ABSTRACT

Title of Dissertation: LONG-TERM BEHAVIOR OF RANDOMLY PERTURBED HAMILTONIAN SYSTEMS: LARGE DEVIATIONS AND AVERAGING

Shuo Yan
Doctor of Philosophy, 2024

Dissertation Directed by: Professor Leonid Koralov
Department of Mathematics

This dissertation concerns various asymptotic problems related to the long-term macroscopic behavior of randomly perturbed Hamiltonian systems, with different types of perturbations and on different time scales. Since the perturbations of the systems are assumed to be small while the systems are observed at large times, non-trivial phenomena can arise due to the interplay between the perturbation size and the temporal and spatial scales, and the systems often demonstrate qualitatively distinct types of behavior depending on the subtle quantitative relation between asymptotic parameters.

More specifically, on a natural time scale, i.e., in time required for the dynamics to move distance of order one, we investigate the dynamical systems with fast-oscillating perturbations and obtain precise estimates on the distribution. In particular, we calculate the exact asymptotics of the distribution in the case of linear dynamical systems. This problem also inspires the study of the local limit theorem for time-inhomogeneous functions of Markov processes. The local

limit theorem is a significant and widely used tool in problems of pure and applied mathematics as well as statistics. This result has been included in [1] and submitted for publication.

On the time scale that is inversely proportional to the effective size of the perturbation, we prove that the evolution of the first integral of the Hamiltonian system with fast-oscillating perturbations converges to a Markov process on the corresponding Reeb graph, with certain gluing conditions specified at the interior vertices. The result is parallel to the celebrated Freildin-Wentzell theory on the averaging principle of additive white-noise perturbations of Hamiltonian systems, and provides a description of the long-term behavior of a system when adopting an alternative approach to modeling random noise. Moreover, the current result provides the first scenario where the motion on a graph and the corresponding gluing conditions appear due to the averaging of a slow-fast system. It allows one to consider, for instance, the long-time diffusion approximation for an oscillator with a potential with more than one well. This result has been submitted for publication [2].

In the last part of the dissertation, we return to the more classical case of additive diffusion-type perturbations, combine the ideas of large deviations and averaging, and establish a large deviation principle for the first integral of the Hamiltonian system on intermediate time scales. Besides representing a new step in large deviations and averaging, this result will have important applications to reaction-diffusion equations and branching diffusions. The latter two concepts concern the evolution of various populations (e.g., in biology or chemical reactions). This result has been published in the journal *Stochastics and Dynamics* [3].

LONG-TERM BEHAVIOR
OF RANDOMLY PERTURBED HAMILTONIAN SYSTEMS:
LARGE DEVIATIONS AND AVERAGING

by

Shuo Yan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Leonid Koralov, Chair/Advisor
Professor Dmitry Dolgopyat
Professor Bassam Fayad
Professor Yu Gu
Professor Michael Fu

Dedication

I dedicate this dissertation to my parents.

Acknowledgments

I would like to express my sincere gratitude to all the people who helped me during my time in graduate school and made this dissertation possible.

First and foremost, I would like to thank my advisor, Professor Leonid Koralov, for his invaluable guidance, support, and encouragement throughout my journey from a fresh college graduate to a professional researcher. He has been an extraordinary mentor to me in mathematics and in life. He has spent countless hours with me, sharing knowledge, discussing research problems, and providing help and advice every time I encountered a difficult situation. I feel incredibly fortunate to have the opportunity to work under his supervision.

I would also like to take this opportunity to thank Professor Dmitry Dolgopyat and Professor Bassam Fayad for being excellent mentors and collaborators; Professor Mark Freidlin for providing interesting research problems and for helpful discussions; Professor Sandra Cerrai and Professor Yu Gu for teaching valuable courses and organizing interesting seminars in probability. Special gratitude is owed to Professor Michael Fu from Department of Decision, Operations and Information Technologies for serving as the Dean's representative during my defense.

I am deeply grateful for all the faculty members in the Department of Mathematics at the University of Maryland, for their contribution to an open and insightful academic atmosphere, allowing me to learn various subjects in mathematics and statistics. The staff members were always patient and helpful. In particular, Trystan Denhard, Cristina Garcia, and Jemma Natanson, in the Math Graduate Office, provided assistance with various matters in the PhD program, and Liliana Gonzalez and Ruth Yun helped me a lot with travel and payroll.

I want to acknowledge financial support from Professor Dmitry Dolgopyat and Professor Leonid Koralov which allowed me to focus on research during my second the fourth years. I also owe gratitude to Graduate School for the Dean's fellowship, Summer Research Fellowship, and Ann G. Wylie Dissertation Fellowship and to Department of Mathematics for Herbert A. Hauptman Endowed Graduate Fellowship and travel grants.

Finally, I wish to express my appreciation to all of my wonderful friends, who have been with me during my ups and downs and have witnessed my journey of pursuing the doctoral degree. And I owe my deepest gratitude to my family, who have always supported me, despite being on the opposite side of the earth.

Table of Contents

Chapter 1: Introduction

1.1 Random perturbations of dynamical systems

Consider the dynamical system defined by the ordinary differential equation with a vector field $b(\cdot)$:

$$dx_t = b(x_t)dt. \tag{1.1}$$

Understanding the evolution of complex dynamical systems is the subject of various important research problems in a wide range of fields, including natural sciences, economics, and engineering. Ideally, the evolution is deterministic since it obeys specific rules related to the system's current state, such as position, time, and temperature, thus a thorough study of the system should enable one to describe its properties and predict its behavior precisely. However, in practice, a purely deterministic description is impossible due to the complexity of the system, lack of information, and underlying uncertainties. This gives rise to probabilistic models that capture those factors as random components to describe real-world systems. For example, geometric Brownian motion is a probabilistic model to describe the evolution of a financial asset's price over time; while branching diffusion is used in biology to model how certain populations grow, or to describe the spread of certain infections.

Among all, we will focus on the situation where the random component serves as small noise influencing the system, i.e., random perturbations with an asymptotically small parameter. For instance, the additive white-noise-type perturbation is the most commonly considered object. In the foundational book [4], M. Freidlin and A. Wentzell consider, among others, Gaussian perturbations of dynamical systems

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}dW_t, \tag{1.2}$$

where W_t is a Brownian motion, and develop the large deviation principle for the measure induced by the process on the space of continuous functions. Namely, they describe the asymptotics of the probability of the event that the sample path stays close to a trajectory that is different from the deterministic process defined in (1.1). The probability is exponentially small in $1/\varepsilon$ with the action functional (also known as the rate function) that depends on the trajectory.

They also provide various averaging results for slow-fast systems. For example, let ξ be a random process, and define the process $\tilde{X}_t^\varepsilon$ by

$$d\tilde{X}_t^\varepsilon = b(\tilde{X}_t^\varepsilon, \xi_{t/\varepsilon})dt.$$

Here, the perturbation is no longer explicitly multiplied by a small parameter. However, with additional mixing assumptions on the process ξ, one can conclude that $\tilde{X}_t^\varepsilon$ can indeed be viewed as a perturbation of the "averaged dynamical system" defined by (1.1) with $b(x) = \langle \pi, b(x, \cdot) \rangle$, where π is the unique invariant measure of the process ξ (cf. [5], [6]). More precisely,

$$\tilde{X}_t^\varepsilon \to x_t \text{ in probability as } \varepsilon \downarrow 0,$$

$$\frac{1}{\sqrt{\varepsilon}}(\tilde{X}_t^\varepsilon - x_t) \to \zeta_t \text{ weakly as } \varepsilon \downarrow 0,$$

where ζ is a Gaussian process. Note that, locally in time, the vector field $b(x_t, \xi_{t/\varepsilon})$ changes primarily due to the fast component $\xi_{t/\varepsilon}$ and is fast oscillating as a result. Thence, we refer to this type of perturbation as the fast-oscillating random perturbation of x_t in (1.1).

In the next section, we will discuss in more detail the two types of perturbations in the special case where (1.1) is given by a Hamiltonian system.

1.2 Hamiltonian systems

Consider a smooth function H and the dynamical system (1.1) on $\mathbb{R}^{2d}$ with $b(x) = \nabla^\perp H(x)$, where

$$\nabla^\perp H(p,q) = (-\nabla_q H, \nabla_p H), \quad x = (p,q).$$

It is clear that the dynamical system preserves the value of H, so a trajectory of the system lies in a single level set of H. In particular, in two-dimensional Hamiltonian systems, the trajectory coincides with a connected component of the level set if the latter is homeomorphic to a circle. See Figure 1.1 for an example in $\mathbb{R}^2$.

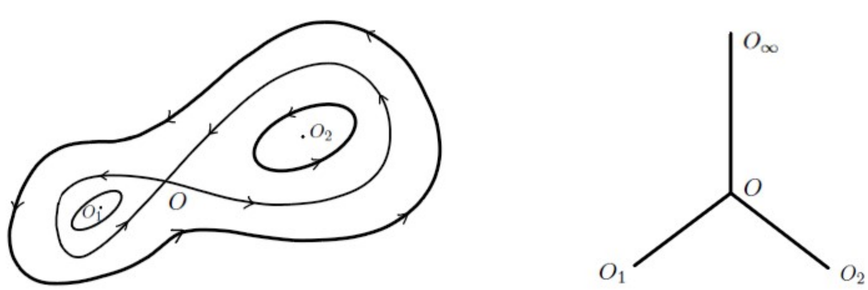


Figure 1.1: Level sets and the Reeb graph.

Hamiltonian systems have been widely used to model dynamics in various physical systems. In [7], non-degenerate additive white-noise perturbations of two-dimensional Hamiltonian systems were considered. Due to the special structure of Hamiltonian systems, new phenomena (compared to the trivial convergence to the unperturbed dynamics (1.1) on finite time intervals) can be observed on the corresponding Reeb graph ([8]) on a larger time scale. Namely, in $\mathbb{R}^2$, the process $h(X^{\varepsilon}_{t/\varepsilon})$ converges weakly to a strong Markov process, where h is the projection onto the Reeb graph. The limiting process on the graph is a diffusion inside the edges, with gluing conditions specified at the interior vertices. The averaging principle can be applied to the perturbed system (1.2) due to the unperturbed system (1.1) being ergodic on the connected components of the level sets of the Hamiltonian H.

This result inspired a sequence of papers that considered similar problems in different and/or more general situations: additive random perturbations of nonlinear oscillators ([9], [10]), additive deterministic perturbations of Hamiltonian systems ([11]), additive small random and deterministic perturbations of dynamical systems and diffusion processes on a multi-dimensional space with a first integral ([12], [13], [14]), etc. There are also results on intermediate time scales in the vicinity of the boundaries ([15], [16]). In Chapter 4, we consider the large deviation principle of the pre-limiting process on the Reeb graph on the intermediate time scales.

In the case of multi-frequency systems with more than one first integral, the limiting process is on the open book space instead of the Reeb graph. The situation is more complicated, and several results in this direction include perturbations of multi-frequency systems outside of the singularities ([17]), perturbations of weakly coupled oscillators ([18], [19]), etc.

Similar problems, including the case where the dynamical system (1.1) is only locally Hamiltonian, have also been considered on symplectic manifolds, see e.g., [20], [21], [22], [23], and [24].

All the results listed above concern additive perturbations. There are relatively few results about fast-oscillating perturbations in the existing literature. Two-dimensional problems were considered in [25] and higher-dimensional problems were considered in [26], both outside of the singularities. In Chapter 2, we deal with the situation with presence of multiple critical points, including saddle points, of the Hamiltonian H in $\mathbb{R}^2$, and prove the weak convergence of the projection of the process on the Reeb graph.

1.3 Plan of the dissertation

In Chapter 2, we establish the averaging principle for fast-oscillating random perturbations of Hamiltonian systems on $\mathbb{R}^2$. The study requires a number of novel technical tools, in particular, a local limit theorem for fast-oscillating random perturbations of linear dynamical systems. It turns out to be a special case of the local limit theorem of time-inhomogeneous functions of Markov processes, which potentially has broader applications than the problems we consider here, and we allocate the whole Chapter 3 to the proof of this general result. Finally, in Chapter 4, we consider additive perturbations of Hamiltonian systems on intermediate time scales. We prove the large deviation result, with the action functional determined by the averaging principle. Our result establishes an important relationship between the two types of asymptotic regimes.

Since we have many objects defined in this dissertation, we need to reuse some notation. Please keep in mind that the same notation usually refers to the same object in the same chapter, unless specified otherwise, but not necessarily in different chapters. Each chapter is relatively self-contained, with required references cited, and can be read independently if one wishes to.

Chapter 2: Fast-oscillating random perturbations of Hamiltonian systems

2.1 Introduction

Consider a diffusion process $(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)$ in $\mathbb{R}^2 \times \mathbb{T}^m$ satisfying

$$
\begin{aligned}
d\boldsymbol{X}_t^\varepsilon &= b(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)dt, \qquad \boldsymbol{X}_0^\varepsilon = x_0 \in \mathbb{R}^2, \\
d\boldsymbol{\xi}_t^\varepsilon &= \frac{1}{\varepsilon} v(\boldsymbol{\xi}_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} \sigma(\boldsymbol{\xi}_t^\varepsilon)dW_t, \quad \boldsymbol{\xi}_0^\varepsilon = y_0 \in \mathbb{T}^m,
\end{aligned}
\tag{2.1}
$$

where ε is a small positive parameter, $\mathbb{T}^m$ is the m-dimensional torus, and W_t is an m-dimensional Brownian motion. In the coupled system, $\boldsymbol{X}_t^\varepsilon$ is the slow motion and $\boldsymbol{\xi}_t^\varepsilon$ is the fast motion, since the generator of the diffusion in the second equation is scaled by ε^{-1}. In this system, all the randomness comes from the second equation and the slow motion depends on the fast one in a deterministic way, and this dependence results in fast-changing velocity for the slow motion. Under natural conditions, the averaging principle holds for the process in (2.1) (cf. [4]). For example, if $\sigma(y)$ is non-degenerate (and thus $\boldsymbol{\xi}_t^\varepsilon$ has a unique invariant measure μ independent of ε), then $\boldsymbol{X}_t^\varepsilon$ converges as $\varepsilon \to 0$ in probability on each finite interval $[0, T]$ to an averaged process defined by the differential equation

$$
d\boldsymbol{x}_t = \bar{b}(\boldsymbol{x}_t)dt, \tag{2.2}
$$

where $\bar{b}(x) = \int_{\mathbb{T}^m} b(x, y) d\mu(y)$. Therefore, $\boldsymbol{X}_t^\varepsilon$ can be viewed as a result of fast-oscillating random perturbations of the deterministic process $\boldsymbol{x}_t$. Moreover, the deviation can be described more precisely: the process $\varepsilon^{-1/2}(\boldsymbol{X}_t^\varepsilon - \boldsymbol{x}_t)$ converges weakly to a Gaussian Markov process on a finite interval $[0, T]$ (cf. [4]), and, if we assume a special type of vector $b(x, y)$, then the local limit theorem holds for $\varepsilon^{-1}(\boldsymbol{X}_t^\varepsilon - \boldsymbol{x}_t)$ at time t ([1]).

If the system (2.2) has a first integral H, then, by the averaging principle, $H(\boldsymbol{X}_t^\varepsilon)$ is nearly constant on finite time intervals when ε is small. Nontrivial behavior can, however, be observed on larger time intervals (of order ε^{-1}). Assume, momentarily, that H has a single critical point. Then it was demonstrated in [25] that $H(\boldsymbol{X}_{t/\varepsilon}^\varepsilon)$ converges weakly in $C([0, T])$, as $\varepsilon \to 0$, to a diffusion process for any finite T, under additional assumptions. A similar result in the case of multiple degrees of freedom was obtained recently in [26] and the main goal there was to overcome difficulties related to resonances, which is typical in the case of multiple degrees of freedom. The result holds in the region where no critical points of the first integrals are present and action-angle-type coordinates can be introduced.

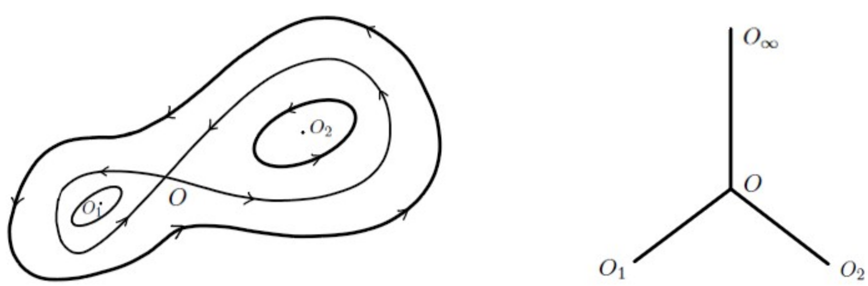


Figure 2.1: Level sets and the Reeb graph.

Let us return to the two-dimensional situation. In the presence of multiple critical points, including saddle points, the problem gets more complicated as we need to consider the Reeb graph in order to describe the evolution of the first integrals denoted by $h = (k, H)$. (For instance, in Figure 2.1, we have two local minima and one saddle point in the space and thus two exterior vertices and one interior vertex on the graph.) In particular, the interior vertices on the graph correspond to the level curves that contain the saddle points, and those level curves are called the separatrices. In this situation, the limiting behavior has already been described for the white-noise-type additive perturbations of dynamical systems: Hamiltonian systems in $\mathbb{R}^2$ ([4]), general dynamical systems with conservation laws in $\mathbb{R}^n$ ([27]), and Hamiltonian systems with an ergodic component on two-dimensional surfaces ([22],[23],[24]).

In this chapter, we consider fast-oscillating random perturbations, as discussed above, of Hamiltonian system in $\mathbb{R}^2$ with multiple critical points and prove that the evolution of the first integrals h converges to a diffusion process defined by an operator $(\mathcal{L}, D(\mathcal{L}))$ on the corresponding Reeb graph. In particular, the exterior vertices turn out to be inaccessible and the behavior of the process near the interior vertices is described in terms of the domain $D(\mathcal{L})$ in the following way: for interior vertex O_i, there are constants p_k such that each function $f \in D(\mathcal{L})$ satisfies

$$\sum_{I_k \sim O_i} p_k \lim_{h_k \to O_i} f'(h_k) = 0. \tag{2.3}$$

Intuitively, the absolute value of p_k is proportional to the probability of entering edge I_k after the process arrives at the vertex O_i. The relation (2.3) is usually referred to as the gluing condition. In the next section, we will formulate the results along with the assumptions more precisely. The coefficients p_k will be calculated explicitly. As we mentioned, similar results hold in case of additive perturbations of Hamiltionian systems. Now the techniques in the proof are more involved and require new ideas with analysis on multiple time scales : $O(\varepsilon^{-1})$, $O(1)$, $O(\varepsilon)$, etc. It is worth noting that our result provides the first example where the motion on a graph and the corresponding gluing conditions appear as a result of averaging of a fast-slow system.

To start with, since our interest is in the long-time behavior of $\boldsymbol{X}_t^\varepsilon$ on $O(\varepsilon^{-1})$ time scales, it is often convenient to consider a temporally re-scaled process $(X_t^\varepsilon, \xi_t^\varepsilon)$:

$$
\begin{aligned}
dX_t^\varepsilon &= \frac{1}{\varepsilon} b(X_t^\varepsilon, \xi_t^\varepsilon)dt, \qquad X_0^\varepsilon = x_0 \in \mathbb{R}^2, \\
d\xi_t^\varepsilon &= \frac{1}{\varepsilon^2} v(\xi_t^\varepsilon)dt + \frac{1}{\varepsilon}\sigma(\xi_t^\varepsilon)dW_t, \quad \xi_0^\varepsilon = y_0 \in \mathbb{T}^m.
\end{aligned} \tag{2.4}
$$

It is clear that $(X_t^\varepsilon, \xi_t^\varepsilon) = (\boldsymbol{X}_{t/\varepsilon}^\varepsilon, \boldsymbol{\xi}_{t/\varepsilon}^\varepsilon)$ in distribution. Thus, it suffices to prove the weak convergence of $h(X_t^\varepsilon)$ in the space $\mathbf{C}([0,T], \mathbb{G})$, where $\mathbb{G}$ is the Reeb graph. The proof of the weak convergence relies on demonstrating that the pre-limiting process asymptotically solves the martingale problem. Namely, we will show that, for each f in a sufficiently large subset of $D(\mathcal{L})$ and $T > 0$,

$$
\mathbf{E}_{(x,y)}[f(h(X_T^\varepsilon)) - f(h(x)) - \int_0^T \mathcal{L}f(h(X_t^\varepsilon))dt] \to 0, \tag{2.5}
$$

as $\varepsilon \to 0$, uniformly in x in any compact set in $\mathbb{R}^2$ and in $y \in \mathbb{T}^m$. The main idea in our proof is to divide the time interval $[0,T]$ into smaller intervals between different visits to the separatrices and show that the contributions from all individual excursions are small and do not accumulate. For example, suppose for now that there is only one saddle point, as shown in Figure 2.2. Let O be the saddle point with $H(O) = 0$, γ be the separatrix, $\gamma' = \{x : |H(x)| = \varepsilon^\alpha\}$ be a set near the separatrix, where $0 < \alpha < 1/2$, and $\sigma \geq 0$ be the first time when the process X_t^ε reaches γ.

Define inductively the two sequences of stopping times:

$$
\begin{aligned}
\sigma_0 &= \sigma, \\
\tau_n &= \inf\{t > \sigma_{n-1} : X_t^\varepsilon \in \gamma'\}, \qquad (2.6)\\
\sigma_n &= \inf\{t > \tau_n : X_t^\varepsilon \in \gamma\},
\end{aligned}
$$

and consequently two Markov chains $(X^\varepsilon_{\tau_n}, \xi^\varepsilon_{\tau_n})$ and $(X^\varepsilon_{\sigma_n}, \xi^\varepsilon_{\sigma_n})$. As pointed out earlier, we wish

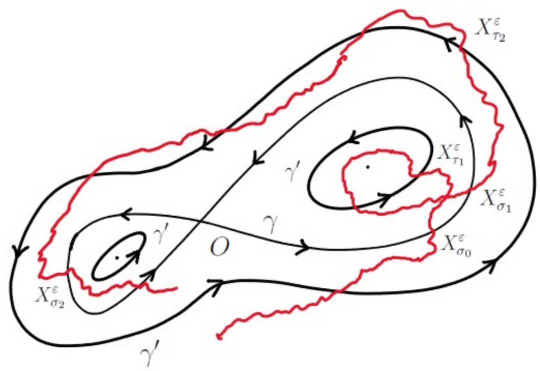


Figure 2.2: Construction of discrete Markov chains.

to prove that the contributions to (2.5) from all individual excursions are small and the sum converges to zero as $\varepsilon \downarrow 0$. Thus, it is sufficient to show that (a) the expectation corresponding to one excursion converges to zero as $\varepsilon \downarrow 0$ uniformly in the initial distribution, (b) the expectation corresponding to one excursion is exactly zero if the process starts with the invariant measure of the Markov chain $(X^\varepsilon_{\sigma_n}, \xi^\varepsilon_{\sigma_n})$ on $\gamma \times \mathbb{T}^m$, and (c) the measures on $\gamma \times \mathbb{T}^m$ induced by $(X^\varepsilon_{\sigma_n}, \xi^\varepsilon_{\sigma_n})$ converge exponentially, as $n \to \infty$, uniformly in ε and in starting point, to the invariant one.

The claim in (a) is an extension of the results outside singularities in [26], and new difficulties arise due to the degenerations occurring on the boundaries. The claim in (b) is true if the gluing conditions are chosen appropriately and there is a common invariant measure for the processes $(X^\varepsilon_t, \xi^\varepsilon_t)$ for all ε. In general, there is no common invariant measure for all ε, and we need to consider a family of auxiliary processes near the separatrix that do have a common invariant measure, and then use the proximity of the auxiliary and the original processes and the Girsanov theorem. The assertion in (c) is harder to verify, and its proof requires new techniques, including a local limit theorem and density estimates for hypoelliptic diffusions that will be discussed in later sections.

The chapter is organized as follows: In Section 2.2, we introduce the notations, state the assumptions, and formulate the main result. In Section 2.3, we construct an auxiliary process and derive diffusion approximations of the processes. In Section 2.4, we prove the averaging principle up to the time when the process reaches the separatrix. In Section 2.5, we construct the Markov chain on the separatrix (see (2.6)) and prove its mixing properties. In Section 2.6, we prove the main result. A few technical results are included in the Section 2.7.

2.2 Main results

Throughout this chapter, $\mathbf{P}$ and $\mathbf{E}$ represent the probability and expectation, respectively, and the subscripts pertain to initial conditions. For brevity, the stopping times' dependence on parameters and initial conditions is not always indicated in the notation when introduced (e.g. (2.6)). ∇ denotes a first order differential operator, i.e., derivative, gradient, Jacobian, etc., depending on the context. χ_A denotes the indicator function of the event A. If A and B are two non-negative functions that depend on an asymptotic parameter, we write $A \lesssim B$ if $A = O(B)$. $\mathbf{C}_0(\mathbb{G})$ is the space of continuous functions on the Reeb graph $\mathbb{G}$ that tend to zero at infinity with uniform norm. h is the projection onto $\mathbb{G}$. In order to formulate the assumptions and results, we introduce some notation:

1. O_i's are the vertices on the graph and are occasionally used to denote the corresponding critical points on the plane when there is no ambiguity. I_k's are the edges on the graph and U_k's are the corresponding two-dimensional domains. Formally, O_∞ is the vertex that corresponds to infinity. A symbol $\sim$ between a vertex and an edge means that the vertex is an endpoint of the edge.

2. Consider the following metric on $\mathbb{G}$: $r(h_1, h_2)$ is the length of the shortest path connecting h_1 and h_2. For example, if $h_1 = (1, H_1)$, $I_1 \sim O_1$, $O_1 \sim I_2$, $I_2 \sim O_2$, $O_2 \sim I_3$ and $h_2 = (3, H_2)$, then $r(h_1, h_2) = |H_1 - H(O_1)| + |H(O_1) - H(O_2)| + |H(O_2) - H_2|$.

3. $\gamma(h) = \{x : H(x) = h\}$ and $\gamma_k(h)$ is the connected component of $\gamma(h)$ in the domain U_k.

4. $b_h(x, y) = \nabla H(x) \cdot b(x, y)$.

5. ξ_t is the diffusion process on $\mathbb{T}^m$ with the generator L, where

$$Lf(y) = v(y) \cdot \nabla_y f(y) + \frac{1}{2} \sum_{i,j} (\sigma\sigma^*)_{ij}(y) \frac{\partial^2}{\partial y_i \partial y_j} f(y). \tag{2.7}$$

6. For h in the interior of I_k, define

$$\begin{aligned}
Q_k(h) &= \int_{\gamma_k(h)} \frac{dl}{|\nabla H(x)|}, \\
A_k(h) &= \frac{2}{Q_k(h)} \int_{\gamma_k(h)} \frac{1}{|\nabla H(x)|} \int_0^\infty \mathbf{E}_\mu b_h(x, \xi_s) b_h(x, \xi_0) ds dl, \\
B_k(h) &= \frac{1}{Q_k(h)} \int_{\gamma_k(h)} \frac{1}{|\nabla H(x)|} \int_0^\infty \mathbf{E}_\mu \nabla_x b_h(x, \xi_s) \cdot (b(x, \xi_0) - \nabla^\perp H(x)) ds dl, \\
L_k f(h) &= \frac{1}{2} A_k(h) f''(h) + B_k(h) f'(h).
\end{aligned}$$

The following conditions are assumed to hold throughout the chapter.

Assumptions on the coefficients:

(H1) $v(y)$ and $\sigma(y)$ are C^∞ functions on $\mathbb{T}^m$. $\sigma(y)$ is $m \times m$ matrix-valued and $\sigma(y)\sigma(y)^\mathsf{T}$ is positive-definite for all $y \in \mathbb{T}^m$.

(H2) $H(x)$ is a C^∞ function from $\mathbb{R}^2$ to $\mathbb{R}$ with bounded second derivatives. $H(x)$ has a finite number of non-degenerate critical points. Each level curve corresponding to a vertex on the Reeb graph contains at most one critical point. As $|x| \to +\infty$, $H(x)/|x| \to +\infty$.

(H3) $b(x, y)$ is a C^∞ function from $\mathbb{R}^2 \times \mathbb{T}^m$ to $\mathbb{R}^2$ such that the averaged process is a Hamiltonian system with H, i.e. $\bar{b}(x) = \nabla^\perp H(x)$.

(H4) The fast-oscillating perturbation is non-degenerate, i.e. $\{b(x, y) - \bar{b}(x) : y \in \mathbb{T}^m\}$ spans $\mathbb{R}^2$ for each $x \in \mathbb{R}^2$, and is uniformly bounded together with its first derivatives.

