$ decay faster than any power of ε due to (25) and the Markov property of the process X^ε_t. Therefore, since f and $\mathcal{L}f$ are bounded,

$$\sup_{x\in\gamma}(|\mathbb{E}_x[f(h(X^\varepsilon_{\tau'})) - \int_0^{\tau'} \mathcal{L}f(h(X^\varepsilon_s))ds]| - |\mathbb{E}_x[f(h(X^\varepsilon_\tau)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]|)$$

decays faster than any power of ε. On the other hand,

$$\sup_{x\in\gamma}\mathbb{E}_x\sum_{n=0}^{\infty}\chi_{\{\sigma_n\leq\rho\}}$$

does not grow faster than a power of ε due to (29), which justifies our claim. Thus it remains to estimate

$$\sup_{x\in\gamma}\mathbb{E}_x\sum_{n=0}^{\infty}\chi_{\{\sigma_n\leq\rho\}}|\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_{\tau'}))-f(h(X^\varepsilon_0))-\int_0^{\tau'}\mathcal{L}f(h(X^\varepsilon_s))ds]|. \tag{32}$$

(e) Let us show that we can replace $\chi_{\{\sigma_n\leq\rho\}}$ in (32) by $\chi_{\{\sigma_n\leq\rho\}\setminus\Omega'}$. Indeed,

$$\mathbb{E}_x\sum_{n=0}^{\infty}\chi_{\{\sigma_n\leq\rho\}\cap\Omega'}\mathbb{E}_{X^\varepsilon_{\sigma_n}}\left|\int_0^{\tau'}\mathcal{L}f(h(X^\varepsilon_s))ds\right|\leq 2\rho\sup|\mathcal{L}f|\mathbb{P}_x(\Omega').$$

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

$$\mathbb{E}_x\sum_{n=0}^{\infty}\chi_{\{\sigma_n\leq\rho\}\cap\Omega'}\mathbb{E}_{X^\varepsilon_{\sigma_n}}|f(h(X^\varepsilon_{\tau'}))-f(h(X^\varepsilon_0))|\leq$$

$$2|f'(0)|\varepsilon^\alpha\sum_{n=1}^{\infty}\mathbb{P}_x(\Omega'\cap\{\sigma_n\leq\rho\}),$$

which can be made smaller than $\delta\rho$ for all sufficiently small ε by taking a sufficiently small ρ due to (30).

We have thus demonstrated that the expression in the left hand side of (27) can be approximated (with the accuracy of $\delta\rho$ with arbitrarily small δ) by

$$\sup_{x\in\gamma}\mathbb{E}_x\sum_{n=0}^{\infty}\chi_{\{\sigma_n\leq\rho\}\setminus\Omega'}|\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_{\tau'}))-f(h(X^\varepsilon_0))-\int_0^{\tau'}\mathcal{L}f(h(X^\varepsilon_s))ds]|. \tag{33}$$

(f) Note that due to the choice of the set Ω', the expectation in (33) differs from the same expectation for the process $\widetilde{X}^\varepsilon_t$ by at most a factor of 2 (with the stopping times corresponding to $\widetilde{X}^\varepsilon_t$). This shows that it is now sufficient to estimate (33) for the process $\widetilde{X}^\varepsilon_t$ (or, equivalently, for the process X^ε_t with $\beta=0$). We can now replace $\chi_{\{\sigma_n\leq\rho\}\setminus\Omega'}$ back by $\chi_{\{\sigma_n\leq\rho\}}$, which only makes the expectation larger, and replace τ' back by τ as above.

(g) For the remainder of the proof we can assume that $\beta=0$ and thus (26) applies. Note that the difference

$$\sum_{n=0}^{\infty}\mathbb{E}_x\chi_{\{\sigma_n\leq\rho\}}|\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau))-f(h(X^\varepsilon_0))-\int_0^{\tau}\mathcal{L}f(h(X^\varepsilon_s))ds]|-$$

$$\sum_{n=0}^{\infty} \mathbb{E}_x \chi_{\{\sigma_n \le \rho\}} |\mathbb{E}_\nu [f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L} f(h(X_s^\varepsilon)) ds]|.$$

can be estimated from above by

$$\sup_{x\in\gamma} |\mathbb{E}_x [f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L} f(h(X_s^\varepsilon)) ds])| \sum_{n=0}^{\infty} \sup_{x\in\gamma} (\mathrm{Var}(P_2^n(x,dy) - \nu(dy))),$$

which tends to zero due to (15) and (24).