(H5) For each x that belongs to one of the separatrices, there exists $y \in \mathbb{T}^m$ such that the process in (2.1) satisfies the parabolic Hörmander condition at (x, y). Namely, with $\varepsilon^{-1}\tilde{v}(y)$ being the drift term in the equation for $\boldsymbol{\xi}_t^\varepsilon$ in the Stratonovich form, we have that

$$\mathrm{Lie}\left(\left\{\begin{pmatrix} 0 \\ \sigma_k(y) \end{pmatrix}, 1 \le k \le m\right\} \bigcup \left\{\left[\begin{pmatrix} b(x,y) \\ \tilde{v}(y) \end{pmatrix}, \begin{pmatrix} 0 \\ \sigma_k(y) \end{pmatrix}\right], 1 \le k \le m\right\}\right)$$

at (x, y) spans $\mathbb{R}^{2+m}$, where $\sigma_k(y)$ is the k-th column of $\sigma(y)$, $[\cdot, \cdot]$ is the Lie bracket, and $\mathrm{Lie}(\cdot)$ is the Lie algebra generated by a set (cf. [28] or Section 2.3.2 of [29]).

Definition 2.2.1. *The domain $D(\mathcal{L})$ consists of functions $f \in \mathbf{C}_0(\mathbb{G})$ satisfying:*

(i) f is twice continuously differentiable in the interior of each edge I_k of $\mathbb{G}$;

(ii) The limits $\lim_{h_k \to O_i} L_k f(h_k)$ *exist and do not depend on the edge* I_k*;*

(iii) For interior vertex O_i*, there are constants* $p_k := \pm \lim_{h \to O_i} A_k(h) Q_k(h)$ *such that*

$$\sum_{I_k \sim O_i} p_k \lim_{h_k \to O_i} f'(h_k) = 0, \tag{2.8}$$

where the sign $+$ *is taken if* O_i *is minimum on* I_k*, and the sign* $-$ *is taken otherwise. The operator* $\mathcal{L}$ *on the Reeb graph is defined by*

$$\mathcal{L} f(h) = L_k f(h) \tag{2.9}$$

for $f \in D(\mathcal{L})$ *and* h *in the interior of* I_k*, and defined as* $\lim_{h \to O_i} \mathcal{L} f(h)$ *at the vertex* O_i*.*

By the Hille-Yosida theorem (see, for example, Theorem 4.2.2 in [30]), one can check that there exists a unique strong Markov process on $\mathbb{G}$ with continuous sample paths that has $\mathcal{L}$ as its generator. Now we are ready to formulate the main result of this chapter.

Theorem 2.2.2. *Let the process* $(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)$ *be defined as in* (2.1) *and the conditions (H1)-(H5) hold. Then* $h(\boldsymbol{X}_{t/\varepsilon}^\varepsilon)$ *converges weakly to the strong Markov process on the Reeb graph* $\mathbb{G}$ *that has the generator* $(\mathcal{L}, D(\mathcal{L}))$ *and the initial distribution* $h(x_0)$*.*

Remark 2.2.3. *The last condition in (H2) can be relaxed without much extra effort since the limiting process defined by* $\mathcal{L}$ *cannot reach infinity in finite time. In addition, as seen from the proofs in Section 2.5 and Remark 2.5.7, assumption (H5) can be relaxed so that it holds for at least one point on each separatrix. Moreover, if the number of Lie brackets needed to generate* $\mathbb{R}^{2+m}$ *in the parabolic Hörmander condition is assumed to be given, then we can relax the assumptions on smoothness of the coefficients.*

To prove the theorem, we need a result on weak convergence of processes, that is Lemma 4.1 in [31] adapted to our case (see also the original statement in [7]):

Lemma 2.2.4. *Let Ψ be a dense linear subspace of $\mathbf{C}_0(\mathbb{G})$ and $\mathcal{D}_{\mathcal{L}}$ be a linear subspace of $D(\mathcal{L})$, and suppose that Ψ and $\mathcal{D}_{\mathcal{L}}$ have the following properties:*

(1) There is a $\lambda > 0$ such that for each $F \in \Psi$ the equation $\lambda f - \mathcal{L}f = F$ has a solution $f \in \mathcal{D}_{\mathcal{L}}$;

(2) For each $T > 0$, each $f \in \mathcal{D}_{\mathcal{L}}$, and each compact $K \subset \mathbb{G}$,

$$\mathbf{E}_{(x,y)}[f(h(X_T^\varepsilon)) - f(h(x)) - \int_0^T \mathcal{L}f(h(X_t^\varepsilon))dt] \to 0, \tag{2.10}$$

uniformly in $x \in h^{-1}(K)$ and $y \in \mathbb{T}^m$.

Suppose that the family of measures on $\mathbf{C}([0,\infty], \mathbb{G})$ induced by the processes $h(X_t^\varepsilon)$, $\varepsilon > 0$, is tight for each $(x,y) \in \mathbb{R}^2 \times \mathbb{T}^m$. Then, for each $(x,y) \in \mathbb{R}^2 \times \mathbb{T}^m$, $h(X_t^\varepsilon)$ converges weakly to the strong Markov process on the Reeb graph $\mathbb{G}$ that has the generator $(\mathcal{L}, D(\mathcal{L}))$ and the initial distribution $h(x)$.

Here we choose Ψ to be all the functions in $\mathbf{C}_0(\mathbb{G})$ that are twice continuously differentiable in the interior of each edge; $\mathcal{D}_{\mathcal{L}}$ to be all the functions in $D(\mathcal{L})$ that are four times continuously differentiable in the interior of each edge. It is easy to check condition (1) holds in Lemma 2.2.4, and the tightness of distributions of $h(X_t^\varepsilon)$ for all $\varepsilon > 0$ is verified in Section 2.7.3. Then the main ingredient of the proof is to verify (2.10) in condition (2) of Lemma 2.2.4.

2.3 Preliminaries

In this section, we explain some technical difficulties and our approach to the proof.

2.3.1 Localization

Considering Theorem 2.2.2 for $\boldsymbol{X}_t^\varepsilon$ in the state space $\mathbb{R}^2$ causes technical difficulties due to the presence of multiple separatrices of the Hamiltonian and to the fact that the process $\boldsymbol{X}_t^\varepsilon$ is not positive recurrent. However, such difficulties can be circumvented by considering the process $\boldsymbol{X}_t^\varepsilon$ locally. Namely, let us cover the plane $\mathbb{R}^2$ by finitely many bounded domains, each containing one of the separatrices and bounded by up to three connected components of level sets of H, and one unbounded domain not containing any critical points. For example, as shown in Figure 2.3, we have different parts of the Reeb graph $\mathbb{G}$ that correspond to the domains in $\mathbb{R}^2$. Every point of $\mathbb{R}^2$ can be assumed to be contained in the interior either one or two domains. Since it takes positive time to travel from the boundary of one domain to the boundary of another, it suffices to prove the result up to time of exit from one domain. To be more precise, let $\{V_k : 1 \le k \le K\}$ be the open cover. Define $\eta_0 = \inf\{t \ge 0 : X_t^\varepsilon \in \bigcup_{1\le k\le K} \partial V_k\}$ and, for k such that $X_{\eta_{n-1}}^\varepsilon \in V_k$, define $\eta_n = \inf\{t > \eta_{n-1} : X_t^\varepsilon \notin V_k\}$, $n \ge 1$. In order to prove (2.10), it suffices to prove instead, uniformly in x in any compact set in $\mathbb{R}^2$ and in $y \in \mathbb{T}^m$, that

$$\mathbf{E}_{(x,y)}[f(h(X_{T\wedge\eta_1}^\varepsilon)) - f(h(x)) - \int_0^{T\wedge\eta_1} \mathcal{L}f(h(X_t^\varepsilon))dt] \to 0, \quad \text{as } \varepsilon \downarrow 0, \tag{2.11}$$

since it also implies that $\mathbf{P}(\eta_n < T) \to 0$ as $n \to \infty$, uniformly in all ε sufficiently small. In the unbounded domain without critical points, (2.11) can be obtained using the result in bounded domain together with the tightness of $h(X_t^\varepsilon)$. It remains to consider the bounded domains. Let V be one of the bounded domains. As explained below, the process X_t^ε in V can be extended beyond the time when it reaches the boundary by embedding V into a compact manifold M with an area form and a Hamiltonian such that there are no other separatrices.

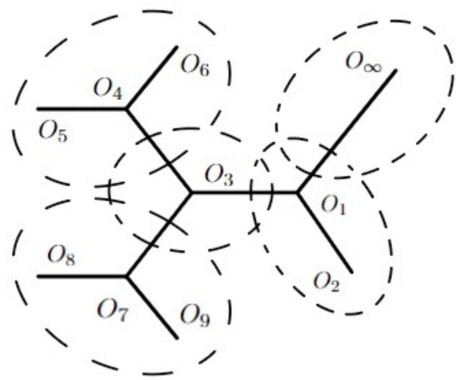
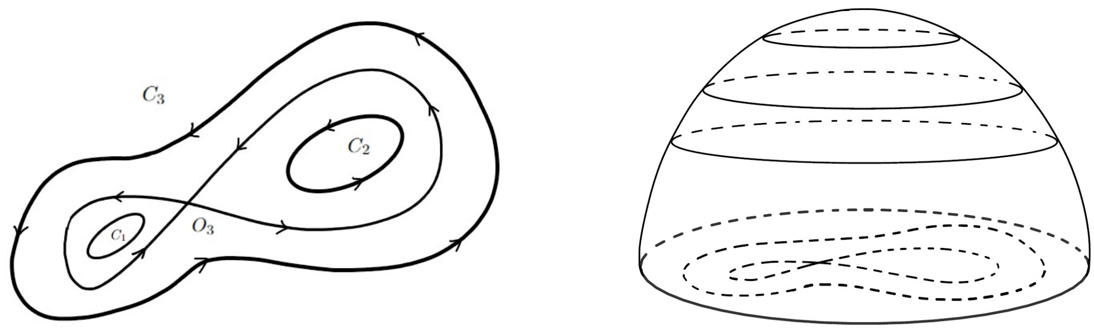


Figure 2.3: Domains projected on the graph.

Consider the case where V is not simply connected - for example, V is the domain that contains O_3 in Figure 2.3. There are three connected components of $\mathbb{R}^2 \setminus V$ as shown in Figure 2.4a (other situations can be treated similarly). Then we can modify the Hamiltonian and the vector field in C_1 and C_2 in such a way that assumptions *H(2)-H(4)* hold locally, and there is only one extremum point of H in C_1 and one in C_2 (this modification is not needed if V is simply connected). The unbounded domain C_3 outside V can be replaced by a compact surface S so that the resulting state space of X_t^ε is, topologically, a sphere $M = V \bigcup C_1 \bigcup C_2 \bigcup S$. Then, the vector field on the surface can be chosen as a smooth extension from V so that the averaged process is a Hamiltonian system on M with respect to an area form ω, which is simply $dx_1 \wedge dx_2$ on V, C_1, and C_2. Moreover, there exists a chart (S, Φ) such that the corresponding vector field $b(x, y)$ on $D := \Phi(S)$ satisfies that $\{b(x, y) - \bar{b}(x) : y \in \mathbb{T}^m\}$ spans $\mathbb{R}^2$ for each $x \in D$ and the averaged process is a Hamiltonian system with respect to $dx_1 \wedge dx_2$ on D. For example, as shown in Figure 2.4b, we can modify the vector field on the plane outside V so that there are two disks with the same center $D_1 \subset D_2$, and the averaged process is a Hamiltonian system in D_2, in particular, rotation between ∂D_1 and ∂D_2. Then D_2 is smoothly glued to a hemisphere and the resulting manifold is M, and the vector field can be extended to the surface in such a way that the averaged process is a rotation with certain constant angular velocity on the level sets.



(a) Three connected components of $\mathbb{R}^2 \setminus V$. (b) Hamiltonian system on manifold M.

Figure 2.4: Localization.

It is clear that, when restricted to $V \times \mathbb{T}^m$, the resulting system defined on $M \times \mathbb{T}^m$ has exactly the same behavior as the original process on $\mathbb{R}^2 \times \mathbb{T}^m$. Therefore, it suffices to prove (2.11) for the new process on $M \times \mathbb{T}^m$. Let us formally restate the corresponding assumptions and formulate the result on $M \times \mathbb{T}^m$. In the remainder of the chapter, all the definitions (e.g. the quantities defined in Section 2.2) and statements on M are understood by locally choosing coordinates so that $\omega = dx_1 \wedge dx_2$. In particular, on S, they are understood in the coordinate Φ, while on the "flat" part that contains V, they are understood in the usual way. Then the assumptions on the coefficients on $M \times \mathbb{T}^m$ are analogous to those introduced earlier, so we only mention the differences:

(H2′) $H(x)$ is a C^∞ function from M to $\mathbb{R}$ that has three extremum points and one saddle point.

(H3′) $b(x, y)$ is a C^∞ function from $M \times \mathbb{T}^m$ to TM such that $\bar{b}(x) = \nabla^\perp H(x)$.

(H4′) $\{b(x, y) - \bar{b}(x) : y \in \mathbb{T}^m\}$ spans TM for all $x \in M$.

From this point on, we denote the process on $M \times \mathbb{T}^m$ as $(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)$, and $(X_t^\varepsilon, \xi_t^\varepsilon)$ on the time scale $O(\varepsilon^{-1})$ (defined by (2.1) and (2.4) with $\mathbb{R}^2$ replaced by M), and assume that the conditions *(H2′)-(H4′)* replacing *(H2)-(H4)* hold. Then (2.10) follows from

Proposition 2.3.1. *For each $f \in \mathcal{D}_{\mathcal{L}}$ and each $T > 0$,*

$$\mathbf{E}_{(x,y)}[f(h(X_\eta^\varepsilon)) - f(h(x)) - \int_0^\eta \mathcal{L}f(h(X_t^\varepsilon))dt] \to 0, \tag{2.12}$$

as $\varepsilon \to 0$, uniformly in $x \in M$, $y \in \mathbb{T}^m$, and $\eta \leq T$ that is a stopping time w.r.t. $\mathcal{F}_t^{X^\varepsilon}$.

2.3.2 Auxiliary process

It turns out that similar results hold for a more general process with a slightly perturbed fast motion:

$$\begin{aligned} d\tilde{\boldsymbol{X}}_t^\varepsilon &= b(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)dt, \\ d\tilde{\boldsymbol{\xi}}_t^\varepsilon &= \frac{1}{\varepsilon}v(\tilde{\boldsymbol{\xi}}_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}}\sigma(\tilde{\boldsymbol{\xi}}_t^\varepsilon)dW_t + c(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)dt, \end{aligned} \tag{2.13}$$

where $c(x, y)$ is infinitely differentiable. Namely, $h(\tilde{\boldsymbol{X}}_{t/\varepsilon}^\varepsilon)$ converges weakly to the Markov process defined by the operator $(\mathcal{L}_c, D(\mathcal{L}_c))$ on the Reeb graph. Here, the subscript c indicates that $\mathcal{L}_c$ depends on the choice of $c(x, y)$. If $c(x, y) = 0$, then it is clear that $(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon) = (\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)$, and thus $\mathcal{L} = \mathcal{L}_c$. However, if $c(x, y) \neq 0$, then we have an additional drift term in (2.13), and thus we need an additional drift term in the generator of the limiting process. While a precise definition of $(\mathcal{L}_c, D(\mathcal{L}_c))$ is deferred to later sections, we observe that the operators replacing L_k depend on $c(x, y)$, and the domain $D(\mathcal{L}_c)$ as well as the linear subspace, denoted by $\mathcal{D}_{\mathcal{L}_c}$, chosen in Lemma 2.2.4 also vary for different $c(x, y)$. Therefore, in order to formulate general results, we consider $\mathcal{D}$, the set of continuous functions on $\mathbb{G}$ that are four-times continuously differentiable inside each edge and satisfy conditions (i) and (iii) in Definition 2.2.1, as well as a weaker form of condition (ii), namely, the limits $\lim_{h_k \to O_i} L_k f(h_k)$ exist but are not necessarily independent of the edge I_k. Note that $\mathcal{D}$ contains $\mathcal{D}_{\mathcal{L}_c}$ for all choices of $c(x,y)$. Define $\mathcal{L}_c$ on $\mathcal{D}$ by applying the differential operator (L_k plus an additional drift corresponding to $c(x,y)$) on each edge separately, with the result not being necessarily continuous at the interior vertices.

As mentioned before, we need to construct a family of auxiliary processes that, on the one hand, have a common invariant measure for all $\varepsilon > 0$ and, on the other hand, are very close to the processes of interest. The auxiliary process on M can in fact be obtained by choosing a special $c(x,y)$ in (2.13). We denote this particular choice of $c(x,y)$ as $\tilde{c}(x,y)$. Now we find $\tilde{c}(x,y)$ such that $\lambda \times \mu$ is the invariant measure for the process with every ε, where λ is the area measure w.r.t. ω and μ is the invariant measure for $\boldsymbol{\xi}_t^\varepsilon$ in $\mathbb{T}^m$. Let $\tilde{L}^\varepsilon$ be the generator of the process $(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)$:

$$\tilde{L}^\varepsilon f(x,y) = b(x,y) \cdot \nabla_x f(x,y) + \tilde{c}(x,y) \cdot \nabla_y f(x,y) + \frac{1}{\varepsilon} L f(x,y).$$

Hence, $\lambda \times \mu$ is the invariant measure if $\tilde{L}^{\varepsilon *} p(y) = 0$, where $\tilde{L}^{\varepsilon *}$ is the adjoint operator of $\tilde{L}^\varepsilon$ and p is the density of μ, i.e.

$$0 = \tilde{L}^{\varepsilon *} p(y) = -\operatorname{div}_x(b(x,y)p(y)) - \operatorname{div}_y(\tilde{c}(x,y)p(y)) + \frac{1}{\varepsilon} L^* p(y), \tag{2.14}$$

where L^* is the adjoint operator of L. Since μ is the invariant measure for $\boldsymbol{\xi}_t^\varepsilon$, the last term vanishes. Hence (2.14) reduces to

$$\operatorname{div}_x b(x,y)p(y) + \operatorname{div}_y(\tilde{c}(x,y)p(y)) = 0. \tag{2.15}$$

To see the existence of the solution, we need the following lemma (cf. Lemma 2.1 in [26]).

Lemma 2.3.2. *Let $\tilde{g}(x,y)$ be a bounded function on $\mathbb{R}^2 \times \mathbb{T}^m$ that is infinitely differentiable, and let $\tilde{L}$ be the generator of a non-degenerate diffusion on $\mathbb{T}^m$ with the unique invariant measure $\tilde{\mu}$ and suppose that $\int_{\mathbb{T}^m} \tilde{g}(x,y)d\tilde{\mu}(y) = 0$ for each $x \in \mathbb{R}^2$. Then there exists a unique solution $\tilde{u}(x,y)$ to the equation*

$$\tilde{L}\tilde{u}(x,y) = -\tilde{g}(x,y), \quad \int_{\mathbb{T}^m} \tilde{u}(x,y)d\tilde{\mu}(y) = 0, \tag{2.16}$$

and $\tilde{u}(x,y)$ is also bounded and infinitely differentiable. Moreover, if $\tilde{g}(x,y)$ has uniformly bounded derivatives up to order K in x (or y), the same holds for $\tilde{u}(x,y)$.

Remark 2.3.3. *The same result also holds for functions on $M \times \mathbb{T}^m$. Thus, the existence of the solution to (2.15) immediately follows from Lemma 2.3.2 applied to $\tilde{g}(x,y) = \mathrm{div}_x b(x,y)p(y)$ and $\tilde{L} = \Delta_y$, and by taking the gradient of the solution in (2.16) w.r.t. y, and dividing it by $p(y)$.*

As in (2.4), we define $(\tilde{X}^\varepsilon_t, \tilde{\xi}^\varepsilon_t) = (\tilde{\boldsymbol{X}}^\varepsilon_{t/\varepsilon}, \tilde{\boldsymbol{\xi}}^\varepsilon_{t/\varepsilon})$ in distribution. Then, a simple corollary can be obtained by using Lemma 2.3.2 and then applying Ito's formula to the corresponding solution $\tilde{u}(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t)$ and $\tilde{u}(\tilde{X}^\varepsilon_t, \tilde{\xi}^\varepsilon_t)$ (cf. Lemma 2.3 in [26]).

Corollary 2.3.4. *Let $\tilde{g}$ satisfy the all the conditions in Lemma 2.3.2 with $\tilde{L} = L$ and $K = 1$, then for fixed $T > 0$*

$$\mathbf{E}_{(x,y)} \left| \int_0^\eta \tilde{g}(\tilde{\boldsymbol{X}}^\varepsilon_s, \tilde{\boldsymbol{\xi}}^\varepsilon_s) ds \right| = O(\sqrt{\varepsilon}), \quad \mathbf{E}_{(x,y)} \left| \int_0^\eta \tilde{g}(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) ds \right| = O(\varepsilon),$$

uniformly in $x \in M$, $y \in \mathbb{T}^m$, and η that is a stopping time bounded by T.

2.3.3 Diffusion approximation

Since $\int_{\mathbb{T}^m}(b(x,y)-\bar{b}(x))d\mu(y) = 0$, by Lemma 2.3.2, there exists a function u that is bounded together with its derivatives such that

$$Lu(x,y) = -(b(x,y)-\bar{b}(x)). \tag{2.17}$$

The equation is understood element-wise. Apply Ito's formula to $u(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)$:

$$\begin{aligned} u(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon) &= u(x_0,y_0) + \frac{1}{\varepsilon}\int_0^t Lu(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \frac{1}{\sqrt{\varepsilon}}\int_0^t \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s \\ &\quad + \int_0^t [\nabla_x u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)b(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)c(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds. \end{aligned} \tag{2.18}$$

Combining (2.13), (2.17), and (2.18), we obtain

$$\begin{aligned} \tilde{\boldsymbol{X}}_t^\varepsilon &= x_0 + \int_0^t \nabla^\perp H(\tilde{\boldsymbol{X}}_s^\varepsilon)ds + \varepsilon\int_0^t [\nabla_x u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)b(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)c(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds \\ &\quad + \sqrt{\varepsilon}\int_0^t \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s + \varepsilon(u(x_0,y_0) - u(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)). \end{aligned} \tag{2.19}$$

Similarly, by applying Ito's formula to $u(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ and repeating the steps above, we have

$$\begin{aligned} \tilde{X}_t^\varepsilon &= x_0 + \frac{1}{\varepsilon}\int_0^t \nabla^\perp H(\tilde{X}_s^\varepsilon)ds + \int_0^t [\nabla_x u(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)]ds \\ &\quad + \int_0^t \nabla_y u(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)\sigma(\tilde{\xi}_s^\varepsilon)dW_s + \varepsilon(u(x_0,y_0) - u(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)). \end{aligned} \tag{2.20}$$

This idea of diffusion approximation is frequently used in the remainder of the chapter, and the function $u(x,y)$ always refers to the solution to (2.17).

2.4 Averaging principle inside one domain

In this section, we consider a general process $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ defined after Remark 2.3.3, which is a faster version of the process in (2.13). The process takes values on $M \times \mathbb{T}^m$. As a result of localization, M is separated into three domains, each bounded by the separatrix or a part of it. This section is devoted to the proof of the averaging principle for $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ on $M \times \mathbb{T}^m$ up to time when $\tilde{X}_t^\varepsilon$ exits from one of the three domains. The domain under consideration will be denoted by U. Therefore, without any ambiguity, the projection h simply reduces to the Hamiltonian H. Let $U(h_1, h_2)$ be the region in U between $\gamma(h_1)$ and $\gamma(h_2)$, O be the saddle point, and O' be the extremum point, and further define stopping times $\tau(h) = \inf\{t : |H(\tilde{X}_t^\varepsilon) - H(O)| = h\}$ and $\eta(h) = \inf\{t : |H(\tilde{X}_t^\varepsilon) - H(O')| = h\}$. Without loss of generality, we assume that $H(O) = 0$ and $H(O') = 1$.

2.4.1 Averaging principle before $\tau(\varepsilon^\alpha) \wedge \eta(\delta)$

We aim to prove the averaging principle between $\gamma(\varepsilon^\alpha)$ and $\gamma(1-\delta)$ with constants $0 < \alpha < 1/4$ and $0 < \delta < 1$. Notice that, for technical reasons, we assume that $0 < \alpha < 1/4$ in this intermediate result and in the proofs that utilize it in this subsection and the next, while we always assume that $0 < \alpha < 1/2$ elsewhere. Let us further define another coordinate ϕ inside this domain U. Let l denote the curve that is tangent to ∇H at each point and connects the saddle point O and the extremum point O', and let $l(h)$ be the intersection of l and $\gamma(h)$. Let $Q(h)$ denote the time it takes for the averaged process $\boldsymbol{x}_t$ to make one rotation on $\gamma(h)$ and $q(x)$ denote the time it takes for $\boldsymbol{x}_t$ starting from $l(H(x))$ to arrive at x. Now we define the coordinate $\phi(x) = q(x)/Q(H(x))$ whose range is $S^1 := \mathbb{R} \pmod 1$. It is easy to see that $\boldsymbol{x}_t$ has constant speed $1/Q(H(\boldsymbol{x}_t))$ in ϕ

coordinate. Since there is logarithmic delay near the saddle point, the coordinate ϕ has exploding derivatives near the separatrix. However, as shown in Section 2.7.1, the order of its derivatives w.r.t. the Euclidean coordinates is under control. Let us denote $\tilde{H}_t^\varepsilon = H(\tilde{X}_t^\varepsilon)$ and $\tilde{\Phi}_t^\varepsilon = \phi(\tilde{X}_t^\varepsilon)$. Along the same lines leading to (2.20), we have the following equations with $u_h = u \cdot \nabla H$, $u_\phi = u \cdot \nabla \phi$, $h_0 = H(x_0)$, and $\phi_0 = \phi(x_0)$:

$$
\begin{aligned}
\tilde{H}_t^\varepsilon = h_0 &+ \int_0^t \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\intercal \sigma(\tilde{\xi}_s^\varepsilon) dW_s + \varepsilon(u_h(x_0, y_0) - u_h(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)) \\
&+ \int_0^t [\nabla_x u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)] ds,
\end{aligned} \tag{2.21}
$$

$$
\begin{aligned}
\tilde{\Phi}_t^\varepsilon = \phi_0 &+ \int_0^t \nabla_y u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\intercal \sigma(\tilde{\xi}_s^\varepsilon) dW_s + \frac{1}{\varepsilon} \int_0^t \frac{1}{Q(\tilde{H}_s^\varepsilon)} ds \\
&+ \int_0^t [\nabla_x u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)] ds \\
&+ \varepsilon(u_\phi(x_0, y_0) - u_\phi(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)),
\end{aligned} \tag{2.22}
$$

for $\varepsilon^\alpha \leq h_0 \leq 1 - \delta$ and $t \leq \tau(\varepsilon^\alpha) \wedge \eta(\delta)$. Define the following coefficients using the original coordinates for all $x \in M$:

$$
\begin{aligned}
A(x) &= \int_{\mathbb{T}^m} |\nabla_y u_h(x, y)^\intercal \sigma(y)|^2 d\mu(y), \\
B_c(x) &= \int_{\mathbb{T}^m} [\nabla_x u_h(x, y) \cdot b(x, y) + \nabla_y u_h(x, y) \cdot c(x, y)] d\mu(y);
\end{aligned} \tag{2.23}
$$

and (h, ϕ) coordinates for $x = (h, \phi)$, where $\varepsilon^\alpha \leq h \leq 1 - \delta$ and $\phi \in S^1$:

$$
\begin{aligned}
A(h, \phi) &= A(x), & \bar{A}(h) &= \int_{S^1} A(h, \phi) d\phi, \\
B_c(h, \phi) &= B_c(x), & \bar{B}_c(h) &= \int_{S^1} B_c(h, \phi) d\phi.
\end{aligned} \tag{2.24}
$$

Define $\mathcal{L}_c$ by $\mathcal{L}_c f = \frac{1}{2}\bar{A}f'' + \bar{B}_c f'$ for $f \in \mathcal{D}$ in the interior of each edge. In particular, when $c(x,y) = 0$, this definition is consistent with that in (2.9). Introduce two processes close to $\tilde{H}_t^\varepsilon, \tilde{\Phi}_t^\varepsilon$:

$$
\begin{aligned}
\hat{H}_t^\varepsilon &= h_0 + \int_0^t \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\tilde{\xi}_s^\varepsilon) dW_s \\
&\quad + \int_0^t [\nabla_x u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)] ds, \qquad (2.25) \\
\hat{\Phi}_t^\varepsilon &= \phi_0 + \int_0^t \nabla_y u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\tilde{\xi}_s^\varepsilon) dW_s + \frac{1}{\varepsilon}\int_0^t \frac{1}{Q(\tilde{H}_s^\varepsilon)} ds \\
&\quad + \int_0^t [\nabla_x u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_\phi(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)] ds. \qquad (2.26)
\end{aligned}
$$

Let $\tau_0 = \tau(\varepsilon^\alpha) \wedge \eta(\delta)$. For each $f \in \mathcal{D}$, $x \in U(\varepsilon^\alpha, 1-\delta)$, $y \in \mathbb{T}^m$, and stopping time $\sigma' \leq T \wedge \tau_0$, by Ito's formula applied to $f(\hat{H}_{\sigma'}^\varepsilon)$, we have

$$
\begin{aligned}
\mathbf{E}_{(x,y)} f(\hat{H}_{\sigma'}^\varepsilon) = f(H(x)) + \mathbf{E}_{(x,y)} \int_0^{\sigma'} \Big(&\frac{1}{2} |\nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\tilde{\xi}_s^\varepsilon)|^2 f''(\hat{H}_s^\varepsilon) \\
&+ \left[\nabla_x u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)\right] f'(\hat{H}_s^\varepsilon) \Big) ds.
\end{aligned}
$$

Since $\sup_{0 \leq t \leq \sigma'} |\tilde{H}_t^\varepsilon - \hat{H}_t^\varepsilon| = O(\varepsilon)$,

$$
\begin{aligned}
\mathbf{E}_{(x,y)} f(\tilde{H}_{\sigma'}^\varepsilon) = f(H(x)) + \mathbf{E}_{(x,y)} \int_0^{\sigma'} \Big(&\frac{1}{2} |\nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\tilde{\xi}_s^\varepsilon)|^2 f''(\tilde{H}_s^\varepsilon) \\
&+ \left[\nabla_x u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot b(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) + \nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon) \cdot c(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)\right] f'(\tilde{H}_s^\varepsilon) \Big) ds + O(\varepsilon).
\end{aligned} \qquad (2.27)
$$

Combining this with (2.23) and (2.24), by Lemma 2.3.2, as in Corollary 2.3.4, we have

$$\mathbf{E}_{(x,y)}\left[f(\tilde{H}^\varepsilon_{\sigma'}) - f(H(x)) - \int_0^{\sigma'}\left(\frac{1}{2}A(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s)f''(\tilde{H}^\varepsilon_s) + B_c(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s)f'(\tilde{H}^\varepsilon_s)\right)ds\right] = O(\varepsilon). \tag{2.28}$$

Lemma 2.4.1. *Let $g(h,\phi)$ be either $A(h,\phi)f''(h)$ or $B_c(h,\phi)f'(h)$, and $\bar{g}(h) = \int_{S^1} g(h,\phi)d\phi$. Then, for every $T > 0$,*

$$\sup_{\substack{x\in U(\varepsilon^\alpha,1-\delta)\\ y\in\mathbb{T}^m}}\ \sup_{\sigma'\le T\wedge\tau_0} \mathbf{E}_{(x,y)}\left|\int_0^{\sigma'}\left[g(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s) - \bar{g}(\tilde{H}^\varepsilon_s)\right]ds\right| \to 0, \ \ \textit{as } \varepsilon\downarrow 0,$$

where the first supremum is taken over all stopping times $\sigma' \le T\wedge\tau_0$.

Proof. Fix $\kappa > 0$. Since, for fixed h, $g(h,\phi) - \bar{g}(h)$ is a function on S^1, we can approximate it by a finite sum of its Fourier series with error less than $\frac{\kappa}{2T}$:

$$g(h,\phi) - \bar{g}(h) \approx \sum_{0<|k|\le K(\varepsilon)} g_k(h,\phi) := \sum_{0<|k|\le K(\varepsilon)} G_k(h)\exp(2\pi i k\phi),$$

for all $\varepsilon^\alpha \le h \le 1-\delta$ and $\phi\in S^1$, where

$$G_k(h) = \int [g(h,\phi) - \bar{g}(h)]\exp(-2\pi i k\phi)d\phi.$$

Since, as shown in Section 2.7.1, $g''_{\phi\phi} = O(|\log h|/h)$, we see that $K(\varepsilon)$ can be chosen as $\varepsilon^{-\alpha}|\log\varepsilon|^2$ for sufficiently small ε. Then it suffices to prove that, for all $0 < |k| \le K(\varepsilon)$

and ε sufficiently small,

$$\sup_{\substack{x\in U(\varepsilon^\alpha,1-\delta)\\ y\in\mathbb{T}^m}}\sup_{\sigma'\le T\wedge\tau_0}\mathbf{E}_{(x,y)}\left|\int_0^{\sigma'} g_k(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s)ds\right| = o\left(\frac{\varepsilon^\alpha}{|\log\varepsilon|^2}\right). \tag{2.29}$$

We define an auxiliary function v for fixed g_k, where $0<|k|\le K(\varepsilon)$:

$$v=\frac{g_k(h,\phi)Q(h)}{2\pi ik},$$

which satisfies that $v'_\phi/Q(h)=g_k(h,\phi)$. We formulate the bounds on ϕ, v, g, and their derivatives, uniformly in all $\varepsilon^\alpha<h<1-\delta$ and $0<|k|\le K(\varepsilon)$ (proved in the Section 2.7.1):

$$\begin{aligned}
&\phi\in[0,1),\ \nabla\phi=O(1/h),\ \nabla^2\phi=O(1/h^2),\\
&v=O(|\log h|),\ v'_\phi=O(|\log h|),\ v''_{\phi\phi}=O(|\log h|^3/h),\\
&v'_h=O(|\log h|^2/h),\ v''_{hh}=O(|\log h|^3/h^3),\ v''_{\phi h}=O(|\log h|^2/h),\\
&g'_h=O(|\log h|/h),\ g''_{hh}=O(|\log h|^2/h^3).
\end{aligned}\tag{2.30}$$

By comparing $(\tilde{H}^\varepsilon_t,\tilde{\Phi}^\varepsilon_t)$ and $(\hat{H}^\varepsilon_t,\hat{\Phi}^\varepsilon_t)$ in (2.21), (2.22), (2.25), and (2.26), and using the bounds in (2.30), we know that for all $\sigma'\le T\wedge\tau_0$,

$$\int_0^{\sigma'}\left|g_k(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s)-\frac{v'_\phi(\hat{H}^\varepsilon_s,\hat{\Phi}^\varepsilon_s)}{Q(\tilde{H}^\varepsilon_s)}\right|ds=\int_0^{\sigma'}\left|\frac{v'_\phi(\tilde{H}^\varepsilon_s,\tilde{\Phi}^\varepsilon_s)-v'_\phi(\hat{H}^\varepsilon_s,\hat{\Phi}^\varepsilon_s)}{Q(\tilde{H}^\varepsilon_s)}\right|ds=O(\varepsilon^{1-2\alpha}|\log\varepsilon|^3). \tag{2.31}$$

Apply Ito's formula to $v(\hat{H}^\varepsilon_{\sigma'},\hat{\Phi}^\varepsilon_{\sigma'})$ and obtain

$$\frac{1}{\varepsilon}\int_0^{\sigma'}\frac{v'_\phi(\hat{H}^\varepsilon_s,\hat{\Phi}^\varepsilon_s)}{Q(\tilde{H}^\varepsilon_s)}ds$$

$$
\begin{aligned}
&= v(\hat{H}^\varepsilon_{\sigma'}, \hat{\Phi}^\varepsilon_{\sigma'}) - v(H(x), \phi(x)) - \int_0^{\sigma'} v'_h(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)\nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s) dW_s \\
&\quad - \int_0^{\sigma'} v'_h(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)[\nabla_x u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot b(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) + \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot c(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)] ds \\
&\quad - \frac{1}{2}\int_0^{\sigma'} v''_{hh}(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)|\nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s)|^2 ds \\
&\quad - \int_0^{\sigma'} v'_\phi(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)\nabla_y u_\phi(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s) dW_s \\
&\quad - \int_0^{\sigma'} v'_\phi(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)[\nabla_x u_\phi(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot b(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) + \nabla_y u_\phi(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot c(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)] ds \\
&\quad - \frac{1}{2}\int_0^{\sigma'} v''_{\phi\phi}(\hat{H}^\varepsilon_s, \hat{\Phi}^\varepsilon_s)|\nabla_y u_\phi(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s)|^2 ds \\
&\quad - \int_0^{\sigma'} v''_{\phi h} \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s)\sigma(\tilde{\xi}^\varepsilon_s)^\mathsf{T} u_\phi(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) ds.
\end{aligned}
$$

By using the estimates in (2.30) and the fact that $0 < \alpha < 1/4$, we know that the expectation of the right-hand side is $o(\frac{\varepsilon^{\alpha-1}}{|\log\varepsilon|^2})$. Combining this with (2.31), we get (2.29). Thus, the desired result follows. □

Now, applying Lemma 2.4.1 to (2.28), we get

Lemma 2.4.2. *For each $f \in \mathcal{D}$, $0 < \alpha < 1/4$, and $0 < \delta < 1$, as $\varepsilon \downarrow 0$,*

$$
\sup_{\substack{x \in U(\varepsilon^\alpha, 1-\delta) \\ y \in \mathbb{T}^m}} \sup_{\sigma' \le T \wedge \eta(\delta) \wedge \tau(\varepsilon^\alpha)} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s)) ds]| \to 0,
$$

where the first supremum is taken over all stopping times $\sigma' \le T \wedge \eta(\delta) \wedge \tau(\varepsilon^\alpha)$.

Remark 2.4.3. *The diffusion process governed by $\mathcal{L}_c$ can reach all points inside the edge and the interior vertex but cannot reach the exterior vertex. For example, in the case considered here, the process can reach all points in $[0, 1)$ but cannot reach 1. The reason is that, for each $\delta > 0$, on $[0, 1-\delta]$, $\bar{B}_c(h)$ is bounded and $1/\bar{A}(h) \lesssim |\log h|$ (see Section 2.7.1 for details). However, for*

$\kappa > 0$ sufficiently small, on $[1-\kappa, 1]$, B_c is uniformly negative while $A(h) \lesssim 1-h$ due to the non-degeneracy of the maximum point (see Lemma 2.7.1 for details).

Lemma 2.4.4. *For each $\delta > 0$ and $0 < \alpha < 1/4$, $\mathbf{E}_{(x,y)}(\eta(\delta) \wedge \tau(\varepsilon^\alpha))$ in uniformly bounded for all $x \in U(\varepsilon^\alpha, 1-\delta)$, $y \in \mathbb{T}^m$, and ε sufficiently small.*

Proof. The solution f^δ to the following equation exists on $[0, 1-\delta]$ due to Remark 2.4.3:

$$\begin{cases} \mathcal{L}_c f^\delta = -1 \\ f^\delta(0) = f^\delta(1-\delta) = 0 \end{cases}$$

Let $\tilde{T} > 3\|f^\delta\|_{\sup}$, then Lemma 2.4.2 implies that, for all $x \in U(\varepsilon^\alpha, 1-\delta)$, $y \in \mathbb{T}^m$, and ε small enough,

$$\mathbf{E}_{(x,y)}(\eta(\delta) \wedge \tau(\varepsilon^\alpha) \wedge \tilde{T}) < \tilde{T}/2.$$

Thus, by Markov inequality and strong Markov property, $\mathbf{E}_{(x,y)}(\eta(\delta) \wedge \tau(\varepsilon^\alpha)) \leq 2\tilde{T}$. □

Lemma 2.4.5. *For each $f \in \mathcal{D}$, $\delta > 0$, and $0 < \alpha < 1/4$, as $\varepsilon \downarrow 0$,*

$$\sup_{\substack{x \in U(\varepsilon^\alpha, 1-\delta) \\ y \in \mathbb{T}^m}} \sup_{\sigma' \leq \eta(\delta) \wedge \tau(\varepsilon^\alpha)} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s)) ds]| \to 0,$$

where the first supremum is taken over all stopping times $\sigma' \leq \eta(\delta) \wedge \tau(\varepsilon^\alpha)$.

Proof. The result can be deduced from Lemma 2.4.2 and Lemma 2.4.4 by choosing a sufficiently large T and using the Markov property. □

2.4.2 Averaging principle before σ

With estimates on the transition times and probabilities between level sets near the critical points, the result from last section can be extended to the time when the process reaches the separatrix, which is the stopping time σ defined earlier. The main result of this subsection is

Proposition 2.4.6. *For each $f \in \mathcal{D}$, as $\varepsilon \downarrow 0$,*

$$\sup_{x\in U, y\in\mathbb{T}^m} \sup_{\sigma'\leq\sigma} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]| \to 0, \tag{2.32}$$

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

We state a simple corollary of Proposition 2.7.3.

Corollary 2.4.7. *For each $0 < \alpha < 1/2$, uniformly in $x \in U(0, \varepsilon^\alpha)$ and $y \in \mathbb{T}^m$,*

$$\mathbf{E}_{(x,y)}\sigma \wedge \tau(\varepsilon^\alpha) = O(\varepsilon^{2\alpha}|\log\varepsilon|).$$

Lemma 2.4.8. *For each $0 < \alpha < 1/2$, uniformly in $x \in U(0, \varepsilon^\alpha)$ and $y \in \mathbb{T}^m$,*

$$|\mathbf{P}_{(x,y)}(\tau(\varepsilon^\alpha) < \sigma) - H(x)\varepsilon^{-\alpha}| = O(\varepsilon^\alpha|\log\varepsilon|).$$

Proof. As in (2.21), write the equation for $H(\tilde{X}^\varepsilon_t) = \tilde{H}^\varepsilon_t$ stopped at $\sigma \wedge \tau(\varepsilon^\alpha)$,

$$\begin{aligned} H(\tilde{X}^\varepsilon_{\sigma\wedge\tau(\varepsilon^\alpha)}) = H(x) + \int_0^{\sigma\wedge\tau(\varepsilon^\alpha)} \nabla_x u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot b(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) + \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot c(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)ds \\ + \int_0^{\sigma\wedge\tau(\varepsilon^\alpha)} \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\mathsf{T} \sigma(\tilde{\xi}^\varepsilon_s)dW_s + O(\varepsilon). \end{aligned}$$

From Corollary 2.4.7, it follows that

$$|\mathbf{P}_{(x,y)}(\tau(\varepsilon^\alpha) < \sigma) - H(x)\varepsilon^{-\alpha}| = \varepsilon^{-\alpha}|\mathbf{E}_{(x,y)}H(\tilde{X}^\varepsilon_{\sigma\wedge\tau(\varepsilon^\alpha)}) - H(x)| = O(\varepsilon^\alpha|\log\varepsilon|). \qquad \square$$

We prove that the process spends finite time (in expectation) inside U. The idea is to use the fact that the process on the graph spends little time near the vertices and exits the edge with positive probability once it gets close enough to the interior vertex.

Lemma 2.4.9. *For each $0 < \alpha < 1/4$, $\mathbf{E}_{(x,y)}\tau(\varepsilon^\alpha)$ is uniformly bounded for all $x \in U$ such that $H(x) \geq \varepsilon^\alpha$, $y \in \mathbb{T}^m$, and ε sufficiently small.*

Proof. By Lemma 2.7.2, fix $\delta > 0$ such that $\mathbf{E}_{(x,y)}\eta(2\delta) \leq 1$ for all ε sufficiently small and all x satisfying $H(x) > 1 - 2\delta$; By Lemma 2.4.5, fix $\kappa > 0$ such that $\mathbf{P}_{(x,y)}(\eta(\delta) < \tau(\varepsilon^\alpha)) < 1 - \kappa$ all x satisfying $H(x) = 1 - 2\delta$, all $y \in \mathbb{T}^m$, and all ε sufficiently small; By Lemma 2.4.4, fix $T > 4(1 + \sup_{x\in U(\varepsilon^\alpha,1-\delta), y\in\mathbb{T}^m} \mathbf{E}_{(x,y)}(\tau(\varepsilon^\alpha) \wedge \eta(\delta)))/\kappa$. For all x with $H(x) > 1 - 2\delta$ and $y \in \mathbb{T}^m$,

$$\begin{aligned}
&\mathbf{P}_{(x,y)}(\tau(\varepsilon^\alpha) > 2T) \\
&\leq \mathbf{P}_{(x,y)}(\eta(2\delta) > T) + \sup_{(x',y')\in\gamma(1-2\delta)\times\mathbb{T}^m} \big(\mathbf{P}_{(x',y')}(\tau(\varepsilon^\alpha)\wedge\eta(\delta) > T) + \mathbf{P}_{(x',y')}(\eta(\delta) < \tau(\varepsilon^\alpha))\big) \\
&\leq 1 - \kappa/2.
\end{aligned}$$

For all $x \in U$ with $\varepsilon^\alpha \leq H(x) \leq 1 - 2\delta$, the estimate above holds without the first term on the second line. Then the uniform boundedness follows from the Markov property. $\square$

We can apply the similar idea near the separatrix. Namely, we choose $0 < \alpha' < \alpha < 1/4$. By Corollary 2.4.7, the process spent little time spent between γ and $\gamma(\varepsilon^{\alpha'})$, and by Lemma 2.4.8, the process is very likely to exit through the separatrix rather than come back to $\gamma(\varepsilon^{\alpha'})$ once it reaches $\gamma(\varepsilon^\alpha)$. Then, with the fact that $\mathbf{E}_{(x,y)}\tau(\varepsilon^\alpha)$ is uniformly bounded, one can prove the following result:

Lemma 2.4.10. $\mathbf{E}_{(x,y)}\sigma$ *is uniformly bounded for all* $x \in U$, $y \in \mathbb{T}^m$, *and* ε *sufficiently small.*

Proof of Proposition 2.4.6. Fix $\kappa > 0$ and $0 < \alpha < 1/4$. By Lemma 2.4.9, let T be large enough such that $\mathbf{P}_{(x,y)}(\tau(\varepsilon^\alpha) > T) < \kappa$ for all $x \in U$ satisfying $H(x) \geq \varepsilon^\alpha$, $y \in \mathbb{T}^m$, and ε sufficiently small. By Lemma 2.7.2, let $\delta > 0$ small enough such that for ε sufficiently small

$$\sup_{\substack{x\in U: H(x)\geq 1-\delta \\ y\in\mathbb{T}^m}} \sup_{\sigma'\leq\eta(\delta)} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]| < \kappa, \tag{2.33}$$

where the first supremum is taken over all stopping times $\sigma' \leq \eta(\delta)$. By Remark 2.4.3 and Lemma 2.4.5, let $\delta' > 0$ small enough such that $\mathbf{P}_{(x,y)}(\eta(\delta') < \tau(\varepsilon^\alpha)) < \kappa$ for all $x \in U(\varepsilon^\alpha, 1-\delta)$, $y \in \mathbb{T}^m$, and ε sufficiently small. For stopping time $\sigma' \leq \tau(\varepsilon^\alpha)$, $x \in U(\varepsilon^\alpha, 1-\delta)$, and $y \in \mathbb{T}^m$,

$$\begin{aligned}
&|\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'}) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]| \\
&\leq |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\eta(\delta')\wedge\sigma'}) - f(H(x)) - \int_0^{\eta(\delta')\wedge\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds] \\
&\quad + \mathbf{P}_{(x,y)}(\eta(\delta') < \sigma') \sup_{\substack{x'\in U: H(x')=\delta' \\ y'\in\mathbb{T}^m}} |\mathbf{E}_{(x',y')}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x')) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]|.
\end{aligned} \tag{2.34}$$

Note that the first term converges to 0 as $\varepsilon \to 0$ by Lemma 2.4.5, the probability in the second term is less than κ, and the supremum is uniformly bounded for all ε by Lemma 2.4.10. Thus, the expression on the left-hand side of (2.34) converges to 0 uniformly. Combining this with (2.33), we obtain

$$\sup_{\substack{x\in U:H(x)\geq\varepsilon^\alpha\\ y\in\mathbb{T}^m}}\sup_{\sigma'\leq\tau(\varepsilon^\alpha)}|\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'}))-f(H(x))-\int_0^{\sigma'}\mathcal{L}_cf(H(\tilde{X}^\varepsilon_s))ds]|\to 0.$$

Finally, let us choose $0<\alpha'<\alpha$. Apply Corollary 2.4.7 and Lemma 2.4.8 to obtain that $\mathbf{E}_{(x,y)}\sigma\wedge\tau(\varepsilon^{\alpha'})<\varepsilon^{\alpha'}$ and $\mathbf{P}_{(x,y)}(\sigma<\tau(\varepsilon^{\alpha'}))>1/2$ for all $x\in\gamma(\varepsilon^\alpha)$, $y\in\mathbb{T}^m$, and ε sufficiently small. As in (2.34), by stopping the process at $\tau(\varepsilon^\alpha)\wedge\sigma'$ and $\tau(\varepsilon^{\alpha'})\wedge\sigma'$ and using the strong Markov property, we can conclude that

$$\sup_{\substack{x\in U:H(x)\geq\varepsilon^\alpha\\ y\in\mathbb{T}^m}}\sup_{\sigma'\leq\sigma}|\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'}))-f(H(x))-\int_0^{\sigma'}\mathcal{L}_cf(H(\tilde{X}^\varepsilon_s))ds]|\to 0.$$

Now (2.32) follows from this by applying Corollary 2.4.7 again. □

2.4.3 Averaging principle starting from $\gamma(\varepsilon^\alpha)$

Fix $0<\alpha<\alpha_1<\alpha_2<1/2$, $r>0$ small enough. Recall $Q(h)$ is the rotation time of $\boldsymbol{x}_t$ on $\gamma(h)$. Our first lemma in this section concerns the typical deviation during one rotation.

Lemma 2.4.11. *For each $\delta>0$ there is $\kappa>0$ such that for all $x\in U(\varepsilon^{\alpha_1},r)$, $y\in\mathbb{T}^m$, and ε sufficiently small,*

$$\mathbf{P}_{(x,y)}\left(\sup_{t\in[0,Q(H(x))]}|H(\tilde{\boldsymbol{X}}^\varepsilon_t)-H(\boldsymbol{x}_t)|>\varepsilon^{1/2-\delta}\right)<\varepsilon^\kappa.$$

There exists $\delta' > 0$ and $\kappa > 0$ such that for all $x \in U(\varepsilon^{\alpha_1}, r)$, $y \in \mathbb{T}^m$, and ε sufficiently small,

$$\mathbf{P}_{(x,y)}\left(\sup_{t\in[0,Q(H(x))]} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| > \varepsilon^{\delta'}\right) < \varepsilon^\kappa.$$

Proof. It suffices to prove the result for $\delta < 1/2-\alpha_1$. Fix $0 < \delta' < \delta'' < 1/2-\alpha_1-\delta$. Recall the definition of q in Section 2.4.1 and consider the coordinates H and q in $U(\varepsilon^{\alpha_2}, 2r)$. As in (2.21) and (2.22), let $q_0 = q(x)$, $u_h = u \cdot \nabla H$, $u_q = u \cdot \nabla q$, and $q_t = q_0 + t$, and write the equations with $\tau^0 = \inf\{t : |\tilde{\boldsymbol{H}}_t^\varepsilon - h_0| > \varepsilon^{1/2-\delta} \text{ or } |\tilde{\boldsymbol{Q}}_t^\varepsilon - q_t| > \varepsilon^{\delta''}\} \wedge Q(h_0)$:

$$\begin{aligned} H(\tilde{\boldsymbol{X}}_{\tau^0}^\varepsilon) &= H(x) + \sqrt{\varepsilon}\int_0^{\tau^0} \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^{\mathsf{T}}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s + \varepsilon(u_h(x,y) - u_h(\tilde{\boldsymbol{X}}_{\tau^0}^\varepsilon, \tilde{\boldsymbol{\xi}}_{\tau^0}^\varepsilon)) \\ &\quad + \varepsilon\int_0^{\tau^0} [\nabla_x u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot b(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot c(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds, \qquad (2.35) \\ q(\tilde{\boldsymbol{X}}_{\tau^0}^\varepsilon) &= q_{\tau^0} + \sqrt{\varepsilon}\int_0^{\tau^0} \nabla_y u_q(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^{\mathsf{T}}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s + \varepsilon(u_q(x,y) - u_q(\tilde{\boldsymbol{X}}_{\tau^0}^\varepsilon, \tilde{\boldsymbol{\xi}}_{\tau^0}^\varepsilon)) \\ &\quad + \varepsilon\int_0^{\tau^0} [\nabla_x u_q(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot b(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \nabla_y u_q(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot c(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds. \qquad (2.36) \end{aligned}$$

In Section 2.7.1, we prove that $|\nabla q| = O(|\nabla H|/H)$. Thus, it is not hard to see, by looking at the inverse of the Jacobian of (H, q) w.r.t x, that $|H(\tilde{\boldsymbol{X}}_t^\varepsilon) - H(\boldsymbol{x}_t)| \le \varepsilon^{1/2-\delta}$ and $|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\delta'}$ for all $t \le \tau^0$. Let S_H and S_Q denote the stochastic integrals in (2.35) and (2.36). Since $\tau^0 \lesssim |\log\varepsilon|$,

$$\mathbf{P}_{(x,y)}(\tau^0 < Q(h(x))) < \mathbf{P}_{(x,y)}(|S_H| > \varepsilon^{1/2-\delta}/2) + \mathbf{P}_{(x,y)}(|S_Q| > \varepsilon^{\delta''}/2).$$

The variance of S_H and S_Q is small:

$$\mathbf{Var}(S_H) = \varepsilon\mathbf{E}(\int_0^{\tau^0} |\nabla_y u_h(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^{\mathsf{T}}\sigma(\tilde{\xi}_s^\varepsilon)|^2 ds) \lesssim \varepsilon\mathbf{E}(\int_0^{\tau^0} |\nabla H(\tilde{X}_s^\varepsilon)|^2 ds) \lesssim \varepsilon|\log\varepsilon|,$$

$$\mathbf{Var}(S_Q) = \varepsilon \mathbf{E}(\int_0^{\tau^0} |\nabla_y u_q(\tilde{X}_s^\varepsilon, \tilde{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\tilde{\xi}_s^\varepsilon)|^2 ds) \lesssim \varepsilon \mathbf{E}(\int_0^{\tau^0} |\nabla q(\tilde{X}_s^\varepsilon)|^2 ds) \lesssim \varepsilon^{1-2\alpha_1} |\log \varepsilon|.$$

Hence both results follow from Chebyshev's inequality with $\kappa < \delta$. □

Let $F(h)$ be the solution to

$$\begin{cases} \mathcal{L}_c F = -1 \\ F(0) = F(2r) = 0 \end{cases} \tag{2.37}$$

Let τ^1 and τ^2 be the first times for $\tilde{X}_t^\varepsilon$ to exit $U(\varepsilon^{\alpha_1}, r)$ and $U(\varepsilon^{\alpha_2}, 2r)$, respectively. Let $x_t^\varepsilon = \boldsymbol{x}_{t/\varepsilon}$.