(h) It remains to note that

$$\sum_{n=0}^{\infty} \mathbb{E}_x \chi_{\{\sigma_n \le \rho\}} |\mathbb{E}_{X_{\sigma_n}^\varepsilon} [f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L} f(h(X_s^\varepsilon)) ds]|$$

tends to zero due to (26) and (29). □

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 (14) is less than η for all sufficiently small positive ε.

First, we replace the time interval $[0, T]$ by a larger one, $[0, \widetilde{\tau}]$, where $\widetilde{\tau}$ is given by (28). Using the Markov property of the process, the difference can be rewritten as follows

$$|\mathbb{E}_x [f(h(X_{\widetilde{\tau}}^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^{\widetilde{\tau}} \mathcal{L} f(h(X_s^\varepsilon)) ds] -$$

$$\mathbb{E}_x [f(h(X_T^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^T \mathcal{L} f(h(X_s^\varepsilon)) ds]| =$$

$$|\mathbb{E}_x \mathbb{E}_{X_T^\varepsilon} [f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L} f(h(X_s^\varepsilon)) ds]|.$$

The latter expression can be made smaller than $\eta/5$ for all sufficiently small ε due to (23). Therefore, it remains to show that

$$|\mathbb{E}_x [f(h(X_{\widetilde{\tau}}^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^{\widetilde{\tau}} \mathcal{L} f(h(X_s^\varepsilon)) ds]| < \frac{4\eta}{5}$$

for all sufficiently small ε. Using the stopping times τ_n and σ_n we can rewrite the expectation in the left hand side of this inequality as follows

$$\mathbb{E}_x [f(h(X_{\widetilde{\tau}}^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^{\widetilde{\tau}} \mathcal{L} f(h(X_s^\varepsilon)) ds] =$$

$$\mathbb{E}_x [f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau \mathcal{L} f(h(X_s^\varepsilon)) ds] +$$

$$\sum_{n=1}^{\infty} \mathbb{E}_x(\chi_{\{\tau_n<\widetilde{\tau}\}} \mathbb{E}_{X^\varepsilon_{\tau_n}}[f(h(X^\varepsilon_\sigma)) - f(h(X^\varepsilon_0)) - \int_0^\sigma \mathcal{L}f(h(X^\varepsilon_s))ds]) + \tag{34}$$
$$\sum_{n=1}^{\infty} \mathbb{E}_x(\chi_{\{\sigma_n<\widetilde{\tau}\}} \mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]),$$

provided that the sums in the right hand side converge absolutely (which follows from the arguments below). Due to (23), the absolute value of the first term on the right hand side of this equality can be made smaller than $\eta/5$ for all sufficiently small ε. Therefore, it remains to estimate the two infinite sums.

Let us start with the first sum. By (22), we can find ε_0, such that for all $\varepsilon < \varepsilon_0$ we have

$$\sup_{x\in\overline{\gamma}} |\mathbb{E}_x[f(h(X^\varepsilon_\sigma)) - f(h(X^\varepsilon_0)) - \int_0^\sigma \mathcal{L}f(h(X^\varepsilon_s))ds]| \leq \frac{\eta\varepsilon^\alpha}{5K}.$$

Therefore, by (29), for $\varepsilon < \varepsilon_0$ we have

$$|\sum_{n=1}^{\infty} \mathbb{E}_x(\chi_{\{\tau_n<\widetilde{\tau}\}} \mathbb{E}_{X^\varepsilon_{\tau_n}}[f(h(X^\varepsilon_\sigma)) - f(h(X^\varepsilon_0)) - \int_0^\sigma \mathcal{L}f(h(X^\varepsilon_s))ds])| \leq$$

$$\sup_{x\in\overline{\gamma}} |\mathbb{E}_x[f(h(X^\varepsilon_\sigma)) - f(h(X^\varepsilon_0)) - \int_0^\sigma \mathcal{L}f(h(X^\varepsilon_s))ds]| \sum_{n=1}^{\infty} \mathbb{E}_x \chi_{\{\tau_n<\widetilde{\tau}\}} \leq \frac{\eta}{5}\ .$$

In order to estimate the second sum in the right hand side of (34), we introduce the stopping times $\overline{\sigma}_n$, where $\overline{\sigma}_0 = \sigma_1$ and σ_k, $1 \leq k \leq [T/\rho]+1$, is the first of the stopping times σ_n which is greater than $k\rho$. Then

$$|\sum_{n=1}^{\infty} \mathbb{E}_x(\chi_{\{\sigma_n<\widetilde{\tau}\}} \mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds])| \leq$$

$$\sum_{k=1}^{[T/\rho]+1} \mathbb{E}_x \sum_{n=0}^{\infty} \mathbb{E}_{X^\varepsilon_{\overline{\sigma}_k}}(\chi_{\{\sigma_n\leq\widetilde{\sigma}\}} |\mathbb{E}_{X^\varepsilon_{\sigma_n}}[f(h(X^\varepsilon_\tau)) - f(h(X^\varepsilon_0)) - \int_0^\tau \mathcal{L}f(h(X^\varepsilon_s))ds]|).$$

By Lemma 2.7, for all sufficiently small ε this does not exceed $([T/\rho]+1)\delta\rho$, where we can take $\delta = \eta/10T$. This provides the desired bound on the second sum in the right hand side of (34). □

3 Asymptotics of Transition Times

In this section we prove Lemmas 2.2 and 2.4.

Proof of Lemma 2.2. Let $\varkappa > 0$ be fixed. Let us note that Lemma 2.2 holds for the process $\widehat{X}^\varepsilon_t$ (instead of X^ε_t). In fact, it was proved in [4] for the case when σ is an identity matrix, but this assumption was not essential in the proof. Therefore, for for the process $\widehat{X}_t^\varepsilon$, any $\delta > 0$ and all sufficiently small ε we have

$$\mathbb{P}_x\left(\overline{\tau} \leq \varepsilon^{\frac{1}{2}-\frac{\varkappa}{2}}\right) \geq 1-\delta, \tag{35}$$

where the stopping time $\overline{\tau}$ for the process $\widehat{X}_t^\varepsilon$ is defined the same way as for the process X_t^ε. Let μ_x^ε be the measure on $C([0,\varepsilon^{\frac{1}{2}-\frac{\varkappa}{2}}],\mathbb{R}^2)$ induced by the process X_t^ε starting at x and $\widehat{\mu}_x^\varepsilon$ be the measure on $C([0,\varepsilon^{\frac{1}{2}-\frac{\varkappa}{2}}],\mathbb{R}^2)$ induced by the process $\widehat{X}_t^\varepsilon$ starting at x. Let p_x^ε be the density of μ_x^ε with respect to $\widehat{\mu}_x^\varepsilon$. By the Girsanov theorem, for each $\delta > 0$ we have $\widehat{\mu}_x^\varepsilon(p_x^\varepsilon \geq 1-\delta) \geq 1-\delta$ for all sufficiently small ε and all $x \in [\mathcal{E}]$. Therefore, since (35) holds for the process $\widehat{X}_t^\varepsilon$, for the process X_t^ε we have

$$\mathbb{P}_x\left(\overline{\tau} \leq \varepsilon^{\frac{1}{2}-\frac{\varkappa}{2}}\right) \geq 1-3\delta.$$

Due to the Markov property of the process X_t^ε, this implies that

$$\mathbb{P}_x\left(\overline{\tau} > n(\varepsilon^{\frac{1}{2}-\frac{\varkappa}{2}})\right) \leq (3\delta)^n.$$

Therefore there exists $\varepsilon_0 > 0$ such that $\mathbb{E}_x\overline{\tau} \leq \varepsilon^{\frac{1}{2}-\varkappa}$ for $\varepsilon \leq \varepsilon_0$ for all $x \in [\mathcal{E}]$. □

We will need the following lemma, which gives us the asymptotics of the time needed to exist the periodic component if the original point is asymptotically close to the boundary. It was proved in [11] (Lemma 4.4) for the case when $\beta = 0$ and σ is an identity matrix. The general case under consideration does not require major modifications to the proof.

Lemma 3.1. *There is a constant k_1, such that for any $\frac{1}{4} < \alpha < \frac{1}{2}$ we have*

$$\lim_{\varepsilon\to 0}\sup_{x\in\overline{\gamma}}\left|\frac{\mathbb{E}_x\sigma}{\varepsilon^\alpha} - k_1\right| = 1. \tag{36}$$

Proof of Lemma 2.4. Formula (16) follows from Lemma 3.1. Let us recall how to identify the constant k_1 (rigorous arguments can be found in [11]). If $x \in U$ were not to depend on ε, then the asymptotics of $\mathbb{E}_x\sigma$ could be obtained using the results of [8] modified to allow for general β and σ. Namely, recall the definition of the differential operator L from Section 2, and let $u(h)$ be the bounded solution of the ordinary differential equation

$$Lu = a(h)u'' + b(h)u' = -1, \quad h \in \mathrm{Int}(I), \tag{37}$$

with the boundary condition $u(0) = 0$. Such a solution exists and is unique (see [8], for example). It is equal to the expectation of the time it takes for the limiting process, starting at h, to reach the end-point of I corresponding to the boundary of the periodic component. It was demonstrated in [8] (Lemma 2.3) that $\lim_{\varepsilon\to 0}\mathbb{E}_x\sigma = u(h(x))$. In particular,

$$\lim_{\varepsilon\to 0}\mathbb{E}_x\sigma = u'(0)h(x) + o(h(x)) \quad \text{as} \quad h(x) \to 0.$$