Lemma 2.4.12. *There exists a function $g(r)$ with $\lim_{r\to 0} g(r) = 0$ such that $|F'(h)| < g(r)$ for all $0 < h < 2r$. There exists $C > 0$ such that $|F''(h)| < C|\log h|$ and $|F'''(h)| < C/h$.*

Proof. The bounds can be verified with the help of estimates for $Q(h)$ in Section 2.7.1. □

Lemma 2.4.13. *There exists a function $g(r)$ with $\lim_{r\to 0} g(r) = 0$ such that for all $x \in \gamma(\varepsilon^\alpha)$, $y \in \mathbb{T}^m$, and ε sufficiently small,*

$$\mathbf{E}_{(x,y)} \tau^1 \leq \varepsilon^\alpha g(r). \tag{2.38}$$

Proof. For $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ starting from (x, y), we define $\bar{\tau}^2 = \varepsilon Q(H(x)) \wedge \tau^2$. As in (2.28):

$$\mathbf{E}_{(x,y)}[F(H(\tilde{X}_{\bar{\tau}^2}^\varepsilon)) - F(H(x)) - \int_0^{\bar{\tau}^2} (\frac{1}{2} A(\tilde{X}_s^\varepsilon) F''(H(\tilde{X}_s^\varepsilon)) + B(\tilde{X}_s^\varepsilon) F'(H(\tilde{X}_s^\varepsilon))) ds] = O(\varepsilon), \tag{2.39}$$

uniformly in $x \in U(\varepsilon^{\alpha_2}, 2r)$ and $y \in \mathbb{T}^m$. By the definition of $\bar{A}(h)$ and $\bar{B}(h)$, one can see that $\varepsilon Q(H(x)) \bar{A}(H(x)) = \int_0^{\varepsilon Q(H(x))} A(x_s^\varepsilon) ds$ and $\varepsilon Q(H(x)) \bar{B}(H(x)) = \int_0^{\varepsilon Q(H(x))} B(x_s^\varepsilon) ds$. Since

F solves (2.37), it follows that

$$\varepsilon Q(H(x)) = -\varepsilon Q(H(x))\mathcal{L}_c F(H(x)) = -\int_0^{\varepsilon Q(H(x))} \frac{1}{2}A(x_s^\varepsilon)F''(H(x_s^\varepsilon)) + B(x_s^\varepsilon)F'(H(x_s^\varepsilon))ds. \tag{2.40}$$

We prove that there exists $K > 0$ such that

$$K\mathbf{E}_{(x,y)}(F(H(\tilde{X}^\varepsilon_{\bar{\tau}^2})) - F(H(x))) \leq -\varepsilon Q(H(x)) \tag{2.41}$$

uniformly in $x \in U(\varepsilon^{\alpha_1}, r)$, $y \in \mathbb{T}^m$, and all ε sufficiently small. Then it follows that

$$\mathbf{E}_{(x,y)}\tau^1 \leq KF(H(x)), \tag{2.42}$$

for $x \in U(\varepsilon^{\alpha_1}, r)$, $y \in \mathbb{T}^m$, and all ε sufficiently small. Indeed, we can define $\bar{\tau}_k^2$, $k \geq 0$ recursively: $\bar{\tau}_0^2 = 0$, $\bar{\tau}_{k+1}^2 = \inf\{t \geq \tau_k^2 : \tilde{X}_t^\varepsilon \notin U(\varepsilon^{\alpha_2}, 2r)\} \wedge (\varepsilon Q(H(\tilde{X}^\varepsilon_{\bar{\tau}_k^2})) + \bar{\tau}_k^2)$, and denote the first k such that $\bar{\tau}_k^2$ exceeds τ^1 as $\boldsymbol{n}$. Then we have

$$\begin{aligned}
&\mathbf{E}_{(x,y)}\left[F(H(\tilde{X}_{\bar{\tau}^2_{\boldsymbol{n}}})) - F(H(x))\right] \\
&= \mathbf{E}_{(x,y)}\sum_{k=0}^\infty \chi_{\bar{\tau}_k^2 < \tau^1}\left[F(H(\tilde{X}_{\bar{\tau}^2_{k+1}})) - F(H(\tilde{X}_{\bar{\tau}^2_k}))\right] \\
&\leq \mathbf{E}_{(x,y)}\sum_{k=0}^\infty \chi_{\bar{\tau}_k^2 < \tau^1} \sup_{(x',y')\in U(\varepsilon^{\alpha_1}, r)\times\mathbb{T}^m} \mathbf{E}_{(x',y')}\left[F(H(\tilde{X}^\varepsilon_{\bar{\tau}^2})) - F(H(x'))\right] \\
&\leq \frac{1}{K}\mathbf{E}_{(x,y)}\sum_{k=0}^\infty \chi_{\bar{\tau}_k^2 < \tau^1}(-\varepsilon Q(\tilde{X}_{\bar{\tau}_k^2})).
\end{aligned}$$

Hence

$$\mathbf{E}_{(x,y)}\tau^1 \leq \mathbf{E}_{(x,y)}\bar{\tau}^2_{\boldsymbol{n}} = \mathbf{E}_{(x,y)}\sum_{k=0}^{\infty}\chi_{\bar{\tau}^2_k<\tau^1}(\bar{\tau}^2_{k+1}-\bar{\tau}^2_k) \leq \varepsilon\mathbf{E}_{(x,y)}\sum_{k=0}^{\infty}\chi_{\bar{\tau}^2_k<\tau^1}Q(\tilde{X}_{\bar{\tau}^2_k}) \leq KF(H(x)).$$

Then (2.38) follows from (2.42) and Lemma 2.4.12 by taking $x \in \gamma(\varepsilon^\alpha)$.

To prove (2.41), it is enough to see that, for $x \in U(\varepsilon^{\alpha_1}, r)$, $y \in \mathbb{T}^m$, and ε sufficiently small,

$$\begin{aligned}
&\varepsilon Q(H(x)) + \mathbf{E}_{(x,y)}F(H(\tilde{X}^\varepsilon_{\bar{\tau}^2})) - F(H(x)) \\
&= -\mathbf{E}_{(x,y)}\int_0^{\varepsilon Q(H(x))}\left(\frac{1}{2}A(x^\varepsilon_s)F''(H(x^\varepsilon_s)) + B(x^\varepsilon_s)F'(H(x^\varepsilon_s))\right)ds \\
&\quad + \mathbf{E}_{(x,y)}\int_0^{\bar{\tau}^2}\left(\frac{1}{2}A(\tilde{X}^\varepsilon_s)F''(H(\tilde{X}^\varepsilon_s)) + B(\tilde{X}^\varepsilon_s)F'(H(\tilde{X}^\varepsilon_s))\right)ds + O(\varepsilon) \\
&= \mathbf{E}_{(x,y)}\int_0^{\bar{\tau}^2}\left(\frac{1}{2}A(\tilde{X}^\varepsilon_s)F''(H(\tilde{X}^\varepsilon_s)) - \frac{1}{2}A(x^\varepsilon_s)F''(H(x^\varepsilon_s))\right)ds \\
&\quad + \mathbf{E}_{(x,y)}\int_0^{\bar{\tau}^2}\left(B(\tilde{X}^\varepsilon_s)F'(H(\tilde{X}^\varepsilon_s)) - B(x^\varepsilon_s)F'(H(x^\varepsilon_s))\right)ds \\
&\quad - \mathbf{E}_{(x,y)}\int_{\bar{\tau}^2}^{\varepsilon Q(H(x))}\left(\frac{1}{2}A(x^\varepsilon_s)F''(H(x^\varepsilon_s)) + B(x^\varepsilon_s)F'(H(x^\varepsilon_s))\right)ds + O(\varepsilon) \\
&= o(\varepsilon Q(H(x))),
\end{aligned}$$

where the first equality is due to (2.39) and (2.40) and the last equality is due to Lemma 2.4.11 and Lemma 2.4.12. □

Similarly to Lemma 2.4.10, we can look at the transitions between $\gamma(\varepsilon^\alpha)$ and $\gamma(\varepsilon^{\alpha_1})$. By the transition probabilities given in Lemma 2.4.8 and transition time given in Corollary 2.4.7 and Lemma 2.4.13, one can obtain the following result using the Strong Markov property.

Corollary 2.4.14. *There exists a function $g(r)$ with $\lim_{r\to 0} g(r) = 0$ such that for all $x \in \gamma(\varepsilon^\alpha)$,*

$y \in \mathbb{T}^m$, and ε sufficiently small,

$$\mathbf{E}_{(x,y)}\tau(r) \wedge \sigma \leq \varepsilon^\alpha g(r).$$

Lemma 2.4.15. *For each $f \in \mathcal{D}$, as $\varepsilon \downarrow 0$,*

$$\sup_{(x,y)\in\gamma(\varepsilon^\alpha)\times\mathbb{T}^m} \sup_{\sigma'\leq\sigma} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]| = o(\varepsilon^\alpha),$$

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

Proof. Fix $\kappa > 0$. By Corollary 2.4.14, we can choose r small enough so that for stopping time $\sigma' \leq \sigma$: $|\mathbf{E}_{(x,y)}[H(\tilde{X}^\varepsilon_{\tau(r)\wedge\sigma'}) - H(x)]| < \kappa\varepsilon^\alpha$ and $|\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\tau(r)\wedge\sigma'})) - f(H(x))]| < \kappa\varepsilon^\alpha$, and

$$\sup_{(x,y)\in\gamma(\varepsilon^\alpha)\times\mathbb{T}^m} \sup_{\sigma'\leq\sigma} |\mathbf{E}_{(x,y)} \int_0^{\tau(r)\wedge\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds| < \kappa\varepsilon^\alpha,$$

for all ε sufficiently small, using similar arguments leading to (2.21) and (2.27). It follows that, $\mathbf{P}_{(x,y)}(H(\tilde{X}^\varepsilon_{\tau(r)\wedge\sigma'}) = r) \leq H(x)/r + \kappa\varepsilon^\alpha/r \leq 2\varepsilon^\alpha/r$. Therefore, uniformly in all $x \in \gamma(\varepsilon^\alpha)$, $y \in \mathbb{T}^m$, and $\sigma' \leq \sigma$,

$$\begin{aligned}
&|\mathbf{E}_{(x,y)}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(\tilde{X}^\varepsilon_s)ds]| \\
&\leq |\mathbf{E}_{(x,y)} f[H(\tilde{X}^\varepsilon_{\tau(r)\wedge\sigma'}) - f(H(x)) - \int_0^{\tau(r)\wedge\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]| \\
&\quad + \mathbf{P}_{(x,y)}(H(X^\varepsilon_{\tau(r)\wedge\sigma'}) = r) \sup_{\substack{x'\in\gamma(r)\\ y'\in\mathbb{T}^m}} |\mathbf{E}_{(x',y')}[f(H(\tilde{X}^\varepsilon_{\sigma'})) - f(H(x)) - \int_0^{\sigma'} \mathcal{L}_c f(H(\tilde{X}^\varepsilon_s))ds]|
\end{aligned}$$

$$\leq 3\kappa\varepsilon^{\alpha},$$

for ε sufficiently small, due to Proposition 2.4.6 and our choice of r. The result follows because κ can be chosen arbitrarily small. $\square$

Lemma 2.4.16. *There is a constant $\kappa > 0$ such that, for all ε sufficiently small,*

$$\sup_{(x,y)\in\gamma(\varepsilon^{\alpha})\times\mathbb{T}^m} \mathbf{E}_{(x,y)} e^{-\sigma} \leq 1 - \kappa\varepsilon^{\alpha}.$$

Proof. By Corollary 2.4.14, as in the proof of Lemma 2.4.8, we can fix $0 < r < 1/3$ such that for all $x \in \gamma(\varepsilon^{\alpha})$, $y \in \mathbb{T}^m$, and ε sufficiently small, $\mathbf{P}_{(x,y)}(\tau(r) < \sigma) \geq \varepsilon^{\alpha}/2r$. Let F be defined as in (2.37) and $t = F(r)/3$, then it follows from Proposition 2.4.6, as $\varepsilon \downarrow 0$,

$$\sup_{(x,y)\in\gamma(r)\times\mathbb{T}^m} \mathbf{E}_{(x,y)}[F(H(\tilde{X}^{\varepsilon}_{\sigma\wedge\tau(2r)\wedge t})) - F(H(x)) - \int_0^{\sigma\wedge\tau(2r)\wedge t} \mathcal{L}_c F(H(\tilde{X}^{\varepsilon}_s))ds] \to 0.$$

Thus, we have that for all $x \in \gamma(r)$, $y \in \mathbb{T}^m$, and ε sufficiently small,

$$\mathbf{E}_{(x,y)} F(H(\tilde{X}^{\varepsilon}_{\sigma\wedge\tau(2r)\wedge t}) > F(r)/2,$$

and it follows that,

$$\mathbf{P}_{(x,y)}(\sigma > t) \geq \mathbf{P}_{(x,y)}(\sigma\wedge\tau(2r) > t) > \frac{\mathbf{E}_{(x,y)} F(H(\tilde{X}^{\varepsilon}_{\sigma\wedge\tau(2r)\wedge t})}{\sup_{[0,2r]} F(h)} > \frac{F(r)}{2\sup_{[0,2r]} F(h)} =: c_1(r).$$

Then, for all $x \in \gamma(r)$, $y \in \mathbb{T}^m$, and ε sufficiently small,

$$\mathbf{E}_{(x,y)} e^{-\sigma} \le \mathbf{P}_{(x,y)}(\sigma \le t) + \mathbf{P}_{(x,y)}(\sigma > t) e^{-t} \le 1 - \mathbf{P}_{(x,y)}(\sigma > t)(1 - e^{-t}) \le 1 - c(r),$$

with $c(r) = (1 - \exp(-F(r)/3)) c_1(r) > 0$, and therefore,

$$\begin{aligned}
\mathbf{E}_{(x,y)} e^{-\sigma} &\le \mathbf{P}_{(x,y)}(\sigma < \tau(r)) + \mathbf{P}_{(x,y)}(\sigma > \tau(r)) \sup_{x' \in \gamma(r), y' \in \mathbb{T}^m} \mathbf{E}_{(x',y')} e^{-\sigma} \\
&\le 1 - \mathbf{P}_{(x,y)}(\sigma > \tau(r))(1 - \sup_{x' \in \gamma(r), y' \in \mathbb{T}^m} \mathbf{E}_{(x',y')} e^{-\sigma}) \\
&\le 1 - \frac{1}{2} c(r) \frac{\varepsilon^\alpha}{r}.
\end{aligned}$$

The result holds with $\kappa = c(r)/2r$. □

Corollary 2.4.17. *For a given $t > 0$, the expected number of excursions before t is $O(\varepsilon^{-\alpha})$:*

$$\sum_{n=0}^{\infty} \mathbf{P}_{(x,y)}(\tau_{n+1} < t) \le \sum_{n=0}^{\infty} \mathbf{P}_{(x,y)}(\sigma_n < t) \le \frac{e^t}{\kappa} \varepsilon^{-\alpha}, \tag{2.43}$$

where κ is the constant chosen in Lemma 2.4.16.

Proof. By Lemma 2.4.16 and the strong Markov property,

$$\sup_{(x,y) \in M \times \mathbb{T}^m} \mathbf{E}_{(x,y)} e^{-\sigma_n} \le (\sup_{(x,y) \in \gamma' \times \mathbb{T}^m} \mathbf{E}_{(x,y)} e^{-\sigma})^n \le (1 - \kappa \varepsilon^\alpha)^n.$$

Thus, by Markov's inequality, for all $n > 0$,

$$\mathbf{P}_{(x,y)}(\tau_{n+1} < t) \le \mathbf{P}_{(x,y)}(\sigma_n < t) \le e^t \mathbf{E}_{(x,y)} e^{-\sigma_n} \le e^t (1 - \kappa \varepsilon^\alpha)^n,$$

and (2.43) follows by taking the sum. $\square$

2.5 Exponential convergence on the separatrix

We fix $0 < \alpha < 1/2$. As in (2.6), we define inductively two sequences of stopping times σ_n, $n \geq 0$, and τ_n, $n \geq 1$, but now for the general process $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ on $M \times \mathbb{T}^m$ with additional drift $c(x, y)$. Without loss of generality, we assume that the saddle point O satisfies $H(O) = 0$. Let $V^\varepsilon = \{x : |H(x)| < \varepsilon^\alpha\}$ and U_1, U_2, U_3 be the three domains separated by γ. We aim to prove that the distribution of Markov chain $(\tilde{X}_{\sigma_n}^\varepsilon, \tilde{\xi}_{\sigma_n}^\varepsilon)$ converges in total variation exponentially fast, uniformly in ε and in the initial distribution. Namely, we have the following lemma.

Lemma 2.5.1. *Let $\nu_{x,y}^{n,\varepsilon}$ denote the measure on $\gamma \times \mathbb{T}^m$ induced by $(\tilde{X}_{\sigma_n}^\varepsilon, \tilde{\xi}_{\sigma_n}^\varepsilon)$ with starting point $(x, y) \in \gamma \times \mathbb{T}^m$. Then there exist a probability measure ν^ε on $\gamma \times \mathbb{T}^m$ and constants $\Xi > 0$ and $0 < c < 1$ such that, for all ε sufficiently small,*

$$\sup_{(x,y)\in\gamma\times\mathbb{T}^m} \mathrm{TV}(\nu_{x,y}^{n,\varepsilon}, \nu^\varepsilon) < \Xi \cdot (1 - c)^n,$$

where TV is the total variation distance of probability measures.

The rest of this section is devoted to the proof of Lemma 2.5.1. Let $\boldsymbol{\sigma}_n$, $n \geq 0$, and $\boldsymbol{\tau}_n$, $n \geq 1$, be the stopping times w.r.t. $(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon)$ that are analogous to σ_n, τ_n w.r.t. $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$. The lemma is equivalent to the exponential convergence in total variation of $(\tilde{\boldsymbol{X}}_{\boldsymbol{\tau}_n}^\varepsilon, \tilde{\boldsymbol{\xi}}_{\boldsymbol{\tau}_n}^\varepsilon)$ on $\gamma' \times \mathbb{T}^m$, uniformly in ε and in the initial distribution. The proof consists of three steps:

1. The process starting on $\gamma' \times \mathbb{T}^m$ hits $I \times \mathbb{T}^m$ before $\boldsymbol{\tau}_1$ with uniformly positive probability, where I is a fixed interval on the separatrix.

2. Let the process starting on $I \times \mathbb{T}^m$ evolve for a certain period of time. Then, by a local limit theorem, we can estimate from below the probabilities of hitting $O(\varepsilon)$-sized boxes in a certain $O(\sqrt{\varepsilon})$-sized region, uniformly w.r.t. the starting point on $I \times \mathbb{T}^m$.

3. By the Hörmander condition *(H5)*, we prove a common lower bound for the density of the distribution of the process starting from each of the $O(\varepsilon)$-sized boxes after a short time.

Let us take care of these steps in order.

Step 1. Let $0 < \beta < 1$, which will be specified later. We prove that the process has a uniformly positive probability of following along the averaged motion and going through a neighborhood of the saddle point without making a deviation more than $\beta\sqrt{\varepsilon}$ in terms of H.

Lemma 2.5.2. *For each fixed $\hat{t} > 0$, $\beta' > 0$,*

$$\mathbf{P}_{(x,y)}\left(\sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \beta'\sqrt{\varepsilon}\right),$$

is uniformly positive for all $(x, y) \in M \times \mathbb{T}^m$ and ε sufficiently small.

Proof. Let the eigenvalues of $\nabla^2 H$ be bounded by K. Recall formula (2.19). By the boundedness of the coefficients, the event

$$E := \left\{\sup_{0\le t\le \tilde{t}} |\int_0^t \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s| \le \frac{1}{2}\beta' e^{-K\tilde{t}}\right\}$$

has positive probability, uniformly in the starting points. By (2.19), we have that on the event E, for $t \le \tilde{t}$ and ε sufficiently small,

$$|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le |\int_0^t (\nabla^\perp H(\tilde{\boldsymbol{X}}_s^\varepsilon) - \nabla^\perp H(\boldsymbol{x}_s)ds| + \sqrt{\varepsilon}|\int_0^t \nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s|$$

$$+\varepsilon\left|\int_0^t[\nabla_x u(\tilde{\boldsymbol{X}}_s^\varepsilon,\tilde{\boldsymbol{\xi}}_s^\varepsilon)b(\tilde{\boldsymbol{X}}_s^\varepsilon,\tilde{\boldsymbol{\xi}}_s^\varepsilon)+\nabla_y u(\tilde{\boldsymbol{X}}_s^\varepsilon,\tilde{\boldsymbol{\xi}}_s^\varepsilon)c(\tilde{\boldsymbol{X}}_s^\varepsilon,\tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds\right|+\varepsilon|u(x,y)-u(\tilde{\boldsymbol{X}}_t^\varepsilon,\tilde{\boldsymbol{\xi}}_t^\varepsilon)|$$
$$\leq K\int_0^t|\tilde{\boldsymbol{X}}_s^\varepsilon-\boldsymbol{x}_s|ds+\beta' e^{-K\tilde{t}}\sqrt{\varepsilon}.$$

Then Grönwall's inequality implies that $|\tilde{\boldsymbol{X}}_t^\varepsilon-\boldsymbol{x}_t|\leq\beta'\sqrt{\varepsilon}$ for all $t\leq\tilde{t}$. Therefore, E implies $\{\sup_{0\leq t\leq\hat{t}}|\tilde{\boldsymbol{X}}_t^\varepsilon-\boldsymbol{x}_t|\leq\beta'\sqrt{\varepsilon}\}$, and the uniform positivity follows. $\square$

Lemma 2.5.3. *For any given $0<c<1$, there exist curves Γ_1 and Γ_2 in U_1 such that*

(i) Γ_1 *and* Γ_2 *have their tangent vectors as ∇H. They intersect with the separatrix on different sides of the saddle point and the averaged motion on the separatrix spends finite time from* Γ_2 *to* Γ_1.

(ii) Let $x\in\Gamma_1$ satisfy $2\beta\sqrt{\varepsilon}\leq|H(x)|\leq 2\sqrt{\varepsilon}$ and $\tau_x=\inf\{t:\tilde{\boldsymbol{X}}_t^\varepsilon\in\Gamma_2\}$. Then for all $y\in\mathbb{T}^m$, $\mathbf{P}_{(x,y)}(\sup_{0\leq t\leq\tau_x}|H(\tilde{\boldsymbol{X}}_t^\varepsilon)-H(x)|\leq\beta\sqrt{\varepsilon})>c$ for all ε sufficiently small.

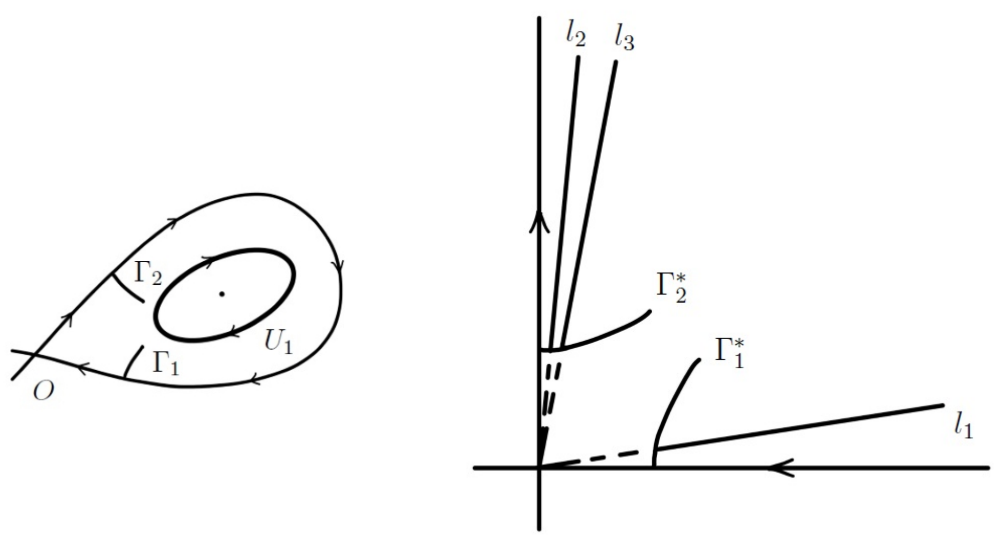


Figure 2.5: Curves in different coordinates.

Proof. Suppose $H(x)>0$ for all $x\in U_1$. By the Morse lemma, there exist neighborhoods U and V of the saddle point O and the origin, respectively, and a diffeomorphism ψ from U to V such that $H(x) = G(\psi(x))$, where $G(z) = z_1 z_2$. Then consider a random change of time by dividing the generator by $D(x) := \det(\nabla_x \psi(x))$:

$$
\begin{aligned}
d\tilde{\boldsymbol{X}}_t^{\varepsilon*} &= \frac{b(\tilde{\boldsymbol{X}}_t^{\varepsilon*}, \tilde{\boldsymbol{\xi}}_t^{\varepsilon*})}{D(\tilde{\boldsymbol{X}}_t^{\varepsilon*})} dt, \\
d\tilde{\boldsymbol{\xi}}_t^{\varepsilon*} &= \frac{1}{\varepsilon} \frac{v(\tilde{\boldsymbol{\xi}}_t^{\varepsilon*})}{D(\tilde{\boldsymbol{X}}_t^{\varepsilon*})} dt + \frac{1}{\sqrt{\varepsilon}} \frac{\sigma(\tilde{\boldsymbol{\xi}}_t^{\varepsilon*})}{\sqrt{D(\tilde{\boldsymbol{X}}_t^{\varepsilon*})}} dW_t + \frac{c(\tilde{\boldsymbol{X}}_t^{\varepsilon*}, \tilde{\boldsymbol{\xi}}_t^{\varepsilon*})}{D(\tilde{\boldsymbol{X}}_t^{\varepsilon*})} dt.
\end{aligned}
\tag{2.44}
$$

Write the equation for $\tilde{Z}_t^{\varepsilon*} := \psi(\tilde{\boldsymbol{X}}_t^{\varepsilon*})$:

$$
d\tilde{Z}_t^{\varepsilon*} = \frac{1}{D(\psi^{-1}(\tilde{Z}_t^{\varepsilon*}))} \cdot \nabla_x \psi(\psi^{-1}(\tilde{Z}_t^{\varepsilon*})) b(\psi^{-1}(\tilde{Z}_t^{\varepsilon*}), \tilde{\boldsymbol{\xi}}_t^{\varepsilon*}) dt =: b^*(\tilde{Z}_t^{\varepsilon*}, \tilde{\boldsymbol{\xi}}_t^{\varepsilon*}) dt.
$$

It is not hard to verify that $b^*(z, y)$ satisfies

$$
\int_{\mathbb{T}^m} b^*(z, y) d\mu(y) = \nabla^{\perp} G(z).
$$

Hence, by Lemma 2.3.2, there exists a bounded solution $u^*(z, y)$ to

$$
Lu^*(z, y) = -(b^*(z, y) - \nabla^{\perp} G(z)) \cdot D(\psi^{-1}(z)).
$$

Consider a local coordinate $G = z_1 z_2$ and $\phi^* = \frac{1}{2} \log(z_2/z_1)$ in V. The averaged motion has constant speed: 0 in G and 1 in ϕ^*. As in (2.21) and (2.22), we have the equations for $\tilde{G}_t^{\varepsilon*} = G(\tilde{Z}_t^{\varepsilon*})$, $\tilde{\Phi}_t^{\varepsilon*} = \phi^*(\tilde{Z}_t^{\varepsilon*})$, by applying Ito's formula to $u_g^* = u^* \cdot \nabla G$ and $u_\phi^* = u^* \cdot \nabla \phi^*$, with

$z = \psi(x)$, $g_0 = G(z)$, and $\phi_0^* = \phi^*(z)$:

$$\tilde{G}_t^{\varepsilon *} = g_0 + \sqrt{\varepsilon}\int_0^t \nabla_y u_g^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *})^\mathsf{T} \frac{\sigma(\tilde{\boldsymbol{\xi}}_s^{\varepsilon *})}{\sqrt{D(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}))}} dW_s - \varepsilon(u_g^*(\tilde{Z}_t^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_t^{\varepsilon *}) - u_g^*(z,y))$$
$$+ \varepsilon \int_0^t \left[\nabla_z u_g^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) \cdot b^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) + \nabla_y u_g^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) \cdot \frac{c(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}), \tilde{\boldsymbol{\xi}}_s^{\varepsilon *})}{D(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}))}\right] ds, \tag{2.45}$$

$$\tilde{\Phi}_t^{\varepsilon *} = \phi_0^* + t + \sqrt{\varepsilon}\int_0^t \nabla_y u_\phi^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *})^\mathsf{T} \frac{\sigma(\tilde{\boldsymbol{\xi}}_s^{\varepsilon *})}{\sqrt{D(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}))}} dW_s - \varepsilon(u_\phi^*(\tilde{Z}_t^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_t^{\varepsilon *}) - u_\phi^*(z,y))$$
$$+ \varepsilon \int_0^t \left[\nabla_z u_\phi^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) \cdot b^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) + \nabla_y u_\phi^*(\tilde{Z}_s^{\varepsilon *}, \tilde{\boldsymbol{\xi}}_s^{\varepsilon *}) \cdot \frac{c(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}), \tilde{\boldsymbol{\xi}}_s^{\varepsilon *})}{D(\psi^{-1}(\tilde{Z}_s^{\varepsilon *}))}\right] ds \tag{2.46}$$