Formula (36) is the corresponding asymptotic formula in the case when $h(x)$ is a function of ε, that is $h(x) = \varepsilon^\alpha$. In particular, $k_1 = u'(0)$. Equation (37) can be solved explicitly using the expressions for the coefficients $a(h)$ and $b(h)$. We obtain that

$$u'(0) = \int_I a^{-1}(t) \exp\left(\int_0^t \frac{b(h)}{a(h)} dh\right) dt,$$

which proves (18).

Next, let us prove (17). Let $\widehat{\gamma} = \{x : h(x) = \varepsilon^{\frac{1}{2}}\}$. We inductively define the following two sequences of stopping times. Let $\overline{\tau}_1 = \overline{\tau}$ be the first time when the process X_t^ε reaches the set $\widehat{\gamma}$. For $n \geq 1$ let $\overline{\sigma}_n$ be the first time following $\overline{\tau}_n$ when the process reaches γ. For $n \geq 2$ let $\overline{\tau}_n$ be the first time following $\overline{\sigma}_{n-1}$ when the process reaches $\widehat{\gamma}$.

First, let us estimate the probability of the event that the process which starts at $x \in \widehat{\gamma}$ reaches $\overline{\gamma}$ before reaching γ. Lemma 4.3 of [11] states that there is a constant c_1, such that for any $x \in \widehat{\gamma}$

$$\left|\mathbb{P}_x(\tau < \sigma) - \varepsilon^{\frac{1}{2}-\alpha}\right| \leq c_1 \varepsilon^\alpha \left|\ln \varepsilon\right|.$$

Since $\alpha > \frac{1}{4}$, this implies that $\mathbb{P}_x(\tau > \sigma) \leq 1 - \frac{1}{2}\varepsilon^{\frac{1}{2}-\alpha}$ for all sufficiently small ε, for all $x \in \widehat{\gamma}$. Using the Markov property of the process, we conclude that

$$\sup_{x\in\gamma} \mathbb{P}_x(\tau > \overline{\sigma}_n) \leq \left(1 - \frac{1}{2}\varepsilon^{\frac{1}{2}-\alpha}\right)^n.$$

We also need to estimate how much time it takes for the process which starts at $x \in \widehat{\gamma}$ to leave V^ε (the region between γ and $\overline{\gamma}$). Lemma 4.2 of [11] states that there is a constant c_2, such that for any $x \in V^\varepsilon$

$$\mathbb{E}_x \min(\tau, \sigma) \leq c_2 \varepsilon^{2\alpha} |\ln \varepsilon|.$$

Since $2\alpha > 1/2 - \varkappa$, the right hand side of this inequality is smaller than $\varepsilon^{\frac{1}{2}-\varkappa}$ for all sufficiently small ε. Therefore, by Lemma 2.2,

$$\sup_{x\in\gamma} \mathbb{E}_x \min(\tau, \overline{\sigma}_1) \leq \sup_{x\in\gamma} \mathbb{E}_x \overline{\tau} + \sup_{x\in\widehat{\gamma}} \mathbb{E}_x \min(\tau, \sigma) \leq 2\varepsilon^{\frac{1}{2}-\varkappa}.$$

Due to the Markov property of the process,

$$\sup_{x\in\gamma} \mathbb{E}_x \tau = \sup_{x\in\gamma} [\mathbb{E}_x \min(\tau, \overline{\sigma}_1) + \sum_{n=1}^\infty (\mathbb{E}_x \min(\tau, \overline{\sigma}_{n+1}) - \mathbb{E}_x \min(\tau, \overline{\sigma}_n))] =$$

$$\sup_{x\in\gamma} [\mathbb{E}_x \min(\tau, \overline{\sigma}_1) + \sum_{n=1}^\infty \mathbb{E}_x(\chi_{\{\tau > \overline{\sigma}_n\}} \mathbb{E}_{X^\varepsilon_{\overline{\sigma}_n}} \min(\tau, \overline{\sigma}_1))] \leq 2\varepsilon^{\frac{1}{2}-\varkappa} \sum_{n=0}^\infty \left(1 - \frac{1}{2}\varepsilon^{\frac{1}{2}-\alpha}\right)^n = 4\varepsilon^{\alpha-\varkappa}.$$

Since $\varkappa$ was arbitrary, this proves (17).

Finally, let us study the asymptotics of $\mathbb{E}_\nu \tau$ for the process $\widetilde{X}_t^\varepsilon$ defined by (9), for which the Lebesgue measure on the torus is invariant. By (16) applied to the process $\widetilde{X}_t^\varepsilon$, we have

$$\mathbb{E}_\mu \sigma = \varepsilon^\alpha (1+o(1)) \int_I a^{-1}(t) \exp\left(\int_0^t \frac{\widetilde{b}(h)}{a(h)} dh \right) dt, \quad \text{as } \varepsilon \to 0. \tag{38}$$

The process $\widetilde{X}_t^\varepsilon$ is ergodic. Applying the Birkhoff ergodic theorem to the process with the initial distribution ν, we obtain that

$$\lim_{n\to\infty} \frac{\int_0^n \chi_{\mathcal{E}}(\widetilde{X}_t^\varepsilon) dt}{n} = \text{Area}(\mathcal{E}) \ \text{ almost surely,}$$

where $\chi_{\mathcal{E}}$ is the indicator function of the set $\mathcal{E}$. Also, from the Birkhoff ergodic theorem we get

$$\lim_{n\to\infty} \frac{\sigma_n}{n} = \mathbb{E}_\nu \sigma_1 = \mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma \ \text{ almost surely.}$$

Using the Birkhoff ergodic theorem again, we can write

$$\lim_{n\to\infty} \frac{\int_0^n \chi_{\mathcal{E}}(\widetilde{X}_t^\varepsilon) dt}{n} = \lim_{n\to\infty} \frac{\int_0^{\sigma_n} \chi_{\mathcal{E}}(\widetilde{X}_t^\varepsilon) dt}{\sigma_n} = \left(\lim_{n\to\infty} \frac{\int_0^{\sigma_n} \chi_{\mathcal{E}}(\widetilde{X}_t^\varepsilon) dt}{n} \right) \left(\lim_{n\to\infty} \frac{n}{\sigma_n} \right) =$$

$$\frac{\mathbb{E}_\nu \int_0^{\sigma_1} \chi_{\mathcal{E}}(\widetilde{X}_t^\varepsilon) dt}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma} \leq \frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma},$$

where the equalities hold almost surely. Therefore,

$$\text{Area}(\mathcal{E}) \leq \frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma}.$$

In exactly the same way we can prove that

$$\text{Area}(\mathcal{E} \cup V^\varepsilon) \geq \frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma}, \quad \text{Area}(U \backslash V^\varepsilon) \leq \frac{\mathbb{E}_\mu \sigma}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma}, \text{ and } \text{Area}(U) \geq \frac{\mathbb{E}_\mu \sigma}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma}.$$

Combining these four estimates, we obtain that

$$\frac{\text{Area}(\mathcal{E})}{\text{Area}(U)} \leq \frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\mu \sigma} \leq \frac{\text{Area}(\mathcal{E} \cup V^\varepsilon)}{\text{Area}(U \backslash V^\varepsilon)}$$

Since $\lim_{\varepsilon\to 0} \text{Area}(V^\varepsilon) = 0$, we obtain

$$\lim_{\varepsilon\to 0} \frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\mu \sigma} = \frac{\text{Area}(\mathcal{E})}{\text{Area}(U)}. \tag{39}$$

Therefore, for the process $\widetilde{X}_t^\varepsilon$,

$$\mathbb{E}_\nu \tau = \varepsilon^\alpha (1+o(1)) \frac{\text{Area}(\mathcal{E})}{\text{Area}(U)} \int_I a^{-1}(t) \exp\left(\int_0^t \frac{\widetilde{b}(h)}{a(h)} dh \right) dt, \quad \text{as } \varepsilon \to 0. \tag{40}$$

Let

$$p(h) = \frac{1}{2}\left(\int_{\gamma(h)} \frac{1}{|\nabla H|} dl\right)^{-1}; \qquad q(h) = \int_{\gamma(h)} \frac{\langle \alpha \nabla H, \nabla H \rangle}{|\nabla H|} dl.$$

Thus, as follows from (10) and (12),

$$a(h) = p(h)q(h), \qquad \widetilde{b}(h) = p(h)q'(h),$$

and (40) can be written as

$$\lim_{\varepsilon \to 0} \frac{\mathbb{E}_\nu \tau}{\varepsilon^\alpha} = \frac{\text{Area}(\mathcal{E})}{\text{Area}(U)} \int_I \frac{1}{p(t)q(0)} dt = 2(\text{Area}(\mathcal{E})) \left(\left| \int_U \text{div}(\alpha \nabla H)(x) dx \right| \right)^{-1}. \tag{41}$$

This completes the proof of the lemma. □

4 The case of several periodic components

Let us assume that each of the domains U_k, $k = 1, ..., n$, contains a single critical point M_k (a maximum or a minimum of H). 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.