To get the lower bound for the desired probability, we will choose the curves Γ_1 and Γ_2 that are close enough to the saddle point. The time it takes to get from Γ_1 to Γ_2 is still of order $|\log\varepsilon|$ since they are chosen independently of ε. In this way, the process starting on Γ_1 and stopped on Γ_2 will be shown to have small variance, hence it is unlikely for the process to have deviations larger than what we wish. With $C > 0$ to be specified later, let $l_1 = \{z : \phi^*(z) = \frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C\}$, $l_2 = \{z : \phi^*(z) = -(\frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C)\}$, and $l_3 = \{z : \phi^*(z) = -(\frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C) - 2\}$. The idea is to look at event that the process stays close to the averaged motion before the latter reaches l_2, which implies that the process does not make a large deviation in G, or equivalently, in H, before reaching l_3. Let Γ_1^*, Γ_2^* be the curves that have tangent vectors as $\nabla_x\psi\circ\psi^{-1}(\nabla_x\psi\circ\psi^{-1})^\mathsf{T}\nabla G$ and go through the points $(e^{-C}, e^{C}\beta\sqrt{\varepsilon})$, $(e^{C+1}\beta\sqrt{\varepsilon}, e^{-C-1})$, respectively. Since ψ is a diffeomorphism, it is easy to see that each z on Γ_1^* or Γ_2^* with $G(z) \geq \beta\sqrt{\varepsilon}$ satisfies that $\frac{1}{4}\log\varepsilon+\frac{1}{2}\log\beta+C \leq \phi^*(z) \leq -(\frac{1}{4}\log\varepsilon+\frac{1}{2}\log\beta+C)-2$. Let Γ_1 and Γ_2 be the pre-images of Γ_1^* and Γ_2^* in U_1. They have ∇H as tangent vectors due to the specific way we construct Γ_1^* and Γ_2^*. Consider the process in (2.44) starting at $x \in \Gamma_1$

satisfying that $2\beta\sqrt{\varepsilon} \le H(x) \le 2\sqrt{\varepsilon}$ with an arbitrary $y \in \mathbb{T}^m$. Let $\phi_t^* = \phi_0^* + t$. Define $t_x = \inf\{t : \phi_t^* = \phi^*(l_2)\}$ and $\tau_x^* = \inf\{t : |\tilde{G}_t^{\varepsilon *} - g_0| = \beta\sqrt{\varepsilon}\} \wedge \inf\{t : |\tilde{\Phi}_t^{\varepsilon *} - \phi_t^*| = 1\} \wedge t_x$. Then it is clear that $\mathbf{P}_{(x,y)}(\sup_{0\le t\le\tau_x^*} |H(\tilde{\boldsymbol{X}}_t^\varepsilon) - H(x)| \le \beta\sqrt{\varepsilon}) \ge \mathbf{P}_{(x,y)}(\tau_x^* = t_x)$. Let S_G and S_ϕ denote the stochastic integrals in (2.45) and (2.46), respectively, with t replaced by τ_x^*. Since $\tau_x^* \lesssim |\log\varepsilon|$, ∇G is bounded, and $\nabla\phi^* \lesssim \varepsilon^{-1/2}$ before τ_x^*, we see that the unwanted deviations happen only if S_G and S_ϕ are large. Namely,

$$\mathbf{P}_{(x,y)}(\tau_x^* < t_x) \le \mathbf{P}_{(x,y)}(|S_G| \ge \beta\sqrt{\varepsilon}/2) + \mathbf{P}_{(x,y)}(|S_\phi| \ge 1/2).$$

Both terms on the right-hand side can be controlled by Chebyshev's inequality. Note that there exists a constant $K > 0$ independent of ε such that

$$\begin{aligned}
\mathbf{Var}(S_G) &\le \varepsilon K\mathbf{E}\int_0^{\tau_x^*} |\nabla G(\tilde{Z}_s^{\varepsilon *})|^2 ds \\
&= \varepsilon K\mathbf{E}\int_0^{\tau_x^*} \tilde{G}_s^{\varepsilon *}(e^{2\tilde{\Phi}_s^{\varepsilon *}} + e^{-2\tilde{\Phi}_s^{\varepsilon *}})ds \\
&\le \varepsilon K\int_0^{\tau_x^*} (2+\beta)\sqrt{\varepsilon}e^2(e^{2\phi_s^*} + e^{-2\phi_s^*})ds \\
&\le 3K\sqrt{\varepsilon^3}e^2\int_0^{-2(\frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C)} (e^{2\phi_s^*} + e^{-2\phi_s^*})ds \\
&= \ 3K\sqrt{\varepsilon^3}e^2\int_{\frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C}^{-(\frac{1}{4}\log\varepsilon + \frac{1}{2}\log\beta + C)} (e^{2\varphi} + e^{-2\varphi})d\varphi \\
&\le \ \frac{3}{\beta}Ke^{2-2C}\varepsilon,
\end{aligned}$$

and, similarly,

$$\mathbf{Var}(S_\phi) \le \varepsilon K\mathbf{E}\int_0^{\tau_x^*} |\nabla\Phi(\tilde{Z}_s^{\varepsilon *})|^2 ds \le \varepsilon K\mathbf{E}\int_0^{\tau_x^*} \frac{1}{\tilde{G}_s^{\varepsilon *}}(e^{2\tilde{\Phi}_s^{\varepsilon *}} + e^{-2\tilde{\Phi}_s^{\varepsilon *}})ds \le \frac{1}{\beta^2}Ke^{2-2C}.$$

Then C can be chosen large enough such that both variances are small enough, and hence $\mathbf{P}_{(x,y)}(|H(\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau^*_x}) - H(x)| \leq \beta\sqrt{\varepsilon}) > c$. $\square$

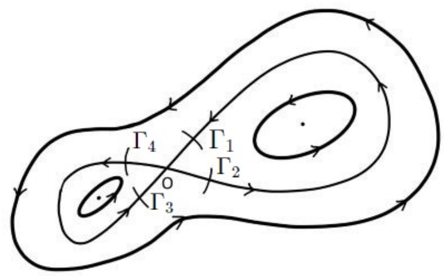


Figure 2.6: Four curves on four directions.

We can choose the corresponding curves in the other regions. As a result, we have four curves corresponding to four different directions, all with positive distance to the saddle point, as shown in Figure 2.6. Moreover, the corresponding transition probabilities near the saddle point have lower bounds analogous to that given in Lemma 2.5.3 (ii). For the rotations happening away from those curves, we will prove that, before the time when the process comes back to the curves, the deviation of H can be large enough to cross the separatrix with positive probability. Let $\Gamma_i(h_1, h_2)$ be the set $\{x \in \Gamma_i : h_1 \leq H(x) \leq h_2\}$.

Lemma 2.5.4. *For each fixed $\hat{t} > 0$,*

$$\mathbf{P}_{(x,y)}\left(\inf_{0\leq t\leq \hat{t}} H(\tilde{\boldsymbol{X}}^{\varepsilon}_t) \leq -\sqrt{\varepsilon}, \sup_{0\leq t\leq \hat{t}} |\tilde{\boldsymbol{X}}^{\varepsilon}_t - \boldsymbol{x}_t| \leq \varepsilon^{\frac{1+2\alpha}{4}}\right)$$

is positive uniformly in $x \in \Gamma_2(0, 2\sqrt{\varepsilon})$, $y \in \mathbb{T}^m$, and all ε sufficiently small.

Proof. By Lemma 2.5.2 and the Markov property, it is enough to consider small $\hat{t}$ such that $\boldsymbol{x}_t$ does not reach Γ_1 before $\hat{t}$. Using formula (2.19) again, we see that $\mathbf{P}_{(x,y)}(\sup_{0\leq t\leq \hat{t}} |\tilde{\boldsymbol{X}}^{\varepsilon}_t - \boldsymbol{x}_t| >$

$\varepsilon^{\frac{1+2\alpha}{4}}) \to 0$ as $\varepsilon \downarrow 0$ uniformly in (x, y). Use formula (2.21) on a shorter time scale:

$$H(\tilde{\boldsymbol{X}}_t^\varepsilon) = H(x) + \sqrt{\varepsilon}\int_0^t \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s + \varepsilon(u_h(x,y) - u_h(\tilde{\boldsymbol{X}}_t^\varepsilon, \tilde{\boldsymbol{\xi}}_t^\varepsilon))$$
$$+ \varepsilon\int_0^t [\nabla_x u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot b(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon) + \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)\cdot c(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)]ds.$$

So, it suffices to show the uniform positivity of

$$\mathbf{P}_{(x,y)}\left(\inf_{0\le t\le \hat{t}}\int_0^t \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s \le -4,\ \sup_{0\le t\le \hat{t}}|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right).$$

Note that there exists another Brownian motion $\tilde{W}$ such that

$$\int_0^t \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)dW_s = \tilde{W}\left(\int_0^t |\nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)|^2 ds\right).$$

Recall that in Section 2.4.1 we defined $A(x) = \int_{\mathbb{T}^m}|\nabla_y u_h(x,y)\sigma(y)|^2 d\mu(y)$. By Corollary 2.3.4,

$$\mathbf{E}_{(x,y)}\left|\int_0^{\hat{t}} |\nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon))|^2 ds - \int_0^{\hat{t}} A(\tilde{\boldsymbol{X}}_s^\varepsilon)ds\right| = O(\sqrt{\varepsilon}). \tag{2.47}$$

Note that on the event $\{\sup_{0\le t\le\hat{t}}|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\}$, $A(\tilde{\boldsymbol{X}}_t^\varepsilon)$ is uniformly positive for $0 \le t \le \hat{t}$.

Let us denote this lower bound as m, which is independent of x, y, and ε. Then

$$\mathbf{P}_{(x,y)}(\int_0^{\hat{t}} A(\tilde{\boldsymbol{X}}_s^\varepsilon)ds > m\hat{t},\ \sup_{0\le t\le\hat{t}}|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}) \to 1.$$

By the L^1 convergence in (2.47), we obtain

$$\mathbf{P}_{(x,y)}\left(\int_0^{\hat{t}} |\nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)|^2 ds > m\hat{t}/2, \sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right) \to 1. \tag{2.48}$$

Suppose $0 < c < \mathbf{P}(\inf_{0\le t\le m\hat{t}/2} \tilde{W}_t < -4)$. Then, for all ε sufficiently small,

$$\begin{aligned}
&\mathbf{P}_{(x,y)}\left(\inf_{0\le t\le \hat{t}} \int_0^t \nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon) dW_s \le -4, \sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right) \\
&= \mathbf{P}_{(x,y)}\left(\inf_{0\le t\le \hat{t}} \tilde{W}\left(\int_0^t |\nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)|^2 ds\right) \le -4, \sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right) \\
&\ge \mathbf{P}_{(x,y)}\left(\inf_{0\le t\le m\hat{t}/2} \tilde{W}_t \le -4, \int_0^{\hat{t}} |\nabla_y u_h(\tilde{\boldsymbol{X}}_s^\varepsilon, \tilde{\boldsymbol{\xi}}_s^\varepsilon)^\mathsf{T}\sigma(\tilde{\boldsymbol{\xi}}_s^\varepsilon)|^2 ds > m\hat{t}/2, \sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right) \\
&\ge c/2.
\end{aligned}$$

$\square$

Remark 2.5.5. *The result in Lemma 2.5.4 also holds for $x \in \Gamma_4(0, 2\sqrt{\varepsilon})$. Similarly, for each fixed $\hat{t} > 0$,*

$$\mathbf{P}_{(x,y)}\left(\sup_{0\le t\le \hat{t}} H(\tilde{\boldsymbol{X}}_t^\varepsilon) \ge \sqrt{\varepsilon}, \sup_{0\le t\le \hat{t}} |\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right)$$

is positive uniformly in $x \in \Gamma_2(-2\sqrt{\varepsilon}, 0) \cup \Gamma_4(-2\sqrt{\varepsilon}, 0)$, $y \in \mathbb{T}^m$, and ε sufficiently small.

Now we can choose $\beta = 1/10$. By the results in Lemma 2.5.2, Lemma 2.5.3, Lemma 2.5.4, and Remark 2.5.5, using the strong Markov property, we obtain the following lemma:

Lemma 2.5.6. *There exists a closed interval I on γ that does not contain the saddle point and a constant $0 < c < 1$ satisfying the following property: if the system (2.13) starts at $(x, y) \in \gamma' \times \mathbb{T}^m$, then for all ε sufficiently small*

$$\mathbf{P}_{(x,y)}(\tilde{\eta}_1 < \boldsymbol{\tau}_1) \ge c$$

where $\tilde{\eta}_1 = \inf\{t : \tilde{\boldsymbol{X}}^\varepsilon_t \in I\}$.

Remark 2.5.7. *In order for us to apply Lemma 2.5.4, we need to choose* I *that contains the intersection of* Γ_2 *and* γ *in its interior. In fact, it is not difficult to show that Lemma 2.5.6 holds for any subset of* γ *with non-empty interior (see Figure 2.7).*

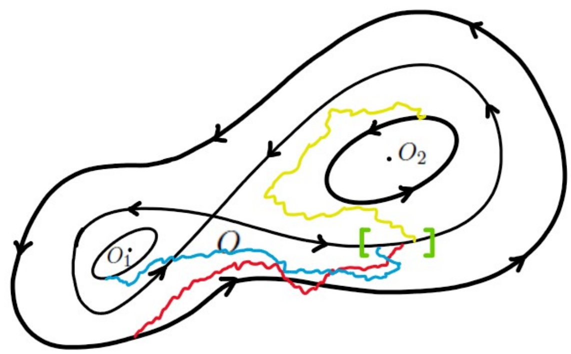


Figure 2.7: The interval on γ can be reached from any points on γ'.

Step 2. Without loss of generality, we assume that if $\boldsymbol{x}_t$ starts at one endpoint of I, then the other endpoint is $\boldsymbol{x}_{1/2}$. In the remainder of this section, $\boldsymbol{x}_t$ always denotes this deterministic motion, irrespective of where $\tilde{\boldsymbol{X}}^\varepsilon_t$ starts. We aim to study the distribution of the process $(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t)$ starting on $I \times \mathbb{T}^m$ with certain $t > 0$. The choice of t will depend on the initial point x being considered (see Figure 2.8), and this will be convenient as we use the strong Markov property later when combining all three steps. To be more precise, for $x \in I$, let $s(x)$ be such that $\boldsymbol{x}_{s(x)} = x$ (so $0 \le s(x) \le 1/2$). We introduce a process $\tilde{\zeta}^\varepsilon_t$, $0 \le t$, as the second term in the expansion of $\tilde{\boldsymbol{X}}^\varepsilon_t$ around the deterministic motion $\boldsymbol{x}_{(s(x)+t)}$:

$$d\tilde{\zeta}^\varepsilon_t = \nabla(\nabla^\perp H)(\boldsymbol{x}_{s(x)+t})\tilde{\zeta}^\varepsilon_t dt + [b(\boldsymbol{x}_{s(x)+t}, \tilde{\boldsymbol{\xi}}^\varepsilon_t) - \nabla^\perp H(\boldsymbol{x}_{s(x)+t})]dt,\ \tilde{\zeta}^\varepsilon_0 = 0.$$

(Note that, for finite t, $\tilde{\zeta}^\varepsilon_t$ is of order $\sqrt{\varepsilon}$.) Then, by standard perturbation arguments and

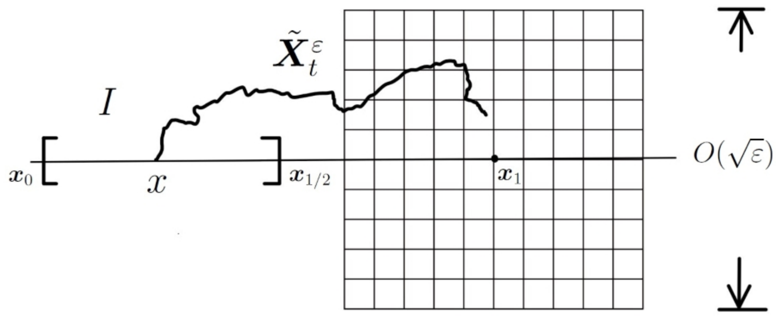


Figure 2.8: Higher order estimate on the distribution of $\tilde{\boldsymbol{X}}_t^\varepsilon$.

Grönwall's inequality, it can be shown that, uniformly in x such that $0 \le s(x) \le 1/2$, $0 \le t \le 1 - s(x)$, and $y \in \mathbb{T}^m$,

$$\mathbf{E}_{(x,y)}|\tilde{\boldsymbol{X}}_t^\varepsilon - \boldsymbol{x}_{s(x)+t} - \tilde{\zeta}_t^\varepsilon| = O(\varepsilon). \tag{2.49}$$

Therefore, understanding of the distribution of $\tilde{\zeta}_{1-s(x)}^\varepsilon$ would help one to understand the distribution of $\tilde{\boldsymbol{X}}_{1-s(x)}^\varepsilon$. However, it is not straightforward to study $\tilde{\zeta}_t^\varepsilon$ since $(\tilde{\boldsymbol{\xi}}_t^\varepsilon, \tilde{\zeta}_t^\varepsilon)$ is not a Markov process. We introduce a related process ζ_t^ε defined using the original Markov process $\boldsymbol{\xi}_t^\varepsilon$, apply the local limit theorem to $(\boldsymbol{\xi}_t^\varepsilon, \zeta_t^\varepsilon)$, and use the Girsanov theorem to get the desired estimate. Namely, let ζ_t^ε, $s \le t$, be defined by:

$$d\zeta_t^\varepsilon = \nabla(\nabla^\perp H)(\boldsymbol{x}_{s(x)+t})\zeta_t^\varepsilon dt + [b(\boldsymbol{x}_{s(x)+t}, \boldsymbol{\xi}_t^\varepsilon) - \nabla^\perp H(\boldsymbol{x}_{s(x)+t})]dt,\ \zeta_0^\varepsilon = 0. \tag{2.50}$$

The following result is a version of the local limit theorem [1] (cf. Theorem 3.6.3) adapted to our case.

Theorem 2.5.8. *Let $g : [0,1] \times \mathbb{T}^m \to \mathbb{R}^2$ be a C^∞ function such that $g(t,\cdot)$ spans $\mathbb{R}^2$ and*

$\int_{\mathbb{T}^m} g(t,y)d\mu(y) = 0$ *for all* $t \geq 0$*, where* μ *is the invariant measure of* $\boldsymbol{\xi}_t^\varepsilon$*. Then a local limit theorem holds for the following random variable as* $\varepsilon \to 0$ *uniformly in* $(x,y) \in I \times \mathbb{T}^m$*,*

$$S^\varepsilon := \frac{1}{\varepsilon}\int_0^{1-s(x)} g(s(x)+t, \boldsymbol{\xi}_t^\varepsilon)dt.$$

Namely, there exists an invertible covariance matrix $B(s)$ *continuous in* s *such that*

$$\lim_{\varepsilon\to 0}\left|\frac{2\pi}{\varepsilon}\sqrt{\det B(s(x))}\cdot \mathbf{P}_{(x,y)}(S^\varepsilon - u \in [0,1)^2) - \exp\!\left(-\frac{\varepsilon\langle B(s(x))^{-1}u, u\rangle}{2}\right)\right| = 0, \tag{2.51}$$

uniformly in $u \in \mathbb{R}^2$*,* $x \in I$*, and* $y \in \mathbb{T}^m$*.*

The second term in (2.51) is non-trivial even when u takes large values (of order $1/\sqrt{\varepsilon}$), which is exactly the situation we are dealing with. Following (2.50), we solve explicitly

$$\zeta_{1-s(x)}^\varepsilon = \int_0^{1-s(x)} U_{s(x)+t,1}(b(\boldsymbol{x}_{s(x)+t}, \boldsymbol{\xi}_t^\varepsilon) - \nabla^\perp H(\boldsymbol{x}_{s(x)+t}))dt,$$

where $U_{t,s}$ solves the differential equation

$$dU_{t,s} = \nabla(\nabla^\perp H)(\boldsymbol{x}_s)U_{t,s}ds,$$

and $U_{t,t}$ is the identity matrix. Since $\boldsymbol{x}_t$ is deterministic, the integrand can be treated as a function only of time t and $\boldsymbol{\xi}_t^\varepsilon$. Moreover, for each t, the integrand has zero mean w.r.t. the invariant measure and spans $\mathbb{R}^2$, since $U_{t,1}$ is deterministic and non-singular and, for each x, $\{b(x,y) -$

$\nabla^\perp H(x) : y \in \mathbb{T}^m\}$ spans $\mathbb{R}^2$ by assumption *(H4′)*. Then Theorem 2.5.8 implies that

$$\mathbf{P}_{(x,y)}\left(\frac{1}{\varepsilon}\zeta^\varepsilon_{1-s(x)} \in [j, j+1) \times [k, k+1)\right) \geq \frac{\varepsilon}{4\pi\sqrt{\det B(s(x))}} \exp\left(-\frac{\varepsilon\langle B(s(x))^{-1}(j,k),(j,k)\rangle}{2}\right) \tag{2.52}$$

for all ε small enough, $-1/\sqrt{\varepsilon} \leq j, k \leq 1/\sqrt{\varepsilon}$, $x \in I$, and $y \in \mathbb{T}^m$. Finally, we compare $(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t, \tilde{\zeta}^\varepsilon_t)$ with $(\boldsymbol{X}^\varepsilon_t, \boldsymbol{\xi}^\varepsilon_t, \zeta^\varepsilon_t)$. Since the added drift $c(x, y)$ in the equation of $\tilde{\boldsymbol{\xi}}^\varepsilon_t$ is small compared to the diffusion term $\frac{1}{\sqrt{\varepsilon}}\sigma(y)$, it is not hard to verify that, using the Girsanov theorem, for all ε small enough, $-1/\sqrt{\varepsilon} \leq j, k \leq 1/\sqrt{\varepsilon}$, $x \in I$, and $y \in \mathbb{T}^m$,

$$\mathbf{P}_{(x,y)}\left(\frac{1}{\varepsilon}\tilde{\zeta}^\varepsilon_{1-s(x)} \in [j, j+1) \times [k, k+1)\right) \geq \frac{1}{2}\mathbf{P}_{(x,y)}\left(\frac{1}{\varepsilon}\zeta^\varepsilon_{1-s(x)} \in [j, j+1) \times [k, k+1)\right). \tag{2.53}$$

Step 3. We proved that $\boldsymbol{x}_1 + \tilde{\zeta}^\varepsilon_{1-s(x)}$ reaches the $O(\varepsilon)-$sized boxes with probabilities bounded from below. We also proved that $\tilde{\boldsymbol{X}}^\varepsilon_{1-s(x)}$ is $O(\varepsilon)$-close to $\boldsymbol{x}_1 + \tilde{\zeta}^\varepsilon_{1-s(x)}$ in L^1. Let us take one generic pair (j, k), let $B^{\varepsilon,K}_{j,k} = \boldsymbol{x}_1 + [(j-K)\varepsilon, (j+1+K)\varepsilon) \times [(k-K)\varepsilon, (k+1+K)\varepsilon)$, and study the distribution of $(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t)$ with the initial point in $B^{\varepsilon,K}_{j,k} \times \mathbb{T}^m$ after time of order $O(\varepsilon)$.

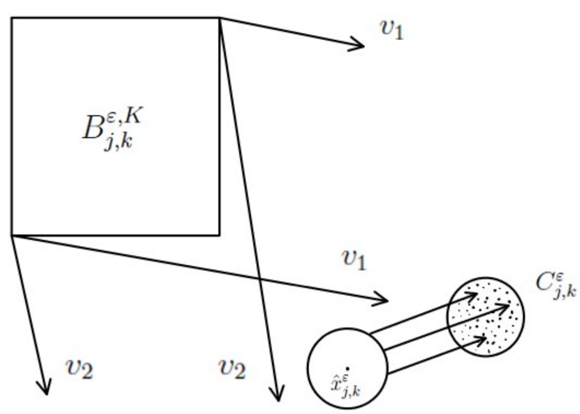


Figure 2.9: Common component of the distributions.

Lemma 2.5.9. *For each $\kappa > 0$, $K > 0$, and $\hat{y} \in \mathbb{T}^m$, there exist $t_2 > 0$, $c > 0$, and, for each pair (j,k), a point $\hat{x}^{\varepsilon}_{j,k}$ such that, for each $(x,y) \in B^{\varepsilon,K}_{j,k} \times \mathbb{T}^m$ and all ε sufficiently small,*

$$\mathbf{P}_{(x,y)}(\tau_\kappa < t_2\varepsilon) \geq c,$$

where $\tau_\kappa = \inf\{t : \tilde{\boldsymbol{X}}^{\varepsilon}_t \in B(\hat{x}^{\varepsilon}_{j,k}, \kappa\varepsilon),\ \tilde{\boldsymbol{\xi}}^{\varepsilon}_t \in B(\hat{y}, \kappa)\}$.

Proof. Recall the definition of $\boldsymbol{x}_1$ at the beginning of Step 2 (see Figure 2.8). By assumption *(H4′)*, $\{b(\boldsymbol{x}_1, y) : y \in \mathbb{T}^m\}$ spans $\mathbb{R}^2$. So there exist $y_1, y_2 \in \mathbb{T}^m$ such that $v_1 := b(\boldsymbol{x}_1, y_1)$ and $v_2 := b(\boldsymbol{x}_1, y_2)$ span $\mathbb{R}^2$. Let us consider the set $S_{j,k} = \bigcap_{x \in B^{\varepsilon,K}_{j,k}} \{x + av_1 + bv_2 : a, b \geq 0\}$. Then it is easy to see that, there exist a constant $t_2 > 0$ and, for each pair (j,k), a point $\hat{x}^{\varepsilon}_{j,k} \in S_{j,k}$ such that for all $x \in B^{\varepsilon,K}_{j,k}$, $\hat{x}^{\varepsilon}_{j,k} = x + a_x\varepsilon v_1 + b_x\varepsilon v_2$ and $0 < a_x, b_x < t_2/5$. There exists $\delta > 0$ such that for each $x \in B(\boldsymbol{x}_1, 2\delta)$ and each y in $B(y_i, 2\delta)$, $|b(x,y) - v_i| < \kappa/t_2$, $i = 1, 2$. Let M be the upper bound of vector $b(x,y)$. For all ε sufficiently small, the probability of the following event, denoted by E, has a lower bound, denoted by c, that only depends on t_2, κ, M, y_1, y_2, $\hat{y}$, δ, and not on the starting point $(x,y) \in B(\boldsymbol{x}_1, \delta) \times \mathbb{T}^m$, thus not on (j,k):

$$E = \left\{ \begin{array}{c} \tau_1 < (t_2 \wedge \kappa/M)\varepsilon/5;\ \tilde{\boldsymbol{\xi}}^{\varepsilon}_{\tau_1+t} \in B(y_1, 2\delta),\ t \in [0, a_x\varepsilon];\ \tau_2 < \tau_1 + a_x + (t_2 \wedge \kappa/M)\varepsilon/5; \\ \tilde{\boldsymbol{\xi}}^{\varepsilon}_{\tau_2+t} \in B(y_2, 2\delta),\ t \in [0, b_x\varepsilon];\ \tau_3 < \tau_2 + b_x\varepsilon + (t_2 \wedge \kappa/M)\varepsilon/5 \end{array} \right\},$$

where $\tau_1 = \inf\{t \geq 0 : \tilde{\boldsymbol{\xi}}^{\varepsilon}_t \in B(y_1, \delta)\}$, $\tau_2 = \inf\{t \geq \tau_1 + a_x\varepsilon : \tilde{\boldsymbol{\xi}}^{\varepsilon}_t \in B(y_2, \delta)\}$, and $\tau_3 = \inf\{t \geq \tau_2 + b_x\varepsilon : \tilde{\boldsymbol{\xi}}^{\varepsilon}_t \in B(\hat{y}, \kappa)\}$. If E is a subset of the event $\{\tau_\kappa < t_2\varepsilon\}$, then the lemma is proved. To show the inclusion, note that on E,

$$|\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_3} - \hat{x}^{\varepsilon}_{j,k}| = |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_3} - (x + a_x\varepsilon v_1 + b_x\varepsilon v_2)|$$

$$
\begin{aligned}
&\leq |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_3} - \tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_2+b_x\varepsilon}| + |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_2+b_x\varepsilon} - (\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_2} + b_x\varepsilon v_2)| + |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_2} - \tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_1+a_x\varepsilon}| \\
&\qquad + |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_1+a_x\varepsilon} - (\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_1} + a_x\varepsilon v_1)| + |\tilde{\boldsymbol{X}}^{\varepsilon}_{\tau_1} - x| \\
&\leq \kappa\varepsilon.
\end{aligned}
$$

Besides, by the definition of τ_3, $\tilde{\boldsymbol{\xi}}^{\varepsilon}_{\tau_3} \in B(\hat{y}, \kappa)$. Thus $\tau_\kappa \leq \tau_3 < t_2\varepsilon$ on E. □

From now on, let $\hat{y}$ be the point in assumption *(H5)* such that the parabolic Hörmander condition holds at $(\boldsymbol{x}_1, \hat{y})$ and let $p^{\varepsilon}_t((x,y),\cdot)$ be the density of $(\tilde{\boldsymbol{X}}^{\varepsilon}_{t\varepsilon}, \tilde{\boldsymbol{\xi}}^{\varepsilon}_{t\varepsilon})$ starting at (x,y).