The phase space of the limiting process is now 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

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

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.

Let a_k, b_k and $\widetilde{b}_k$ be given by formulas (10), (11) and (12) (where $\gamma(h)$ is replaced by $\gamma^k(h_k)$, which is defined for each of the periodic components and has the same meaning as in the case of one periodic component). Let

$$p_k = \pm\frac{1}{2}(\text{Area}(\mathcal{E}))^{-1} \int_{\gamma_k} \frac{\langle \alpha \nabla H, \nabla H \rangle}{|\nabla H|} dl = \pm\frac{1}{2}(\text{Area}(\mathcal{E}))^{-1} \left| \int_{U_k} \text{div}(\alpha \nabla H)(x) dx \right|, \tag{42}$$

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

As in the case of one periodic component, we define the limiting process via its generator $\mathcal{L}$. For each k we define the differential operator $L_k f(h_k) = a_k(h_k) f'' + b_k(h_k) f'$ on the interior of I_k. The domain of $\mathcal{L}$ 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

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

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.

As in the case of one periodic component, we have the following theorem

Theorem 2. *The measure on 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 $h(X_0^\varepsilon)$.*

The proof of this theorem requires some modifications to the proof of Theorem 1. We sketch these modifications without providing some of the technical details.

(I) Recall the definition of the Markov chains ξ_n^1 and ξ_n^2 from Section 2. In the case of several periodic components, the state spaces for these Markov chains will be slightly different. Namely, we replace the curves γ and $\overline{\gamma}$ defined in Section 2 by the curves $\gamma = \bigcup_k \gamma_k$ and $\overline{\gamma} = \bigcup_k \overline{\gamma}_k$, where $\overline{\gamma}_k = \{|H_k| = \varepsilon^\alpha\}$. In the proof of Lemma 2.4, U will now stand for the union $U = \bigcup_k U_k$ of the periodic components. Let μ and ν be the invariant measures on $\overline{\gamma}$ and γ, respectively. Let us study the asymptotics $\mu(\overline{\gamma}_k)$ for different k.

Let μ_k be the normalized restriction of the measure μ to $\overline{\gamma}_k$, that is

$$\mu_k(A) = \mu(A)/\mu(\overline{\gamma}_k),$$

for each measurable subset A of $\overline{\gamma}_k$. Let us prove that

$$\mu(\overline{\gamma}_k) = \left| \int_{U_k} \mathrm{div}(\alpha \nabla H)(x) dx \right| \left(\sum_{i=1}^{n} \left| \int_{U_i} \mathrm{div}(\alpha \nabla H)(x) dx \right| \right)^{-1} (1 + o(1)), \tag{43}$$

as $\varepsilon \to 0$. Using arguments similar to those in the proof of Lemma 2.4, it is not difficult to show that we can replace the process X_t^ε by the process $\widetilde{X}_t^\varepsilon$ when studying the asymptotics of $\mu(\overline{\gamma}_k)$. Instead of (38) we now have

$$\mathbb{E}_{\mu_k} \sigma = 2(\mathrm{Area}(U_k)) \left| \int_{U_k} \mathrm{div}(\alpha \nabla H)(x) dx \right|^{-1}. \tag{44}$$

Let $\tau^{(k)}$ be the time when the process $\widetilde{X}_t^\varepsilon$ visits $\overline{\gamma}_k$ for the first time. Similarly to (39), we obtain

$$\frac{\mathbb{E}_\nu \tau}{\mathbb{E}_\mu \sigma} \sim \frac{\mathrm{Area}(\mathcal{E})}{\mathrm{Area}(U)} \quad \text{and} \tag{45}$$

$$\frac{\mathbb{E}_\nu \tau^{(k)}}{\mathbb{E}_{\mu_k} \sigma} \sim \frac{\mathrm{Area}(\mathbb{T}^2 - U_k)}{\mathrm{Area}(U_k)}, \quad \text{as } \varepsilon \to 0. \tag{46}$$

Let $N_k(T)$ be the number of times the process $\widetilde{X}_t^\varepsilon$ travels from γ to $\overline{\gamma}_k$ before time T, and $M(T)$ be the number of times the process travels from γ to $\overline{\gamma}$ before time T. Note that

$$\lim_{T\to\infty} \frac{N_k(T)}{M(T)} = \mu(\overline{\gamma}_k)(1+o(1)), \quad \text{as } \varepsilon \to 0.$$

By the Birghoff ergodic theorem

$$\lim_{T\to\infty} \frac{T}{N_k(T)} = (\mathbb{E}_\nu \tau^{(k)} + \mathbb{E}_{\mu_k}\sigma)(1+o(1)), \quad \text{as } \varepsilon \to 0,$$

and

$$\lim_{T\to\infty} \frac{T}{M(T)} = (\mathbb{E}_\nu \tau + \mathbb{E}_{\mu}\sigma)(1+o(1)), \quad \text{as } \varepsilon \to 0.$$

Therefore,

$$\lim_{\varepsilon\to 0} \frac{\mu(\overline{\gamma}_k)(\mathbb{E}_\nu \tau^{(k)} + \mathbb{E}_{\mu_k}\sigma)}{\mathbb{E}_\nu \tau + \mathbb{E}_\mu \sigma} = 1.$$

Now (44) and (46) imply that the expression $\mu(\overline{\gamma}_k)|\int_{U_k} \mathrm{div}(\alpha\nabla H)(x)dx|^{-1}$ is asymptotically independent of k, thus proving (43).

(II) Let us use (43) to justify (26). Near the root we have

$$f(h_k) = f(O) + \lim_{h_k\to 0} f'(h_k)h_k + o(h_k).$$

Let $r_k = 1$ if M_k is a maximum, and $r_k = -1$ if M_k is a minimum. Observe that, since μ and ν are invariant, $\nu(X_\tau^\varepsilon \in \overline{\gamma}_k) = \mu(\overline{\gamma}_k)$. Therefore,

$$\mathbb{E}_\nu \left(f(h(X_\tau^\varepsilon)) - f(h(X_0^\varepsilon)) - \int_0^\tau (Lf)(X_s^\varepsilon)ds \right) \sim \sum_k r_k \mu(\overline{\gamma}_k) \lim_{h_k\to 0} f'(h_k) - q\mathbb{E}_\nu\tau + o(\varepsilon^\alpha). \tag{47}$$

We can use the asymptotic expression for $\mathbb{E}_\nu\tau$ with the process $\widetilde{X}_t^\varepsilon$ instead of X_t^ε. From (44), (43), and (45) it follows that

$$\mathbb{E}_\nu \tau = 2\mathrm{Area}(\mathcal{E}) \left(\sum_{i=1}^n \left| \int_{U_k} \mathrm{div}(\alpha\nabla H)(x)dx \right| \right)^{-1} \varepsilon^\alpha(1+o(1)), \quad \text{as } \varepsilon \to 0.$$

Combining this with (43), we see that the right hand side of (47) is of order $o(\varepsilon^\alpha)$, thus justifying (26).

(III) The rest of the arguments follow the proof of Theorem 1. The proof of Lemma 2.3 (for which we referred to [4]) is somewhat more complicated in the case of several periodic components. Its proof in the case of several periodic components can also be easily modified from [4].

5 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 4 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

$$a_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \int_{\gamma^k(h_k)} \frac{\langle \varkappa \alpha \nabla H, \nabla H \rangle}{|\nabla H|} dl \quad \text{and} \tag{48}$$

$$b_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \int_{\gamma^k(h_k)} \frac{2\langle \beta + \varkappa u, \nabla H \rangle + \varkappa \alpha \cdot H''}{|\nabla H|} dl, \tag{49}$$

where

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

is the period of the unperturbed system. Note that

$$a_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1} \varkappa \overline{\varphi}_k (1 + o(1)), \quad |h_k| \downarrow 0, \tag{50}$$

$$b_k^\varkappa(h_k) = \frac{1}{2}(T_k(h_k))^{-1}(\overline{\psi}_k(1+o(1)) + \varkappa O(\ln(|h_k|))), \quad |h_k| \downarrow 0, \tag{51}$$

and therefore

$$\frac{b_k^\varkappa(h_k)}{a_k^\varkappa(h_k)} = \frac{\overline{\psi}_k}{\varkappa \overline{\varphi}_k}(1+o(1)) + O(\ln(|h_k|)), \quad |h_k| \downarrow 0. \tag{52}$$

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

$$\varkappa \sum_{k=1}^n p_k \lim_{h_k \to 0} f'(h_k) = q^\varkappa, \tag{53}$$

where p_k are given by (42).

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

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

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

$$\mu = \sum_{k=1}^n r_k. \tag{54}$$

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

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

Thus Z_t moves deterministically along the edge I_k of the graph with the speed $\bar{b}_k(h_k)$. 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 3. *The measure on 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 $dZ_t/dt = \bar{b}_k(Z_t)$, one can easily obtain the following statements:

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

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

and for each $T < \infty$,

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

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 3 will follow from the two lemmas below.