Lemma 2.5.10. *There exists $\kappa > 0$ such that for each $\hat{x} \in B(\boldsymbol{x}_1, \kappa)$ and all ε sufficiently small, there is a domain $C^{\varepsilon}_{\hat{x},\hat{y}} \subset V^{\varepsilon} \times \mathbb{T}^m$ with $\lambda(C^{\varepsilon}_{\hat{x},\hat{y}}) > \kappa\varepsilon^2$ and $p^{\varepsilon}_1((x,y),\cdot) > \kappa/\varepsilon^2$ on $C^{\varepsilon}_{\hat{x},\hat{y}}$ for $(x,y) \in B(\hat{x}, \kappa\varepsilon) \times B(\hat{y}, \kappa)$.*

Proof. Consider the stochastic processes that depend on the parameters (ε, x, y):

$$
\begin{aligned}
&d\theta^{\varepsilon,x,y}_t = b(x + \varepsilon\theta^{\varepsilon,x,y}_t, y + \eta^{\varepsilon,x,y}_t)dt,\ \theta^{\varepsilon,x,y}_0 = 0 \in \mathbb{R}^2, \\
&d\eta^{\varepsilon,x,y}_t = v(y + \eta^{\varepsilon,x,y}_t)dt + \varepsilon c(x + \varepsilon\theta^{\varepsilon,x,y}_t, y + \eta^{\varepsilon,x,y}_t)dt + \sigma(y + \eta^{\varepsilon,x,y}_t)dW_t,\ \eta^{\varepsilon,x,y}_0 = 0 \in \mathbb{R}^m.
\end{aligned}
\tag{2.54}
$$

Since, by assumption *(H5)*, the parabolic Hörmander condition for equation (2.1) holds at $(\boldsymbol{x}_1, \hat{y})$, it is not hard to see that, if (x,y) is close to $(\boldsymbol{x}_1, \hat{y})$ and ε is small, the parabolic Hörmander condition holds for (2.54) at 0 and the distribution of $(\theta^{\varepsilon,x,y}_t, \eta^{\varepsilon,x,y}_t)$ is absolutely continuous w.r.t. the Lebesgue measure ([29]). Moreover, if the density function, denoted by $\tilde{p}^{\varepsilon,x,y}_1(\theta,\eta)$, exists, it is continuous in $\varepsilon, x, y, \theta$, and η. Let $\hat{\theta}$ and $\hat{\eta}$ satisfy that $\tilde{p}^{0,\boldsymbol{x}_1,\hat{y}}_1(\hat{\theta},\hat{\eta}) > 0$. Then there exists $0 < \delta < 1$ such that $\tilde{p}^{\varepsilon,x,y}_1(\theta,\eta)$ exists and is greater than δ for all $0 < \varepsilon < \delta$, $x \in B(\boldsymbol{x}_1, \delta)$, $y \in B(\hat{y}, \delta)$, $\theta \in B(\hat{\theta}, \delta)$, and $\eta \in B(\hat{\eta}, \delta)$. For $\hat{x} \in B(\boldsymbol{x}_1, \delta/2)$, define $C^{\varepsilon}_{\hat{x},\hat{y}} = B(\hat{x} + \varepsilon\hat{\theta}, \varepsilon\delta/2) \times$

$B(\hat{y}+\hat{\eta},\delta/2)$. Then, for $(x,y)\in B(\hat{x},\varepsilon\delta/2)\times B(\hat{y},\delta/2)$, and $(x',y')\in C^{\varepsilon}_{\hat{x},\hat{y}}$, and $0<\varepsilon<\delta$, we have that

$$p_1^{\varepsilon}((x,y),(x',y'))=\frac{1}{\varepsilon^2}\tilde{p}^{\varepsilon,x,y}\left(\frac{x'-x}{\varepsilon},y'-y\right)>\frac{\delta}{\varepsilon^2}.$$

The result holds with $\kappa=(\delta/2)^{m+2}$. $\square$

Lemma 2.5.11. *For each $K>0$, there exist constants $c>0$ and $t_1>0$ such that for all $-1/\sqrt{\varepsilon}\le j,k\le 1/\sqrt{\varepsilon}$, there exists a measure $\pi^{\varepsilon}_{j,k}$ and a stopping time $\tilde{\eta}_3^{j,k}<t_1\varepsilon$ such that for each $(x,y)\in B^{\varepsilon,K}_{j,k}\times\mathbb{T}^m$, the distribution of $(\tilde{\boldsymbol{X}}^{\varepsilon}_{\tilde{\eta}_3^{j,k}},\tilde{\boldsymbol{\xi}}^{\varepsilon}_{\tilde{\eta}_3^{j,k}})$ starting at (x,y) has $\pi^{\varepsilon}_{j,k}$ as a component and $\pi^{\varepsilon}_{j,k}(V^{\varepsilon}\times\mathbb{T}^m)>c$ for all ε sufficiently small.*

Proof. We fix constant $\kappa>0$ such that the statements in Lemma 2.5.10 hold. Then, for the fixed κ, by Lemma 2.5.9, we fix $t_2>0$, $c'>0$, and the point $\hat{x}^{\varepsilon}_{j,k}$ for each pair (j,k) such that for all $(x,y)\in B^{\varepsilon,K}_{j,k}\times\mathbb{T}^m$ and ε small, $\mathbf{P}_{(x,y)}(\tau_\kappa<t_2\varepsilon)\ge c'$, where $\tau_\kappa=\inf\{t:\tilde{\boldsymbol{X}}^{\varepsilon}_t\in B(\hat{x}^{\varepsilon}_{j,k},\kappa\varepsilon),\ \tilde{\boldsymbol{\xi}}^{\varepsilon}_t\in B(\hat{y},\kappa)\}$. It follows from Lemma 2.5.10 that there is a domain $C^{\varepsilon}_{j,k}\subset V^{\varepsilon}\times\mathbb{T}^m$ with $\lambda(C^{\varepsilon}_{j,k})>\kappa\varepsilon^2$ and $p_1^{\varepsilon}((x,y),\cdot)>\kappa/\varepsilon^2$ on $C^{\varepsilon}_{j,k}$ for all $(x,y)\in B(\hat{x}^{\varepsilon}_{j,k},\kappa\varepsilon)\times B(\hat{y},\kappa)$. Then the result follows if we define $c=c'\kappa^2$, $\pi^{\varepsilon}_{j,k}=c'\kappa/\varepsilon^2\cdot\chi_{\{C^{\varepsilon}_{j,k}\}}\lambda$, $t_1=t_2+2$, and $\tilde{\eta}_3^{j,k}=\tau_\kappa\wedge t_2\varepsilon+\varepsilon<t_1\varepsilon$. $\square$

Now let us combine Step 2 and Step 3 together to get the following result concerning the total variation distance of $(\tilde{\boldsymbol{X}}_{\boldsymbol{\tau}_1},\tilde{\boldsymbol{\xi}}_{\boldsymbol{\tau}_1})$ with different starting points on $I\times\mathbb{T}^m$:

Lemma 2.5.12. *For each $(x,y)\in I\times\mathbb{T}^m$, let $\tilde{\mu}^{\varepsilon}_{x,y}$ be the measure induced by $(\tilde{\boldsymbol{X}}_{\boldsymbol{\tau}_1},\tilde{\boldsymbol{\xi}}_{\boldsymbol{\tau}_1})$ starting at (x,y). Then there exists $c>0$ such that $\mathrm{TV}(\tilde{\mu}^{\varepsilon}_{x,y},\tilde{\mu}^{\varepsilon}_{x',y'})<1-c$ for any $(x,y),(x',y')\in I\times\mathbb{T}^m$ and all ε sufficiently small.*

Proof. It suffices to show that there exist $c>0$ and a stopping time $\tilde{\eta}\le\boldsymbol{\tau}_1$ such that the total variation distance of $(\tilde{\boldsymbol{X}}_{\tilde{\eta}}, \tilde{\boldsymbol{\xi}}_{\tilde{\eta}})$ with different starting points on $I \times \mathbb{T}^m$ is no more than $1-c$. Recall the definitions of $s(x)$ and $\tilde{\zeta}^\varepsilon_t$ in Step 2. For the process $(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t)$ starting at $(x, y) \in I \times \mathbb{T}^m$, define

$$A^\varepsilon_{j,k} = \{\frac{1}{\varepsilon}\tilde{\zeta}^\varepsilon_{1-s(x)} \in [j, j+1) \times [k, k+1)\},$$

$$E^\varepsilon_K = \{|\tilde{\boldsymbol{X}}^\varepsilon_{1-s(x)} - \boldsymbol{x}_1 - \tilde{\zeta}^\varepsilon_{1-s(x)}| > K\varepsilon\} \cup \{\sup_{0 \le t \le 1-s(x)} |H(\tilde{\boldsymbol{X}}^\varepsilon_t)| > K\sqrt{\varepsilon}\}.$$

Using (2.52) and (2.53), we can find a constant $c' > 0$ such that, for all $x \in I$, $y \in \mathbb{T}^m$, ε sufficiently small, and $-1/\sqrt{\varepsilon} \le j, k \le 1/\sqrt{\varepsilon}$, $\mathbf{P}_{(x,y)}(A^\varepsilon_{j,k}) \ge c'\varepsilon$. And using (2.19) and (2.49) we can choose K large enough such that, for all $x \in I$, $y \in \mathbb{T}^m$, and ε sufficiently small, $\mathbf{P}_{(x,y)}(E^\varepsilon_K) < c'/100$. Let $\tilde{\eta}_2 = 1 - s(x) \wedge \boldsymbol{\tau}_1$. Then it is not hard to see that

$$\sum_{-1/\sqrt{\varepsilon} \le j,k \le 1/\sqrt{\varepsilon}} \mathbf{P}_{(x,y)}(A^\varepsilon_{j,k} \cap \{\tilde{\boldsymbol{X}}_{\tilde{\eta}_2} \notin B^{\varepsilon,K}_{j,k}\}) < c'/100.$$

Now let us define, for $(x, y) \in I \times \mathbb{T}^m$,

$$R^\varepsilon_{x,y} = \{(j,k) : -1/\sqrt{\varepsilon} \le j, k \le 1/\sqrt{\varepsilon}, \mathbf{P}_{(x,y)}(A^\varepsilon_{j,k} \cap \{\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}_2} \in B^{\varepsilon,K}_{j,k}\}) < c'\varepsilon/2\}.$$

Then we know that $|R^\varepsilon_{x,y}| < \frac{1}{50\varepsilon}$ since, for every $(j,k) \in R^\varepsilon_{x,y}$,

$$\mathbf{P}_{(x,y)}(A^\varepsilon_{j,k} \cap \{\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}_2} \notin B^{\varepsilon,K}_{j,k}\}) \ge \mathbf{P}_{(x,y)}(A^\varepsilon_{j,k}) - \mathbf{P}_{(x,y)}(A^\varepsilon_{j,k} \cap \{\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}_2} \in B^{\varepsilon,K}_{j,k}\}) \ge c'\varepsilon/2.$$

Let the constants $c'' > 0$, $t_1 > 0$, the stopping time $\tilde{\eta}^{j,k}_3 < t_1\varepsilon$, and $\pi^\varepsilon_{j,k}$ be defined as in

Lemma 2.5.11. Define

$$\pi^\varepsilon = \frac{1}{2}c'\varepsilon \sum_{-1/\sqrt{\varepsilon}\le j,k\le 1/\sqrt{\varepsilon}} \pi^\varepsilon_{j,k}, \qquad \hat{\pi}^\varepsilon_{x,y,x',y'} = \frac{1}{2}c'\varepsilon \sum_{(j,k)\in R^\varepsilon_{x,y}\cup R^\varepsilon_{x',y'}} \pi^\varepsilon_{j,k}.$$

In order to define the desired stopping time, we first run the process starting on $I \times \mathbb{T}^m$ for time $\tilde{\eta}_2$ (with overwhelming probability, it is the time for the deterministic motion with the same stating point to reach $\boldsymbol{x}_1$). Then we use the locations of both $\tilde{\zeta}^\varepsilon_{\tilde{\eta}_2}$ and $\tilde{X}^\varepsilon_{\tilde{\eta}_2}$ to determine whether the process continues and, if it continues, we choose the stopping time based on Lemma 2.5.11. Namely, we define

$$\tilde{\eta} = \tilde{\eta}_2 + \sum_{-1/\sqrt{\varepsilon}\le j,k\le 1/\sqrt{\varepsilon}} \chi(A^\varepsilon_{j,k} \cap \{\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}_2^{j,k}} \in B^{\varepsilon,K}_{j,k}\}) \cdot \tilde{\eta}_3^{j,k}(\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}_2}, \tilde{\boldsymbol{\xi}}^\varepsilon_{\tilde{\eta}_2}),$$

where $\tilde{\eta}_3^{j,k}(x,y)$ denotes the stopping time with initial condition (x,y). Then it follows from previous results that, for any pair $(x,y),(x',y') \in I \times \mathbb{T}^m$, there is a common component $\pi^\varepsilon - \hat{\pi}^\varepsilon_{x,y,x',y'}$ of the distributions of $(\tilde{\boldsymbol{X}}^\varepsilon_{\tilde{\eta}}, \tilde{\boldsymbol{\xi}}^\varepsilon_{\tilde{\eta}})$ starting from (x,y) and (x',y'), respectively. Moreover, $(\pi^\varepsilon - \hat{\pi}^\varepsilon_{x,y,x',y'})(V^\varepsilon \times \mathbb{T}^m) > c'c''$ since $|R^\varepsilon_{x,y}| < \frac{1}{50\varepsilon}$ and $|R^\varepsilon_{x',y'}| < \frac{1}{50\varepsilon}$. Therefore, the total variation is no more than $1 - c'c''$. □

Finally, we combine the result we just obtained with Step 1 to prove Lemma 2.5.1.

Proof of Lemma 2.5.1. As we discussed, the result is equivalent to the exponential convergence in total variation of $(\tilde{\boldsymbol{X}}^\varepsilon_{\tau_n}, \tilde{\boldsymbol{\xi}}^\varepsilon_{\tau_n})$ on $\gamma' \times \mathbb{T}^m$, uniformly in ε and in the initial distribution. Let $\mu^\varepsilon_{x,y}$ denote the measure on $\gamma' \times \mathbb{T}^m$ induced by $(\tilde{\boldsymbol{X}}^\varepsilon_{\tau_1}, \tilde{\boldsymbol{\xi}}^\varepsilon_{\tau_1})$ with the starting point $(x,y) \in \gamma' \times \mathbb{T}^m$. Then it suffices to prove that there exists $c > 0$ such that, for every pair (x,y), $(x',y') \in \gamma' \times \mathbb{T}^m$ and all ε sufficiently small, $\mathrm{TV}(\mu^\varepsilon_{x,y}, \mu^\varepsilon_{x',y'}) < 1 - c$, which follows from Lemma 2.5.6 and

Lemma 2.5.12. □

2.6 Proof of the main result

Since we deal with both the original and the auxiliary processes in this section, certain notation needs clarifying to avoid possible ambiguity: the process $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ represents not a generic process with arbitrary bounded $c(x,y)$ but only the auxiliary process with $\tilde{c}(x,y)$ satisfying (2.15); σ_n, τ_n are defined as in (2.6), and $\tilde{\sigma}_n$, $\tilde{\tau}_n$ represent the corresponding stopping time w.r.t. $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$. In (2.24), we defined $\mathcal{L}_c$ on each edge for a generic $c(x,y)$. Here we give a more explicit definition of $\mathcal{L}_{\tilde{c}} = \tilde{L}_k$ on the edge I_k:

$$
\begin{aligned}
\tilde{L}_k f(h) &= \frac{1}{2} A_k(h) f''(h) + \tilde{B}_k(h) f'(h),\\
A_k(h) &= \frac{2}{Q_k(h)} \int_{\gamma_k(h)} \frac{1}{|\nabla H(x)|} \int_0^\infty \mathbf{E}_\mu b_h(x,\xi_s) b_h(x,\xi_0) ds dl,\\
\tilde{B}_k(h) &= \frac{1}{Q_k(h)} \int_{\gamma_k(h)} \frac{1}{|\nabla H(x)|} \int_0^\infty \mathbf{E}_\mu \mathrm{div}_x (b_h(x,\xi_s)(b(x,\xi_0) - \nabla^\perp H(x))) ds dl.
\end{aligned}
$$

One can easily check that this is consistent with the definitions of $\bar{A}$ and $\bar{B}_c$ in (2.24), which are the generalizations of the coefficients defined in Section 2.2, and

$$
\frac{1}{2}[A_k(h_k) Q_k(h_k) f'(h_k)]' = \frac{1}{2} A_k(h_k) Q_k(h_k) f''(h_k) + \tilde{B}_k(h_k) Q_k(h_k) f'(h_k). \tag{2.55}
$$

Lemma 2.6.1. *For each $f \in \mathcal{D}$ and all ε sufficiently small, we have $\mathbf{E}_{\nu^\varepsilon} \int_0^{\sigma_1} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}_t^\varepsilon)) dt = 0$.*

Proof. Since the process $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$ on $M \times \mathbb{T}^m$ is recurrent, and the measure $\lambda \times \mu$ is the invariant measure, by Theorem 2.1 in [32], we have that for any measurable set $A \subset M$,

$$\int_M \chi_A(x) d\lambda(x) = \lambda(A) = (\lambda \times \mu)(A \times \mathbb{T}^m) = \mathbf{E}_{\nu^\varepsilon} \int_0^{\sigma_1} \chi_A(\tilde{X}_t^\varepsilon) dt.$$

Thus,

$$\int_M \mathcal{L}_{\tilde{c}} f(h(x)) d\lambda(x) = \mathbf{E}_{\nu^\varepsilon} \int_0^{\sigma_1} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}_t^\varepsilon)) dt.$$

So, it suffices to show that the left-hand side is zero. By (2.55) and (2.8),

$$\begin{aligned}
\int_M \mathcal{L}_{\tilde{c}} f(h(x)) d\lambda(x) &= \sum_{k=1}^{3} \int_{I_k} \tilde{L}_k f(h_k) Q_k(h_k) dh_k \\
&= \sum_{k=1}^{3} \int_{I_k} (\frac{1}{2} A_k(h_k) f''(h_k) + \tilde{B}_k(h_k) f'(h_k)) Q_k(h_k) dh_k \\
&= \sum_{k=1}^{3} \int_{I_k} \frac{1}{2} [A_k(h_k) Q_k(h_k) f'(h_k)]' dh_k \\
&= \frac{1}{2} \sum_{k=1}^{3} p_k \lim_{h_k \to O} f'(h_k) \\
&= 0. \qquad \square
\end{aligned}$$

Let us verify the analogue of (2.12) in the case of the auxiliary process $(\tilde{X}_t^\varepsilon, \tilde{\xi}_t^\varepsilon)$.

Proposition 2.6.2. *For each $f \in \mathcal{D}$ and $T > 0$,*

$$\mathbf{E}_{(x,y)}[f(h(\tilde{X}_\eta^\varepsilon)) - f(h(x)) - \int_0^\eta \mathcal{L}_{\tilde{c}} f(h(\tilde{X}_t^\varepsilon)) dt] \to 0$$

as $\varepsilon \to 0$, uniformly in $x \in M$, $y \in \mathbb{T}^m$, and η is a stopping time bounded by T.

Proof. We divide the time interval $[0,\eta]$ into visits to the separatrix. Since $\sigma_n \to \infty$,

$$
\begin{aligned}
&|\mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\eta})) - f(h(x)) - \int_0^{\eta} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt]| \\
&\le |\lim_{n\to\infty} \mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}_n})) - f(h(x)) - \int_0^{\tilde{\sigma}_n} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt]| \\
&\quad + |\lim_{n\to\infty} \mathbf{E}_{(x,y)}\mathbf{E}_{(\tilde{X}^{\varepsilon}_{\eta},\tilde{\xi}^{\varepsilon}_{\eta})}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}_n})) - f(h(\tilde{X}^{\varepsilon}_0)) - \int_0^{\tilde{\sigma}_n} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt]| \\
&\le 2 \sup_{(x,y)\in M\times\mathbb{T}^m} |\mathbf{E}_{(x,y)}[f(H(\tilde{X}^{\varepsilon}_{\tilde{\sigma}})) - f(H(x)) - \int_0^{\sigma} \mathcal{L}_{\tilde{c}} f(H(\tilde{X}^{\varepsilon}_s))ds]| \qquad (2.56) \\
&\quad + 2 \lim_{n\to\infty} \sup_{(x,y)\in\gamma\times\mathbb{T}^m} |\mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}_n})) - f(h(x)) - \int_0^{\sigma_n} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt]|. \qquad (2.57)
\end{aligned}
$$

Note that (2.56) converges to 0 due to Proposition 2.4.6, and (2.57) also converges to 0 since

$$
\begin{aligned}
&\lim_{n\to\infty} \sup_{(x,y)\in\gamma\times\mathbb{T}^m} |\mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}_n})) - f(h(x)) - \int_0^{\tilde{\sigma}_n} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt]| \\
&\le \lim_{n\to\infty} \sup_{(x,y)\in\gamma\times\mathbb{T}^m} \sum_{k=0}^{n-1} |\mathbf{E}_{(x,y)} \int_{\tilde{\sigma}_k}^{\tilde{\sigma}_{k+1}} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt| \\
&\le \lim_{n\to\infty} \sup_{(x,y)\in\gamma\times\mathbb{T}^m} \sum_{k=0}^{n} \left(2\cdot \mathrm{TV}(\nu^{k,\varepsilon}_{x,y}, \nu^{\varepsilon}) \cdot \sup_{(x',y')\in\gamma\times\mathbb{T}^m} |\mathbf{E}_{(x',y')} \int_0^{\tilde{\sigma}_1} \mathcal{L}_{\tilde{c}} f(h(\tilde{X}^{\varepsilon}_t))dt| \right) \\
&= 0,
\end{aligned}
$$

where the second inequality is due to Lemma 2.6.1 and the last equality follows from Proposition 2.4.6, Lemma 2.5.1, and Proposition 2.7.3. Thus, the desired result holds. $\square$

A technical result is needed to verify (2.12) for the original process $(X^{\varepsilon}_t, \xi^{\varepsilon}_t)$ on $M \times \mathbb{T}^m$.

Lemma 2.6.3. *For each $f \in \mathcal{D}$ and $\delta > 0$ there is $0 < \rho < 1$ such that, for all $x \in \gamma$, $y \in \mathbb{T}^m$,* *and all ε sufficiently small,*

$$\begin{aligned}&\sup_{\sigma'\leq\rho}|\mathbf{E}_{(x,y)}\sum_{n=0}^{\infty}\chi_{\{\sigma_n<\sigma'\}}[f(h(X^{\varepsilon}_{\tau_{n+1}}))-f(h(X^{\varepsilon}_{\sigma_n}))-\int_{\sigma_n}^{\tau_{n+1}}\mathcal{L}f(h(X^{\varepsilon}_s))ds]|\\&\quad\leq\delta\rho+\varepsilon^{\alpha}\delta\sum_{n=0}^{\infty}\mathbf{P}_{(x,y)}(\sigma_n<\rho),\end{aligned}\tag{2.58}$$

where σ' is a stopping time w.r.t. $\mathcal{F}_t^{X^{\varepsilon}_{\cdot}}$.

Proof. The result holds either with or without the integral term since nearly all of the time is spent from τ_n to σ_n. To be precise, by the strong Markov property, Corollary 2.4.17, and Proposition 2.7.3,

$$\begin{aligned}&\sup_{(x,y)\in\gamma\times\mathbb{T}^m}\sup_{\sigma'\leq\rho}|\mathbf{E}_{(x,y)}\sum_{n=0}^{\infty}\chi_{\{\sigma_n<\sigma'\}}\int_{\sigma_n}^{\tau_{n+1}}\mathcal{L}f(h(X^{\varepsilon}_s))ds|\\&\lesssim\sup_{(x,y)\in\gamma\times\mathbb{T}^m}\sup_{\sigma'\leq\rho}\sum_{n=0}^{\infty}|\mathbf{E}_{(x,y)}\chi_{\{\sigma_n<\sigma'\}}\mathbf{E}_{(X^{\varepsilon}_{\sigma_n},\xi^{\varepsilon}_{\sigma_n})}\tau_1|=O(\varepsilon^{\alpha}|\log\varepsilon|).\end{aligned}\tag{2.59}$$

Thus, it suffices to prove for all ε sufficiently small

$$\sup_{\sigma'\leq\rho}|\mathbf{E}_{(x,y)}\sum_{n=0}^{\infty}\chi_{\{\sigma_n<\sigma'\}}[f(h((X^{\varepsilon}_{\tau_{n+1}}))-f(h((X^{\varepsilon}_{\sigma_n}))|\leq\delta\rho+\varepsilon^{\alpha}\delta\sum_{n=0}^{\infty}\mathbf{P}_{(x,y)}(\sigma_n<\rho).\tag{2.60}$$

Let us prove this for $\tilde{X}^{\varepsilon}_t$ first using Proposition 2.6.2, then apply the Girsanov theorem to get the result for X^{ε}_t. Let $\tilde{\sigma}'$ be the analogue of σ' w.r.t. $\mathcal{F}_t^{\tilde{X}^{\varepsilon}_{\cdot}}$. Divide the time interval $[0,\tilde{\sigma}']$ into excursions using stopping times $\tilde{\sigma}_n$ and $\tilde{\tau}_n$:

$$\mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}'}))-f(h(x))-\int_0^{\tilde{\sigma}'}\mathcal{L}f(h(\tilde{X}^{\varepsilon}_t))dt]\tag{2.61}$$

$$=\mathbf{E}_{(x,y)}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}'\wedge\tilde{\sigma}}))-f(h(x))-\int_0^{\tilde{\sigma}'\wedge\tilde{\sigma}}\mathcal{L}f(h(\tilde{X}^{\varepsilon}_t))dt]\tag{2.62}$$

$$+\sum_{n=0}^{\infty}\mathbf{E}_{(x,y)}\left(\chi_{\{\tilde\sigma_n<\tilde\sigma'\}}[f(h(\tilde X^\varepsilon_{\tilde\tau_{n+1}\wedge\tilde\sigma'}))-f(h(\tilde X^\varepsilon_{\tilde\sigma_n}))-\int_{\tilde\sigma_n}^{\tilde\tau_{n+1}\wedge\tilde\sigma'}\mathcal{L}f(h(\tilde X^\varepsilon_t))dt]\right)\tag{2.63}$$
$$+\sum_{n=1}^{\infty}\mathbf{E}_{(x,y)}\left(\chi_{\{\tilde\tau_n<\tilde\sigma'\}}[f(h(\tilde X^\varepsilon_{\tilde\sigma_n\wedge\tilde\sigma'}))-f(h(\tilde X^\varepsilon_{\tilde\tau_n}))-\int_{\tilde\tau_n}^{\tilde\sigma_n\wedge\tilde\sigma'}\mathcal{L}f(h(\tilde X^\varepsilon_t))dt]\right).\tag{2.64}$$

Thus, (2.63) converges to 0 uniformly for all $x\in\gamma$ and $\tilde\sigma'\le\rho$ due to the convergence of (2.61), (2.62), and (2.64), by Proposition 2.6.2, Proposition 2.4.6, and Lemma 2.4.15 with Corollary 2.4.17, respectively. Note that (2.59) also holds for $\tilde X^\varepsilon_t$, hence we conclude that

$$\sup_{(x,y)\in\gamma\times\mathbb{T}^m}\sup_{\tilde\sigma'\le\rho}\sum_{n=0}^{\infty}\mathbf{E}_{(x,y)}\left(\chi_{\{\tilde\sigma_n<\tilde\sigma'\}}[f(h(\tilde X^\varepsilon_{\tilde\tau_{n+1}\wedge\tilde\sigma'}))-f(h(\tilde X^\varepsilon_{\tilde\sigma_n}))]\right)\to 0.\tag{2.65}$$

To apply the Girsanov theorem, we choose a sufficiently small time interval and use the fact that the transition probability of $(X^\varepsilon_t,\xi^\varepsilon_t)$ is similar to that of $(\tilde X^\varepsilon_t,\tilde\xi^\varepsilon_t)$ in the sense that they are absolutely continuous with density close to 1. More precisely, for any fixed $\delta'>0$, by the Girsanov theorem, we can choose a constant ρ_1 such that for all $0<\rho<\rho_1$,

$$\mu^\varepsilon_{x,y}\left(\left|\frac{d\tilde\mu^\varepsilon_{x,y}}{d\mu^\varepsilon_{x,y}}-1\right|<\delta'\right)\ge 1-\rho^2,$$

where $\mu^\varepsilon_{x,y}$ and $\tilde\mu^\varepsilon_{x,y}$ are the measures on $\mathbf{C}[0,\rho]$ induced by $(X^\varepsilon_t,\xi^\varepsilon_t)$ and $(\tilde X^\varepsilon_t,\tilde\xi^\varepsilon_t)$. Define

$$C'=\left\{\left|\frac{d\tilde\mu^\varepsilon_{x,y}}{d\mu^\varepsilon_{x,y}}-1\right|<\delta\right\}\subset\mathbf{C}[0,\rho],\ \Omega'=\{(X^\varepsilon_t,\ t\in[0,\rho])\in C'\}.$$

Note that the quantity in (2.60) primarily depends on the behavior of the processes on time interval $[0,\sigma']$ and event Ω'. Indeed, we can replace the stopping times τ_n by $\tau_n\wedge\sigma'$ with $O(\varepsilon^\alpha)$ errors. To replace Ω with Ω', we need several additional results that control the difference.

As in Corollary 2.4.17, we fix $\kappa > 0$ and choose a large constant $C > 0$ independent of ρ such that

$$\sum_{n=[C\log(C/\rho)\varepsilon^{-\alpha}]}^{\infty} \mathbf{P}_{(x,y)}(\sigma_n < \rho) \leq \sum_{n=[C\log(C/\rho)\varepsilon^{-\alpha}]}^{\infty} e^{\rho}(1-\kappa\varepsilon^{\alpha})^n \leq \delta'\rho\varepsilon^{-\alpha}.$$

Now we choose $\rho_2 > 0$ such that, for all $0 < \rho < \rho_2$, $C\rho\log(C/\rho) < \delta'$. Hence, for all $\sigma' \leq \rho$,

$$\sum_{n=0}^{\infty} \mathbf{P}_{(x,y)}(\{\sigma_n < \sigma'\} \setminus \Omega') \leq C\rho^2\log(C/\rho)\varepsilon^{-\alpha} + \delta'\rho\varepsilon^{-\alpha} \leq 2\delta'\rho\varepsilon^{-\alpha}.$$

Thus, with $K := \max_{I_k \sim O} |\lim_{h_k \in I_k, h_k \to O} f'(h_k)|$, we obtain

$$|\mathbf{E}_{(x,y)} \sum_{n=0}^{\infty} \chi_{\{\sigma_n < \sigma'\} \setminus \Omega'}[f(h(X^{\varepsilon}_{\tau_{n+1}\wedge\sigma'})) - f(h(X^{\varepsilon}_{\sigma_n}))]| \leq 2(K+1)\delta'\rho.$$

By following the same steps, we can choose $\rho_3 > 0$ such that for all $0 < \rho < \rho_3$,

$$|\mathbf{E}_{(x,y)} \sum_{n=0}^{\infty} \chi_{\{\tilde{\sigma}_n < \tilde{\sigma}'\} \setminus \Omega'}[f(h(\tilde{X}^{\varepsilon}_{\tilde{\tau}_{n+1}\wedge\tilde{\sigma}'})) - f(h(\tilde{X}^{\varepsilon}_{\tilde{\sigma}_n}))]| \leq 2(K+1)\delta'\rho. \tag{2.66}$$

It remains to consider

$$|\mathbf{E}_{(x,y)} \sum_{n=0}^{\infty} \chi_{\{\sigma_n < \sigma'\} \cap \Omega'}[f(h(X^{\varepsilon}_{\tau_{n+1}\wedge\sigma'})) - f(h(X^{\varepsilon}_{\sigma_n}))]|, \tag{2.67}$$

which can be written and estimated as, with F denoting the functional on $\mathbf{C}[0,\rho]$ found inside the expectation in (2.67),

$$\begin{aligned}\left|\int_{C'} F d\mu^{\varepsilon}_{x,y}\right| &= \left|\int_{C'} F d\tilde{\mu}^{\varepsilon}_{x,y} - \int_{C'} F\left(\frac{d\tilde{\mu}^{\varepsilon}_{x,y}}{d\mu^{\varepsilon}_{x,y}} - 1\right) d\mu^{\varepsilon}_{x,y}\right| \\ &\leq \left|\int_{C'} F d\tilde{\mu}^{\varepsilon}_{x,y}\right| + \delta' \int_{C'} |F| d\mu^{\varepsilon}_{x,y}.\end{aligned}$$

The first term is bounded by $2(K+2)\delta'\rho$ due to (2.65) and (2.66). The second term is simply bounded by $(K+1)\delta'\varepsilon^{\alpha}\sum_{n=0}^{\infty}\mathbf{P}_{(x,y)}(\sigma_n < \rho)$. Thus, we see that the left-hand side in (2.60) is no more than $(4K+6)\delta'(\rho + \varepsilon^{\alpha}\sum_{n=0}^{\infty}\mathbf{P}_{(x,y)}(\sigma_n < \rho))$ with finite K independent of δ' and ρ for all ε sufficiently small. It remains to take $\delta' = \delta/(4K+6)$. □

Proof of Proposition 2.3.1. Fix arbitrary $\delta > 0$. We divide the time interval $[0, \eta]$ into excursions from γ to γ' and from γ' to γ by using stopping times σ_n and τ_n:

$$\begin{aligned}&\mathbf{E}_{(x,y)}[f(h(X^{\varepsilon}_{\eta})) - f(h(x)) - \int_0^{\eta} \mathcal{L}f(h(X^{\varepsilon}_t))dt] \\ &= \mathbf{E}_{(x,y)}[f(h(X^{\varepsilon}_{\eta\wedge\sigma})) - f(h(x)) - \int_0^{\eta\wedge\sigma} \mathcal{L}f(h(X^{\varepsilon}_t))dt] \qquad (2.68)\\ &+ \sum_{n=0}^{\infty} \mathbf{E}_{(x,y)}\left(\chi_{\{\sigma_n<\eta\}}[f(h(X^{\varepsilon}_{\tau_{n+1}\wedge\eta})) - f(h(X^{\varepsilon}_{\sigma_n})) - \int_{\sigma_n}^{\tau_{n+1}\wedge\eta} \mathcal{L}f(h(X^{\varepsilon}_t))dt]\right) \qquad (2.69)\\ &+ \sum_{n=1}^{\infty} \mathbf{E}_{(x,y)}\left(\chi_{\{\tau_n<\eta\}}[f(h(X^{\varepsilon}_{\sigma_n\wedge\eta})) - f(h(X^{\varepsilon}_{\tau_n})) - \int_{\tau_n}^{\sigma_n\wedge\eta} \mathcal{L}f(h(X^{\varepsilon}_t))dt]\right). \qquad (2.70)\end{aligned}$$

Here (2.68) converges to 0 by Proposition 2.4.6 and (2.70) converges to 0 by Lemma 2.4.15 and Corollary 2.4.17. It remains to consider (2.69) and it suffices to consider instead

$$\sum_{n=0}^{\infty} \mathbf{E}_{(x,y)}\left(\chi_{\{\sigma_n<\eta\}}[f(h(X^{\varepsilon}_{\tau_{n+1}})) - f(h(X^{\varepsilon}_{\sigma_n})) - \int_{\sigma_n}^{\tau_{n+1}} \mathcal{L}f(h(X^{\varepsilon}_t))dt]\right) \qquad (2.71)$$

because the difference converges to 0 by Proposition 2.7.3. By Proposition 2.6.2, we choose

$0 < \rho < 1$ such that (2.58) holds for δ and all ε sufficiently small. We introduce the stopping times $\hat{\sigma}_n$ by letting $\hat{\sigma}_0 = \sigma$ and $\hat{\sigma}_n$ be the first of σ_k such that $\sigma_k - \hat{\sigma}_{n-1} \geq \rho$. It is clear that $\hat{\sigma}_{[T/\rho]} \geq T \geq \eta$. Hence, we can replace (2.71) by

$$\sum_{n=0}^{[T/\rho]-1} \mathbf{E}_{(x,y)}(\chi_{\{\hat{\sigma}_n<\eta\}}\mathbf{E}_{(X^\varepsilon_{\hat{\sigma}_n},\xi^\varepsilon_{\hat{\sigma}_n})}\sum_{k=0}^{\infty}\chi_{\{\sigma_k<\rho\}}[f(h(X^\varepsilon_{\tau_{n+1}}))-f(h(X^\varepsilon_{\sigma_n}))-\int_{\sigma_n}^{\tau_{n+1}}\mathcal{L}f(h(X^\varepsilon_t))dt])$$

and, by the strong Markov property, the difference is no more than

$$\sup_{(x,y)\in\gamma\times\mathbb{T}^m}\sup_{\sigma'\leq\rho}|\mathbf{E}_{(x,y)}\sum_{n=0}^{\infty}\chi_{\{\sigma_n<\sigma'\}}[f(h((X^\varepsilon_{\tau_{n+1}}))-f(h((X^\varepsilon_{\sigma_n}))-\int_{\sigma_n}^{\tau_{n+1}}\mathcal{L}f(h(X^\varepsilon_t))dt]|,$$

where σ' is a stopping time w.r.t. $\mathcal{F}_t^{X^\varepsilon_\cdot}$. Both of them can be bounded by $O(\delta)$ due to Lemma 2.6.3 and Corollary 2.4.17. $\square$

2.7 Technical proofs

2.7.1 Derivatives of the action-angle coordinates

In this section, we carefully estimate the first and second derivatives of $q(x)$, $Q(h)$, $\phi(x)$, $A(h,\phi)$, and $B_c(h,\phi)$ in order to prove (2.30). Our main tool is the Morse lemma. Note that we only need to verify the bounds near the separatrix, since the derivatives are uniformly bounded inside the domain. For $x, y \in \mathbb{R}^2$, let $x \to y$ denote the line segment connecting x and y. For $x, y \in \gamma(h)$ for certain h, let $x \xrightarrow{\gamma} y$ denote the piece of $\gamma(h)$ connecting x to y along the direction of $\nabla^\perp H$. Recall that $q(x) = \int_{l(H(x))\xrightarrow{\gamma}x}\frac{1}{|\nabla H|}dl$. To start with, using the Morse lemma, one can compute that $q(x) \lesssim |\log H(x)|$ and $Q(h) \lesssim |\log h|$. Then we make use of two special deterministic motions in the directions of $\nabla^\perp H$ and ∇H to calculate the first derivatives of $q(x)$ precisely:

$$d\boldsymbol{x}_t = \nabla^\perp H(\boldsymbol{x}_t)dt,$$

$$d\boldsymbol{y}_t = \nabla H(\boldsymbol{y}_t)dt.$$

It follows that

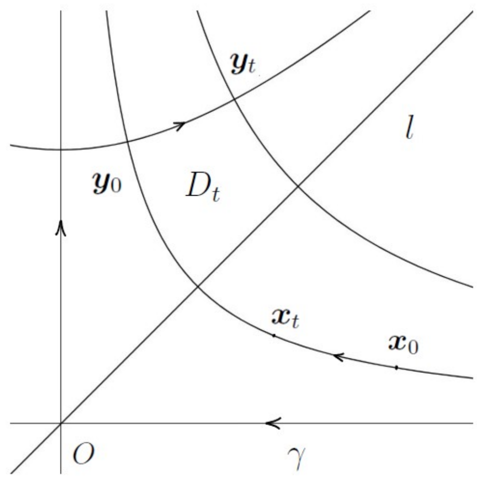


Figure 2.10: Motions tangent and perpendicular to the level curve.

$$q(\boldsymbol{x}_t) = q(\boldsymbol{x}_0) + t, \tag{2.72}$$

$$q(\boldsymbol{y}_t) = q(\boldsymbol{y}_0) + \int_{\partial D_t} \frac{\nabla H}{|\nabla H|^2} \cdot \boldsymbol{n} dl = q(\boldsymbol{y}_0) + \int_{D_t} \mathrm{div}(\frac{\nabla H}{|\nabla H|^2})dS, \tag{2.73}$$

where D_t is the region bounded by l, trajectory of $\boldsymbol{y}_s$, $0 \leq s \leq t$, $\gamma(H(\boldsymbol{y}_0))$, and $\gamma(H(\boldsymbol{y}_t))$, as shown in Figure 2.10. Thus, by differentiating (2.72) and (2.73) in t, we have the following equations:

$$\begin{aligned} &\nabla q(x) \cdot \nabla^\perp H(x) = 1, \\ &\nabla q(x) \cdot \nabla H(x) = |\nabla H(x)|^2 \int_{l(H(x)) \xrightarrow{\gamma} x} \mathrm{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{1}{|\nabla H|} dl. \end{aligned} \tag{2.74}$$

Therefore, with subscripts denoting the partial derivatives, by solving the linear system,

$$
\begin{aligned}
q_1' &= \frac{-H_2'}{{H_1'}^2 + {H_2'}^2} + H_1' p, \\
q_2' &= \frac{H_1'}{{H_1'}^2 + {H_2'}^2} + H_2' p,
\end{aligned}
$$

where $p(x) = \int_{l(H(x)) \xrightarrow{\gamma} x} \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{1}{|\nabla H|} dl$. Using the Morse lemma, one can compute $p(x) = O(1/H(x))$, since

$$
\left| \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \right| \lesssim \frac{1}{|\nabla H|^2}.
$$

Note that the non-degeneracy of the saddle point implies that $|H| \lesssim |\nabla H|^2$, and hence $\nabla q = O(|\nabla H|/H)$. The next step is to estimate ∇p. For all x, y close enough, with D denoting the region bounded by l, $x \to y$, $\gamma(H(x))$, and $\gamma(H(y))$, we have

$$
\begin{aligned}
|p(x) - p(y)| &= \left| \int_{\partial D} \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{\nabla H}{|\nabla H|^2} \cdot \mathbf{n} dl - \int_{x \to y} \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{\nabla H}{|\nabla H|^2} \cdot \mathbf{n} dl \right| \\
&\leq \int_D \left| \operatorname{div} \left[\operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{\nabla H}{|\nabla H|^2} \right] \right| dS + \int_{x \to y} \left| \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{1}{|\nabla H|} \right| dl \\
&\lesssim \int_D \frac{1}{|\nabla H|^4} dS + \int_{x \to y} \frac{1}{|\nabla H|^3} dl.
\end{aligned}
$$

Then one can obtain $|\nabla p| \lesssim |\nabla H|/H^2$ by using the Morse lemma again and, as a result, $|q_{ij}''| \lesssim |\nabla H|^2/H^2$, $1 \leq i, j \leq 2$. Similarly, we can estimate the derivatives of $Q(h)$. In fact, $Q'(h) = \int_{\gamma(h)} \operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{1}{|\nabla H|} dl = O(1/h)$ because of the estimate we had on $p(x)$. In addition,

$$
|Q''(h)| = \left| \int_{\gamma(h)} \operatorname{div} \left[\operatorname{div}(\frac{\nabla H}{|\nabla H|^2}) \frac{\nabla H}{|\nabla H|^2} \right] \frac{1}{|\nabla H|} dl \right| \lesssim \int_{\gamma(h)} \frac{1}{|\nabla H|^5} dl = O(1/h^2).
$$

Since $\phi(x) = q(x)/Q(H(x))$, $|\nabla\phi| \lesssim |\nabla H|/H$ and $|\phi''_{ij}| \lesssim |\nabla H|^2/H^2$, $1 \le i, j \le 2$. Finally, we estimate the derivatives w.r.t. h of a general function $F(h, \phi) = \tilde{F}(x_1, x_2)$ with $\tilde{F}$ having bounded first and second derivatives. By computing the inverse of the Jacobian of (H, ϕ) and using the first equation in (2.74),

$$F'_h = \frac{\tilde{F}'_1\phi'_2 - \tilde{F}'_2\phi'_1}{H'_1\phi'_2 - H'_2\phi'_1} = (\tilde{F}'_1\phi'_2 - \tilde{F}'_2\phi'_1)Q(H(x_1, x_2)). \tag{2.75}$$

We deduce that $F'_h = O(|\log h|/h)$ and, using F'_h instead of F in (2.75), $F''_{hh} = O(|\log h|^2/h^3)$. Similarly, one can obtain $F'_\phi = O(|\log h|)$ and $F''_{\phi\phi} = O(|\log h|/h)$.

2.7.2 Exit from neighborhoods of critical points

In this section, we obtain estimates for the exit time from the neighborhoods of the critical points, including extremum points and saddle points. Recall the notation in Section 2.4: O is a saddle point, O' is an extremum point, U is a domain bounded by the separatrix, and $\eta(h) = \inf\{t : |H(\tilde{X}^\varepsilon_t) - H(O')| = h\}$. Recall the function u defined in (2.17) and let us define

$$\boldsymbol{A}(x) = \int_{\mathbb{T}^m} \nabla_y u(x, y)\sigma(y)\sigma(y)^\intercal \nabla_y u(x, y)^\intercal d\mu(y). \tag{2.76}$$

Using assumption *(H4′)*, one can see that $\boldsymbol{A}(x)$ is positive-definite. Indeed, if there exist a point x and a vector $v \ne 0$ such that $v^\intercal \boldsymbol{A}(x)v = 0$, then, since $\sigma\sigma^\intercal$ is positive-definite, $v^\intercal \nabla_y u(x, y) = 0$ for all $y \in \mathbb{T}^m$. Namely, $v^\intercal u(x, y)$ is constant, and $v^\intercal(b(x, y) - \bar{b}(x)) = L(v^\intercal u(x, y)) = 0$ on $\mathbb{T}^m$, which contradicts with *(H4′)*.

Lemma 2.7.1. *Recall the definition of $B_c(x)$ in (2.23). If O' is a minimum point, then $B_c(O') >$*

0*; if* O' *is a maximum point, then* $B_c(O') < 0$.

Proof. We prove the result in the case of minimum point. The other case can be treated similarly. Since O' is a critical point and $Lu(x,y) = -(b(x,y) - \bar{b}(x))$, we have $\nabla H(O') = 0$ and

$$
\begin{aligned}
B_c(O') &= \int_{\mathbb{T}^m} [\nabla_x u_h(O',y) b(O',y) + \nabla_y u_h(O',y) c(x,y)] d\mu(y) \\
&= \int_{\mathbb{T}^m} u(O',y)^\intercal \nabla^2 H(O') b(O',y) d\mu(y) \\
&= -\int_{\mathbb{T}^m} u(O',y)^\intercal \nabla^2 H(O') Lu(O',y) d\mu(y) \\
&= -\sum_{1\le i,j\le 2} \frac{\partial^2}{\partial x_i \partial x_j} H(O') \int_{\mathbb{T}^m} u_i(O',y) Lu_j(O',y) d\mu(y) \\
&= -\sum_{1\le i,j\le 2} \frac{\partial^2}{\partial x_i \partial x_j} H(O') \int_{\mathbb{T}^m} \frac{1}{2}(u_i Lu_j + u_j Lu_i)(O',y) d\mu(y) \\
&= \frac{1}{2} \sum_{1\le i,j\le 2} \frac{\partial^2}{\partial x_i \partial x_j} H(O') \left(\boldsymbol{A}(O')_{i,j} - \int_{\mathbb{T}^m} L(u_i u_j) d\mu(y) \right) \\
&= \frac{1}{2} \sum_{1\le i,j\le 2} \frac{\partial^2}{\partial x_i \partial x_j} H(O') \boldsymbol{A}(O')_{i,j}.
\end{aligned}
$$

This is positive since both $\nabla^2 H$ and $\boldsymbol{A}$ are positive definite at O'. $\square$

Lemma 2.7.2. *For each* $\kappa > 0$*, there exists* $\delta > 0$ *such that*

$$\mathbf{E}_{(x,y)} \eta(\delta) \le \kappa$$

for all x *satisfying* $|H(x) - H(O')| < \delta$*,* $y \in \mathbb{T}^m$*, and* ε *sufficiently small.*

Proof. Without loss of generality, we assume O' to be a minimum point. Similarly to (2.21), we apply Ito's formula to $u_h(\tilde{X}^\varepsilon_{\eta(\delta)\wedge 1}, \tilde{\xi}^\varepsilon_{\eta(\delta)\wedge 1})$ and we obtain

$$H(\tilde{X}^\varepsilon_{\eta(\delta)\wedge 1}) = H(x) + \int_0^{\eta(\delta)\wedge 1} \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)^\top \sigma(\tilde{\xi}^\varepsilon_s) dW_s + \varepsilon(u_h(x,y) - u_h(\tilde{X}^\varepsilon_{\eta(\delta)\wedge 1}, \tilde{\xi}^\varepsilon_{\eta(\delta)\wedge 1}))$$
$$+ \int_0^{\eta(\delta)\wedge 1} [\nabla_x u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot b(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) + \nabla_y u_h(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s) \cdot c(\tilde{X}^\varepsilon_s, \tilde{\xi}^\varepsilon_s)] ds.$$

By taking the expectation on both sides and using Corollary 2.3.4, we obtain

$$\mathbf{E}_{(x,y)} \int_0^{\eta(\delta)\wedge 1} B_c(\tilde{X}^\varepsilon_s) ds < \delta + O(\varepsilon). \tag{2.77}$$

Due to Lemma 2.7.1, $B_c(O') > 0$. Hence, we can choose δ to be small enough such that $B_c(\tilde{X}^\varepsilon_s) > B_c(O')/2 > 0$ before $\eta(\delta)$. Thus, it follows from (2.77) that, for all ε sufficiently small,

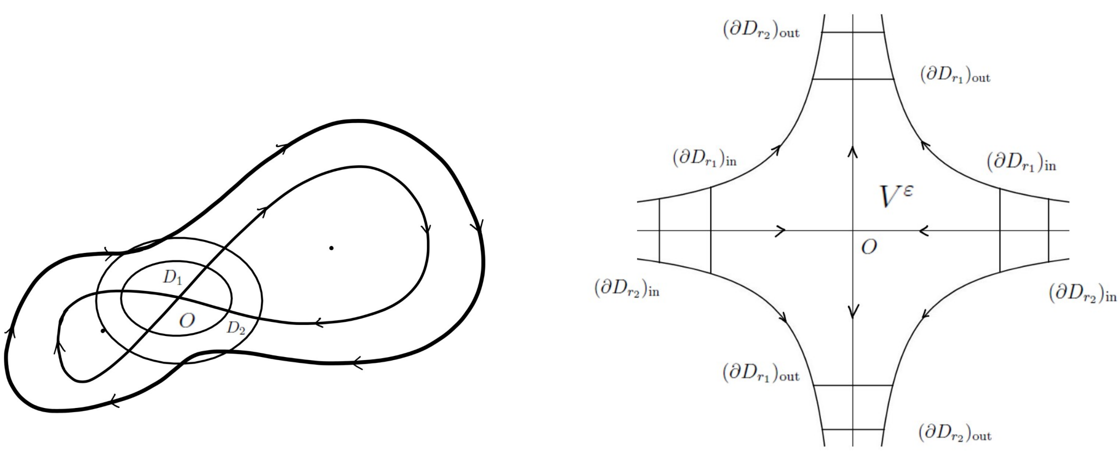
$$\mathbf{E}_{(x,y)}(\eta(\delta) \wedge 1) \leq 4\delta / B_c(O').$$

Then $\mathbf{P}_{(x,y)}(\eta(\delta) \geq 1) \leq 4\delta/B_c(O')$ for all x satisfying $|H(x) - H(O')| < \delta$, $y \in \mathbb{T}^m$ by Markov's inequality. Then, one can obtain the desired result using the Markov property by the fact that $\mathbf{E}_{(x,y)}\eta(\delta) \leq \mathbf{E}_{(x,y)}(\eta(\delta)\wedge 1) + \mathbf{E}_{(x,y)}(\eta(\delta)\chi_{\{\eta(\delta)>1\}})$. □

Proposition 2.7.3. *Let $0 < \alpha < 1/2$, $V^\varepsilon = \{x : |H(x) - H(O)| < \varepsilon^\alpha\}$, and $\tau = \inf\{t : \tilde{X}^\varepsilon_t \notin V^\varepsilon\}$. Then, uniformly in $0 < \alpha < 1/2$ and $(x,y) \in V^\varepsilon \times \mathbb{T}^m$, as $\varepsilon \downarrow 0$,*

$$\mathbf{E}_{(x,y)}\tau = O(\varepsilon^{2\alpha}|\log\varepsilon|).$$

To prove Proposition 2.7.3, it is more convenient to consider the process $(\tilde{\boldsymbol{X}}^\varepsilon_t, \tilde{\boldsymbol{\xi}}^\varepsilon_t)$, define the stopping time $\boldsymbol{\tau} = \inf\{t : \tilde{\boldsymbol{X}}^\varepsilon_t \notin V^\varepsilon\}$, and then prove that $\mathbf{E}_{(x,y)}\boldsymbol{\tau} = O(\varepsilon^{2\alpha-1}|\log\varepsilon|)$. We need careful analysis of the behavior of the processes near the saddle point and away from the saddle point. The latter is easier to deal with since there is no degeneracy, while the former needs us to, again, use the Morse Lemma to make concrete computations. For simplicity, we prove the result in the case of $(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)$ without the additional drift $c(x, y)$ since it can be seen in the proof that the additional terms induced by $c(x, y)$ are always relatively small. We prove that there exist two neighborhoods, $D_1 \subset D_2$, of O (as shown in Figure 2.11a), such that, in V^ε, it takes the process $O(|\log \varepsilon|)$ time to escape from D_2, and $O(1)$ time to return to D_1. Since $x \in V^\varepsilon$ is two-dimensional, we denote $x = (p, q)$. To avoid confusion brought by convoluted formulas, we assume the saddle point is the origin and the Hamiltonian $H(x) = pq$ in a small neighborhood of O. This assumption is not restrictive because, as in the proof of Lemma 2.5.3, we can use the Morse Lemma and perform a random change of time with the multiplier bounded from below and above, which will not change the order of the expected time. For $r > 0$, we denote D_r to be the region in V^ε with $|p| \leq r$ and $|q| \leq r$, $(\partial D_r)_{\text{in}} = \{|p| = r\} \cap V^\varepsilon$, $(\partial D_r)_{\text{out}} = \{|q| = r\} \cap V^\varepsilon$, and choose $r_3 > 0$ small enough that $H(x) = pq$ in D_{r_3}.



(a) Two neighborhoods of O.

(b) D_{r_1} and D_{r_2}.

Figure 2.11: Transitions between two boundaries.

Lemma 2.7.4. *There exist $r_1, r_2 > 0$ such that, uniformly in $(x, y) \in D_{r_1} \times \mathbb{T}^m$, as $\varepsilon \downarrow 0$,*

$$\mathbf{E}_{(x,y)}\bar{\tau} = O(|\log \varepsilon|),$$

where $\bar{\tau} = \inf\{t : \boldsymbol{X}_t^\varepsilon \notin D_{r_2}\}$.