Lemma 5.1. *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 $\mathbb{G}^\delta = \{(i, h_i) \in \mathbb{G} : |h_i| \le \delta\}$. 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 $|h_i| \in (0, \delta)$, $1 \le i \le n$;

(b) The limits $\lim_{h_i \to 0} f'_{k,\varkappa}(h_i)$ exist and (53) 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 $i \ne k$.

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 $f_{i,\varkappa}(h_i)$, $0 \le |h_i| \le \delta$, $1 \le i \le n$. 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

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

where

$$I_i(\varkappa) = \int_0^\delta \exp\left(- \int_0^s \frac{b_i^\varkappa(u)}{a_i^\varkappa(u)} du \right) ds. \tag{55}$$

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

$$f_{k,\varkappa}(O) = \frac{p_k I_k^{-1}(\varkappa)}{\sum_{i=1}^n p_i I_i^{-1}(\varkappa)}. \tag{56}$$

From (52) it easily follows that

$$I_i(\varkappa) = (\varkappa + o(\varkappa)) \frac{\overline{\varphi}_i}{\overline{\overline{\psi}}_i} \quad \text{when} \quad \varkappa \downarrow 0 \quad \text{if} \quad s_i = 1; \tag{57}$$

$$I_i(\varkappa) \to \infty \quad \text{when} \quad \varkappa \downarrow 0 \quad \text{if} \quad s_i = 0. \tag{58}$$

Substituting this into (56), 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 5.2. *Let $\lambda \ge 0$, $Z_0^\varkappa = O$ and $r_k \ne 0$ for some k. Let A_m denote the event that $Z^\kappa_{\sigma^\varkappa(\delta)} \in I_m$. Then*

$$\mathbb{E}\left(\chi_{A_m} \exp(-\lambda \sigma^\varkappa(\delta))\right) = \frac{r_m}{\mu + \lambda}(1 + \xi_m(\lambda, \delta, \varkappa)) + \frac{\lambda \eta_m(\lambda, \delta, \varkappa)}{\mu + \lambda}, \tag{59}$$

where μ is defined in (54), $\lim_{\varkappa\downarrow 0}\xi_m(\lambda,\delta,\varkappa)) = 0$ *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

$$\mathbb{E}\exp(-\lambda\sigma^{\varkappa}(\delta)) = \frac{\mu}{\mu+\lambda}(1+\xi(\lambda,\delta,\varkappa)) + \frac{\lambda\eta(\lambda,\delta,\varkappa)}{\mu+\lambda}, \tag{60}$$

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

Proof. Let us prove (59). Let $f_\varkappa(h)$, $h \in \mathbb{G}^\delta$, be equal to the expectation in the left hand side of (59), 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 (53) 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,

$$f'_\varkappa(h_k) = \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)}du\right) ds + c_{k,\varkappa}\right) \exp\left(-\int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)}ds\right), \tag{61}$$

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

$$\sum_{k=1}^n p_k c_{k,\varkappa} = g_\varkappa(O)/\varkappa. \tag{62}$$

Upon integrating (61) 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

$$J_k(\delta,\lambda,\varkappa) = \int_0^\delta \left(\int_0^{h_k} \frac{g_\varkappa(s)}{a_k^\varkappa(s)} \exp\left(\int_0^s \frac{b_k^\varkappa(u)}{a_k^\varkappa(u)}du\right) ds\right) \exp\left(-\int_0^{h_k} \frac{b_k^\varkappa(s)}{a_k^\varkappa(s)}ds\right) dh_k, \tag{63}$$

$I_k(\varkappa)$ was defined in (55) 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 (62), 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 (57) and (58), 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 (52)

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

converges to zero when $\delta \downarrow 0$, while the second factor inside the integral in (63) 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 (52) to show that $J_k(\delta, \lambda, \varkappa)$ tends to zero when $\delta \downarrow 0$. This proves (59). Now (60) follows from (59) by summing over m. □

Finally, as it was already mentioned, the case where some of the periodic components contains saddles could be treated using the analysis of [3]. 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 $|\bar{\psi}_k|$.

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

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

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

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

[7] Freidlin M. I., Wentzell A. D. *Random perturbations of dynamical systems,* Second edition. Grundlehren der Mathematischen Wissenschaften **260** (1998) Springer, New York.

[8] Freidlin M. I., Wentzell A. D. *Random Perturbations of Hamiltonian Systems,* Memoirs of the American Math. Society. **109** (523), May 1994

[9] 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.

[10] Khinchin A. Ya. *Continued fractions,* Dover Publications, Inc., Mineola, NY, 1997.

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

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

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