Proof. We denote $\eta_t^\varepsilon = (\boldsymbol{X}_t^\varepsilon)_2$ and study the behavior of η_t^ε inside $B(0, r_3)$. All the computations below concern $\boldsymbol{X}_t^\varepsilon$ before leaving D_{r_3}. As in (2.19), we can write the equation for η_t^ε,

$$\begin{aligned}\eta_t^\varepsilon = q + \int_0^t \eta_s^\varepsilon ds + \sqrt{\varepsilon}\int_0^t \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s \\ + \varepsilon\int_0^t \nabla_x u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds + \varepsilon(u_2(x, y) - u_2(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)).\end{aligned}$$

Introduce $\hat{\eta}_t^\varepsilon$, which is close to η_t^ε:

$$\hat{\eta}_t^\varepsilon = q + \int_0^t \eta_s^\varepsilon ds + \sqrt{\varepsilon}\int_0^t \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s + \varepsilon\int_0^t \nabla_x u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds.$$

Let $F(q) = \int_0^q e^{-z^2}\int_0^z e^{w^2}dwdz$, which satisfies $F(q) \sim \frac{1}{2}\log q$, as $p \to \infty$, and has bounded derivatives up to the third order. Then we can choose a large $C > 0$, such that $|F'|, |F''|, |F'''|$, $|u(x, y)|, |\nabla u(x, y)|, |\nabla^2 u(x, y)|$ are bounded by C. Recall from (2.76) that the vector-valued function $\nabla_y u_2(x, y)^\mathsf{T}\sigma(y)$ has non-degenerate average w.r.t. μ, in the sense that $\boldsymbol{A}_{22}(x) > 0$. Let $A_0 = \boldsymbol{A}_{22}(O) > 0$ and let $0 < r_1 < r_2 < r_3$ be such that $A_0(1 - 1/(2C)) < \boldsymbol{A}_{22}(x) < A_0(1 + 1/(2C))$ in D_{r_2}, as shown in Figure 2.11b. Let $\bar{r}_2 = \frac{r_3 + r_2}{2}$. Define function $f(q)$ (that depends on ε) and compute its derivatives:

$$
\begin{aligned}
f(q) &= 2(F(\frac{\bar{r}_2}{\sqrt{A_0\varepsilon}}) - F(\frac{q}{\sqrt{A_0\varepsilon}})), \quad f'(q) = -\frac{2}{\sqrt{A_0\varepsilon}}F'(\frac{q}{\sqrt{A_0\varepsilon}}), \\
f''(q) &= -\frac{2}{A_0\varepsilon}F''(\frac{q}{\sqrt{A_0\varepsilon}}), \qquad\qquad f'''(q) = -\frac{2}{(\sqrt{A_0\varepsilon})^3}F'''(\frac{q}{\sqrt{A_0\varepsilon}}).
\end{aligned} \tag{2.78}
$$

Furthermore, f satisfies the differential equations:

$$
\begin{cases}
\frac{1}{2}A_0\varepsilon f'' + qf' = -1 \\
f(-\bar{r}_2) = f(\bar{r}_2) = 0
\end{cases}. \tag{2.79}
$$

By Lemma 2.3.2, there is a function $v_2(x,y)$ that is bounded together with its derivatives such that

$$
Lv_2(x,y) = |\nabla_y u_2(x,y)\sigma(y)|^2 - \boldsymbol{A}_{22}(x).
$$

where L is the generator of the process ξ_t (see (2.7)). Since $|\eta_t^\varepsilon - \hat{\eta}_t^\varepsilon| = O(\varepsilon)$ and $\bar{r}_2 > r_2$, we can apply Ito's formula to $v_2(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)f''(\eta_t^\varepsilon)$ for $0 \le t \le \bar{\tau}$:

$$
\begin{aligned}
v_2(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)f''(\eta_t^\varepsilon) =& v_2(x,y)f''(q) + \int_0^t \nabla_x(v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)f''(\eta_s^\varepsilon)) \cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds \\
&+ \frac{1}{\varepsilon}\int_0^t Lv_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)f''(\eta_s^\varepsilon)ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^{\mathsf{T}} f''(\eta_s^\varepsilon)\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s.
\end{aligned}
$$

Hence it follows that

$$
\begin{aligned}
&\int_0^t (|\nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 - \boldsymbol{A}_{22}(\boldsymbol{X}_s^\varepsilon)) f''(\eta_s^\varepsilon) ds \\
&\quad = \varepsilon \left(v_2(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon) f''(\eta_t^\varepsilon) - v_2(x,y) f''(q)\right) - \varepsilon \int_0^t \nabla_x (v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) f''(\eta_s^\varepsilon)) \cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) ds \\
&\qquad - \sqrt{\varepsilon} \int_0^t f''(\eta_s^\varepsilon) \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s \\
&\quad = O(1) + O(\frac{1}{\sqrt{\varepsilon}}) \cdot t - \sqrt{\varepsilon} \int_0^t f''(\eta_s^\varepsilon) \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s.
\end{aligned}
\tag{2.80}
$$

Now apply Ito's formula to $f(\hat{\eta}_t^\varepsilon)$ for $0 \le t \le \bar{\tau}$:

$$
\begin{aligned}
f(\hat{\eta}_t^\varepsilon) &= f(q) + \int_0^t f'(\hat{\eta}_s^\varepsilon) \eta_s^\varepsilon ds + \frac{\varepsilon}{2} \int_0^t f''(\hat{\eta}_s^\varepsilon) |\nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds \\
&\quad + \varepsilon \int_0^t f'(\hat{\eta}_s^\varepsilon) \nabla_x u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) \cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) ds + \sqrt{\varepsilon} \int_0^t f'(\hat{\eta}_s^\varepsilon) \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s \\
&= f(q) + \int_0^t f'(\hat{\eta}_s^\varepsilon) \hat{\eta}_s^\varepsilon ds + O(\sqrt{\varepsilon}) \cdot t + \frac{\varepsilon}{2} \int_0^t f''(\hat{\eta}_s^\varepsilon) \boldsymbol{A}_{22}(\boldsymbol{X}_s^\varepsilon) ds \\
&\quad + \frac{\varepsilon}{2} \int_0^t (|\nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 - \boldsymbol{A}_{22}(\boldsymbol{X}_s^\varepsilon)) f''(\eta_s^\varepsilon) ds + O(\sqrt{\varepsilon}) \cdot t \\
&\quad + \sqrt{\varepsilon} \int_0^t f'(\hat{\eta}_s^\varepsilon) \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_t^\varepsilon) dW_s \\
&= f(q) + \int_0^t [f'(\hat{\eta}_s^\varepsilon) \hat{\eta}_s^\varepsilon + \frac{A_0 \varepsilon}{2} f''(\hat{\eta}_s^\varepsilon)] ds + \frac{\varepsilon}{2} \int_0^t f''(\hat{\eta}_s^\varepsilon) (\boldsymbol{A}_{22}(\boldsymbol{X}_s^\varepsilon) - A_0) ds \\
&\quad + \frac{\varepsilon}{2} (O(1) + O(\frac{1}{\sqrt{\varepsilon}}) \cdot t - \sqrt{\varepsilon} \int_0^t f''(\eta_s^\varepsilon) \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s) + O(\sqrt{\varepsilon}) \cdot t \\
&\quad + \sqrt{\varepsilon} \int_0^t f'(\hat{\eta}_s^\varepsilon) \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s \\
&\le f(q) - t + \frac{t}{2} + O(\varepsilon) + O(\sqrt{\varepsilon}) \cdot t - \frac{\sqrt{\varepsilon^3}}{2} \int_0^t f''(\eta_s^\varepsilon) \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s \\
&\quad + \sqrt{\varepsilon} \int_0^t f'(\hat{\eta}_s^\varepsilon) \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s,
\end{aligned}
$$

where the equalities follow from (2.78) and (2.80), and the last inequality holds since f solves

(2.79) and $|\boldsymbol{A}_{22}(\boldsymbol{X}_s^\varepsilon) - A_0| < A_0/(2C)$. Let $\tilde{\tau} = \bar{\tau} \wedge 1/\varepsilon$. Then $\hat{\eta}_{\tilde{\tau}}^\varepsilon \in [-\bar{r}_2, \bar{r}_2]$ because $|\hat{\eta}_{\tilde{\tau}}^\varepsilon - \eta_{\tilde{\tau}}^\varepsilon| = O(\varepsilon)$. The previous calculation reduces to

$$f(\hat{\eta}_{\tilde{\tau}}^\varepsilon) \leq f(q) - \tilde{\tau}/2 + O(\varepsilon) + O(\sqrt{\varepsilon}) \cdot \tilde{\tau} - \frac{\sqrt{\varepsilon^3}}{2} \int_0^{\tilde{\tau}} f''(\eta_s^\varepsilon) \nabla_y v_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s$$
$$+ \sqrt{\varepsilon} \int_0^{\tilde{\tau}} f'(\hat{\eta}_s^\varepsilon) \nabla_y u_2(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s.$$

By taking the expectation, we have for all $x \in D_{r_2}$, $y \in \mathbb{T}^m$, and ε small enough

$$\mathbf{E}_{(x,y)} \tilde{\tau} \leq 5 \sup_{-\bar{r}_2 \leq q' \leq \bar{r}_2} f(q') = O(|\log \varepsilon|).$$

Then, by Markov's inequality and the Markov property, we obtain that $\mathbf{E}_{(x,y)} \bar{\tau} = O(|\log \varepsilon|)$. □

Lemma 2.7.5. *Let $r_1, r_2, \bar{\tau}$ be defined as in Lemma 2.7.4. Then, uniformly in $(x,y) \in D_{r_1} \times \mathbb{T}^m$,*

$$\mathbf{P}_{(x,y)}(\boldsymbol{X}_{\bar{\tau}}^\varepsilon \in (\partial D_{r_2})_{\mathrm{in}}) \to 0 \text{ as } \varepsilon \downarrow 0.$$

Proof. Again, we denote $x = (p,q)$ and, for simplicity, we assume that the saddle point is the origin and that $H(x) = pq$ in a small neighborhood of O. We extend the function $b(x,y)$ in the vertical direction in such a way that it is bounded together with its partial derivatives and the first component of $\bar{b}(x)$ is $-p$ in the region $\{x : |p| \leq r_2\}$. We denote $\zeta_t^\varepsilon = (\boldsymbol{X}_t^\varepsilon)_1$ and show that it takes significantly longer than $|\log \varepsilon|$ for ζ_t^ε to reach $\pm r_2$, hence it is unlikely for $\boldsymbol{X}_t^\varepsilon$ to exit D_{r_2} through $(\partial D_{r_2})_{\mathrm{in}}$. All the computations below concern $\boldsymbol{X}_t^\varepsilon$ before ζ_t^ε reaches $\pm r_2$. As in (2.19), we can write the equation for ζ_t^ε:

$$\zeta_t^\varepsilon = p - \int_0^t \zeta_s^\varepsilon ds + \sqrt{\varepsilon}\int_0^t \nabla_y u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s$$
$$+ \varepsilon\int_0^t \nabla_x u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds + \varepsilon(u_1(x,y) - u_1(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)).$$

Introduce $\hat{\zeta}_t^\varepsilon$, which is close to ζ_t^ε:

$$\hat{\zeta}_t^\varepsilon = p - \int_0^t \zeta_s^\varepsilon ds + \sqrt{\varepsilon}\int_0^t \nabla_y u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s + \varepsilon\int_0^t \nabla_x u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds.$$

Since $b(x,y)$ and its partial derivatives are bounded in $\{x : |p| \le r_2\} \times \mathbb{T}^m$, we can choose $C > 0$ such that

$$\sup_{x:|p|\le r_2, y\in\mathbb{T}^m} \left(|\nabla_y u_1(x,y)^\mathsf{T}\sigma(y)|^2 \vee 2|u_1(x,y)| \vee |\nabla_x u_1(x,y)\cdot b(x,y)|\right) < C/2. \tag{2.81}$$

Let us define $\bar{r}_2 = \frac{r_1+r_2}{2}$, $\hat{\tau}_2 = \inf\{t : |\hat{\zeta}_t^\varepsilon| > \bar{r}_2\}$, and function $f(p) = \exp(p^2/(C\varepsilon))$. Then it follows that

$$\frac{C\varepsilon}{2}f'' - pf' - f = 0. \tag{2.82}$$

Note that $|\zeta_t^\varepsilon| \le r_2$ for $0 \le t \le \hat{\tau}_2$ since $|\zeta_t^\varepsilon - \hat{\zeta}_t^\varepsilon| \le C\varepsilon/2$. Apply Ito's formula to $\exp(-t/2)f(\hat{\zeta}_t^\varepsilon)$ for $0 \le t \le \hat{\tau}_2$ and obtain using (2.81):

$$e^{-t/2}f(\hat{\zeta}_t^\varepsilon) = f(p) - \frac{1}{2}\int_0^t e^{-s/2}f(\hat{\zeta}_s^\varepsilon)ds - \int_0^t e^{-s/2}f'(\hat{\zeta}_s^\varepsilon)\zeta_s^\varepsilon ds$$
$$+ \varepsilon\int_0^t e^{-s/2}f'(\hat{\zeta}_s^\varepsilon)\nabla_x u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)ds$$
$$+ \frac{\varepsilon}{2}\int_0^t e^{-s/2}f''(\hat{\zeta}_s^\varepsilon)|\nabla_y u_1(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds$$

$$
\begin{aligned}
&+\sqrt{\varepsilon}\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\nabla_y u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s \\
&= f(p)-\frac{1}{2}\int_0^t e^{-s/2} f(\hat{\zeta}_s^\varepsilon)ds \\
&\quad+\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\left(-\frac{1}{2}\hat{\zeta}_s^\varepsilon+\left[(\hat{\zeta}_s^\varepsilon-\zeta_s^\varepsilon)+\varepsilon\nabla_x u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)-\frac{1}{2}\hat{\zeta}_s^\varepsilon\right]\right)ds \\
&\quad+\frac{\varepsilon}{2}\int_0^t e^{-s/2} f''(\hat{\zeta}_s^\varepsilon)|\nabla_y u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds \\
&\quad+\sqrt{\varepsilon}\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\nabla_y u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s \\
&\le f(p)-\frac{1}{2}\int_0^t e^{-s/2} f(\hat{\zeta}_s^\varepsilon)ds-\frac{1}{2}\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\hat{\zeta}_s^\varepsilon ds+\frac{C\varepsilon}{4}\int_0^t e^{-s/2} f''(\hat{\zeta}_s^\varepsilon)ds \\
&\quad+\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\left[(\hat{\zeta}_s^\varepsilon-\zeta_s^\varepsilon)+\varepsilon\nabla_x u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)\cdot b(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)-\frac{1}{2}\hat{\zeta}_s^\varepsilon\right]ds \\
&\quad+\sqrt{\varepsilon}\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\nabla_y u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s \\
&\le f(p)+18C\varepsilon(1-e^{-t/2})+\sqrt{\varepsilon}\int_0^t e^{-s/2} f'(\hat{\zeta}_s^\varepsilon)\nabla_y u_1(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)^\mathsf{T}\sigma(\boldsymbol{\xi}_s^\varepsilon)dW_s.
\end{aligned}
$$

The last inequality follows from (2.82) and the fact that the integrand on the second line is either negative, when $|\hat{\zeta}_s^\varepsilon|\ge 2C\varepsilon$, or small and bounded by $9C\varepsilon e^{-s/2}$, when $|\hat{\zeta}_s^\varepsilon|\le 2C\varepsilon$. By replacing t by the stopping time $\hat{\tau}_2$ and taking expectation, we obtain

$$
\mathbf{E}_{(x,y)}e^{-\hat{\tau}_2/2}\le 2e^{(r_1^2-\bar{r}_2^2)/(C\varepsilon)}.
$$

Let $\bar{\tau}_2=\inf\{t:|\zeta_t^\varepsilon|>r_2\}$. Then, since $|\zeta_t^\varepsilon-\hat{\zeta}_t^\varepsilon|\le C\varepsilon/2$, it follows that

$$
\mathbf{P}_{(x,y)}(\bar{\tau}_2<|\log\varepsilon|/\sqrt{\varepsilon})\le\mathbf{P}_{(x,y)}(\hat{\tau}_2<|\log\varepsilon|/\sqrt{\varepsilon})\le 2\exp(\frac{r_1^2-\bar{r}_2^2}{C\varepsilon}+\frac{|\log\varepsilon|}{2\sqrt{\varepsilon}})\to 0,
$$

as $\varepsilon \downarrow 0$. However, by Lemma 2.7.4 and Markov's inequality, we have

$$\mathbf{P}_{(x,y)}(\bar{\tau} > |\log \varepsilon|/\sqrt{\varepsilon}) \to 0,$$

as $\varepsilon \downarrow 0$. Thus, the desired result follows. $\square$

Lemma 2.7.6. *Let $\bar{\bar{\tau}} = \inf\{t : \boldsymbol{X}_t^\varepsilon \in D_{r_1}\} \wedge \boldsymbol{\tau}$. Then there exists $a > 0$ such that, uniformly in $(x, y) \in (\partial D_{r_2})_{\text{out}} \times \mathbb{T}^m$,*

$$\mathbf{P}_{(x,y)}(\bar{\bar{\tau}} < \boldsymbol{\tau}, \int_0^{\bar{\bar{\tau}}} |\nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds < a) \to 0 \text{ as } \varepsilon \to 0.$$

Furthermore, $\mathbf{E}_{(x,y)}\bar{\bar{\tau}}$ is bounded uniformly in $(x, y) \in (\partial D_{r_2})_{\text{out}} \times \mathbb{T}^m$.

Proof. Let $\hat{t} > 0$ and $\check{t} > 0$ be the lower bound and the upper bound of time spent by $\boldsymbol{x}_t$ to get from $(\partial D_{r_2})_{\text{out}}$ to D_{r_1}, respectively. Then, similarly to (2.48), there exists $a > 0$ such that

$$\mathbf{P}_{(x,y)}\left(\int_0^{\hat{t}/2} |\nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon))|^2 ds > a, \sup_{0 \le t \le 2\hat{t}} |\boldsymbol{X}_t^\varepsilon - \boldsymbol{x}_t| \le \varepsilon^{\frac{1+2\alpha}{4}}\right) \to 1.$$

Hence

$$\begin{aligned}
&\mathbf{P}_{(x,y)}\left(\bar{\bar{\tau}} < \boldsymbol{\tau}, \int_0^{\bar{\bar{\tau}}} |\nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds < a\right) \\
&\le \mathbf{P}_{(x,y)}\left(\hat{t}/2 \le \bar{\bar{\tau}} < \boldsymbol{\tau}, \int_0^{\bar{\bar{\tau}}} |\nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds < a\right) + \mathbf{P}_{(x,y)}(\bar{\bar{\tau}} < \boldsymbol{\tau}, \bar{\bar{\tau}} < \hat{t}/2) \\
&\le \mathbf{P}_{(x,y)}\left(\int_0^{\hat{t}/2} |\nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon)|^2 ds < a\right) + \mathbf{P}_{(x,y)}\left(\sup_{0 \le t \le 2\hat{t}} |\boldsymbol{X}_t^\varepsilon - \boldsymbol{x}_t| > \varepsilon^{\frac{1+2\alpha}{4}}\right) \\
&\to 0.
\end{aligned}$$

Similarly, it is easy to see that $\mathbf{P}_{(x,y)}(\bar{\bar{\tau}} > 2\check{t}) < \mathbf{P}_{(x,y)}\left(\sup_{0\le t\le 2\check{t}} |\boldsymbol{X}_t^\varepsilon - \boldsymbol{x}_t| > \varepsilon^{\frac{1+2\alpha}{4}}\right) \to 0$, and the desired result follows from the Markov property. $\square$

Proof of Proposition 2.7.3. As in (2.21):

$$
\begin{aligned}
H(\boldsymbol{X}_t^\varepsilon) = H(x) &+ \varepsilon \int_0^t \nabla_x u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) \cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) ds \\
&+ \sqrt{\varepsilon} \int_0^t \nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s + \varepsilon(u_h(x,y) - u_h(\boldsymbol{X}_t^\varepsilon, \boldsymbol{\xi}_t^\varepsilon)).
\end{aligned}
$$

The change in $H(\boldsymbol{X}_t^\varepsilon)$ is mainly due to the stochastic integral while the other terms are of order $O(\varepsilon)$ and $O(t\cdot\varepsilon)$ and can be controlled. For each $t(\varepsilon) > 0$,

$$
\begin{aligned}
\{\boldsymbol{\tau} < t(\varepsilon)\} \supset &\left\{ \sup_{[0,t(\varepsilon)]} \left| \sqrt{\varepsilon} \int_0^{t(\varepsilon)} \nabla_y u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)^\mathsf{T} \sigma(\boldsymbol{\xi}_s^\varepsilon) dW_s \right| > 3\varepsilon^\alpha \right\} \\
&\bigcap \left\{ \varepsilon \int_0^{t(\varepsilon)} |\nabla_x u_h(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon) \cdot b(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)| ds < \varepsilon^\alpha/2 \right\}.
\end{aligned} \tag{2.83}
$$

Note that if we choose $t(\varepsilon) = o(\varepsilon^{\alpha-1})$, then the second event is always true. Now we recursively define stopping times:

$$
\begin{aligned}
&\theta_0^1 = 0, \\
&\theta_k^2 = \inf\{t \ge \theta_{k-1}^1 : \boldsymbol{X}_t^\varepsilon \in \partial D_{r_2}\} \wedge \boldsymbol{\tau}, \\
&\theta_k^1 = \inf\{t \ge \theta_k^2 : \boldsymbol{X}_t^\varepsilon \in \partial D_{r_1}\} \wedge \boldsymbol{\tau}.
\end{aligned}
$$

We denote $D(x,y) = \nabla_y u_h(x,y)^\mathsf{T}\sigma(y)$. Note that once the process leaves V^ε, the stopping times stay constant afterwards. The main idea of the proof is to show that after a sufficiently long time $t(\varepsilon)$, the stochastic integral will accumulate enough variance to exit from V^ε. Let us bound the probability of variance being small:

$$\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau} \geq t(\varepsilon), \int_0^{t(\varepsilon)} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds < 9\varepsilon^{2\alpha-1}\right)$$
$$\leq \mathbf{P}_{(x,y)}\left(\boldsymbol{\tau} \geq t(\varepsilon) > \theta^1_{n(\varepsilon)}, \int_0^{t(\varepsilon)} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds < 9\varepsilon^{2\alpha-1}\right) + \mathbf{P}_{(x,y)}(\theta^1_{n(\varepsilon)} \geq t(\varepsilon)), \tag{2.84}$$

where the integer $n(\varepsilon)$ will be specified later. Let $\bar{\tau}$, $\bar{\bar{\tau}}$, and a be defined as in Lemma 2.7.4 and Lemma 2.7.6. Then

$$\begin{aligned}
&\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau} \geq t(\varepsilon) > \theta^1_{n(\varepsilon)}, \int_0^{t(\varepsilon)} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds < 9\varepsilon^{2\alpha-1}\right) \\
&\quad \leq \exp(9\varepsilon^{2\alpha-1}/a)\mathbf{E}_{(x,y)}\left(\chi_{\{\boldsymbol{\tau} > \theta^1_{n(\varepsilon)}\}} \exp\left(-\frac{1}{a}\int_0^{\theta^1_{n(\varepsilon)}} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds\right)\right) \\
&\quad \leq \exp(9\varepsilon^{2\alpha-1}/a)\left[\sup_{(x,y)\in\partial D_{r_1}\times\mathbb{T}^m} \mathbf{E}_{(x,y)}\left(\chi_{\{\boldsymbol{\tau} > \theta^1_1\}} \exp\left(-\frac{1}{a}\int_0^{\theta^1_1} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds\right)\right)\right]^{n(\varepsilon)-1}.
\end{aligned} \tag{2.85}$$

Now let us deal with one excursion from D_{r_2} to D_{r_1}. For $(x,y) \in (\partial D_{r_2})_{\text{out}} \times \mathbb{T}^m$,

$$\begin{aligned}
&\mathbf{E}_{(x,y)}\left(\chi_{\{\bar{\bar{\tau}} < \boldsymbol{\tau}\}} \exp\left(-\frac{1}{a}\int_0^{\bar{\bar{\tau}}} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds\right)\right) \\
&\quad \leq \mathbf{P}_{(x,y)}(\bar{\bar{\tau}} < \boldsymbol{\tau}, \int_0^{\bar{\bar{\tau}}} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds < a) + \mathbf{P}_{(x,y)}(\bar{\bar{\tau}} < \boldsymbol{\tau}, \int_0^{\bar{\bar{\tau}}} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds \geq a)/e \\
&\quad \leq e^{-0.99},
\end{aligned} \tag{2.86}$$

for all ε sufficiently small, by Lemma 2.7.6. For $(x,y) \in \partial D_{r_1} \times \mathbb{T}^m$:

$$\begin{aligned}
&\mathbf{E}_{(x,y)}\left(\chi_{\{\boldsymbol{\tau} > \theta^1_1\}} \exp\left(-\frac{1}{a}\int_0^{\theta^1_1} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds\right)\right) \\
&\leq \mathbf{E}_{(x,y)}\left(\chi_{\{\boldsymbol{\tau} > \theta^1_1, \boldsymbol{X}^\varepsilon_{\bar{\tau}} \in (\partial D_{r_2})_{\text{out}}\}} \exp\left(-\frac{1}{a}\int_0^{\theta^1_1} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds\right)\right) + \mathbf{P}_{(x,y)}(\boldsymbol{X}^\varepsilon_{\bar{\tau}} \in (\partial D_{r_2})_{\text{in}})
\end{aligned}$$

$$
\leq \sup_{(x',y')\in(\partial D_{r_2})_{\text{out}}\times\mathbb{T}^m} \mathbf{E}_{(x',y')}\left(\chi_{\{\bar{\bar{\tau}}<\boldsymbol{\tau}\}}\exp\left(-\frac{1}{a}\int_0^{\bar{\bar{\tau}}}|D(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)|^2ds\right)\right)+\mathbf{P}_{(x,y)}(\boldsymbol{X}_{\bar{\tau}}^\varepsilon\in(\partial D_{r_2})_{\text{in}})
$$

$$
\leq e^{-0.98},
$$

by Lemma 2.7.5 and (2.86). Now we can come back to (2.85) and have

$$
\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau}\geq t(\varepsilon)>\theta_n^1,\int_0^{t(\varepsilon)}|D(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)|^2ds<9\varepsilon^{2\alpha-1}\right)\leq\exp(9\varepsilon^{2\alpha-1}/a-0.98(n(\varepsilon)-1)). \tag{2.87}
$$

The second probability on the right-hand side of (2.84) can be bounded by Lemmas 2.7.4 and 2.7.6 with certain $K>0$:

$$
\begin{aligned}
\mathbf{P}_{(x,y)}(\theta_n^1\geq t(\varepsilon)) &\leq \mathbf{E}_{(x,y)}\theta_n^1/t(\varepsilon)\\
&\leq\left(\sup_{(x',y')\in\partial D_{r_1}\times\mathbb{T}^m}\mathbf{E}_{(x',y')}\bar{\tau}+\sup_{(x',y')\in\partial D_{r_2}\times\mathbb{T}^m}\mathbf{E}_{(x',y')}\bar{\bar{\tau}}\right)\cdot\frac{n(\varepsilon)}{t(\varepsilon)}\\
&\leq\frac{n(\varepsilon)K|\log\varepsilon|}{t(\varepsilon)}.
\end{aligned} \tag{2.88}
$$

Choose $n(\varepsilon)=[10\varepsilon^{2\alpha-1}/a+2]$. Then the quantity in (2.87) converges to 0. Choose $t(\varepsilon)=100K\varepsilon^{2\alpha-1}|\log\varepsilon|/a$, then the quantity on the right-hand side of (2.88) converges to 0.1. Therefore, the quantity on the right-hand side of (2.84) converges to 0.1. Moreover, since $t(\varepsilon)=o(\varepsilon^{\alpha-1})$, it follows from (2.83) that, for all $x\in V^\varepsilon$, $y\in\mathbb{T}^m$, and ε sufficiently small,

$$
\begin{aligned}
&\mathbf{P}_{(x,y)}(\boldsymbol{\tau}\geq t(\varepsilon))\\
&=\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau}\geq t(\varepsilon),\sup_{[0,t(\varepsilon)]}\left|\sqrt{\varepsilon}\int_0^t D(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)dW_s\right|\leq 3\varepsilon^\alpha\right)\\
&\leq\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau}\geq t(\varepsilon),\int_0^{t(\varepsilon)}|D(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)|^2ds>9\varepsilon^{2\alpha-1},\sup_{[0,t(\varepsilon)]}\left|\sqrt{\varepsilon}\int_0^t D(\boldsymbol{X}_s^\varepsilon,\boldsymbol{\xi}_s^\varepsilon)dW_s\right|\leq 3\varepsilon^\alpha\right)
\end{aligned}
$$

$$+\mathbf{P}_{(x,y)}\left(\boldsymbol{\tau} \geq t(\varepsilon), \int_0^{t(\varepsilon)} |D(\boldsymbol{X}_s^\varepsilon, \boldsymbol{\xi}_s^\varepsilon)|^2 ds < 9\varepsilon^{2\alpha-1}\right)$$
$$\leq 0.69 + 0.11 = 0.8,$$

since the stochastic integral in the last inequality can be represented as time-changed Brownian motion. Finally, we have by the Markov property

$$\mathbf{E}_{(x,y)}\boldsymbol{\tau} \leq 5t(\varepsilon) = O(\varepsilon^{2\alpha-1}|\log \varepsilon|). \qquad \square$$

2.7.3 Tightness

In this section, we verify the tightness for the original process (2.4) on $\mathbb{R}^2 \times \mathbb{T}^m$. By the Arzelà–Ascoli theorem, it suffices to check the following two conditions

$$(i) \lim_{R\to+\infty} \mathbf{P}_{(x,y)}\left(\sup_{0\leq t\leq T} |H(X_t^\varepsilon)| \geq R\right) \to 0, \tag{2.89}$$

$$(ii) \lim_{\delta\downarrow 0} \mathbf{P}_{(x,y)}\left(\sup_{\substack{0\leq s<t\leq T\\ |s-t|<\delta}} r(h(X_t^\varepsilon), h(X_s^\varepsilon)) > \kappa\right) \to 0, \tag{2.90}$$

hold for all $\kappa > 0$, uniformly in all $0 < \varepsilon < 1$. As in (2.21), we can also write the equation for $H(X_t^\varepsilon)$, where we consider $(X_t^\varepsilon, \xi_t^\varepsilon)$ to be a process on $\mathbb{R}^2 \times \mathbb{T}^m$:

$$\begin{aligned} H(X_t^\varepsilon) = H(x) &+ \int_0^t \nabla_y u_h(X_s^\varepsilon, \xi_s^\varepsilon)^\mathsf{T} \sigma(\xi_s^\varepsilon) dW_s + \varepsilon(u_h(x_0, y_0) - u_h(X_t^\varepsilon, \xi_t^\varepsilon)) \\ &+ \int_0^t \nabla_x u_h(X_s^\varepsilon, \xi_s^\varepsilon) \cdot b(X_s^\varepsilon, \xi_s^\varepsilon) ds. \end{aligned} \tag{2.91}$$

By assumption *(H4)* and Lemma 2.3.2, $u(x,y)$ is bounded together with its first derivatives. Besides, by assumption *(H2)*, $H(x)/|x| \to +\infty$ as $|x| \to +\infty$, hence $\sup\{|x| : H(x) \le R\}/R \to 0$ as $R \to +\infty$. Also, there exists an $K > 0$ such that and $H(x) > -K$ for all $x \in \mathbb{R}^2$. For $R > K$, let $\tau_R = \inf\{t : |H(X_t^\varepsilon)| = R\} \wedge T$. Then, by Markov's inequality and boundedness of second derivatives of H,

$$
\begin{aligned}
\mathbf{P}_{(x,y)}(\sup_{0\le t\le T} |H(X_t^\varepsilon)| \ge R) &= \mathbf{P}_{(x,y)}(H(X_{\tau_R}^\varepsilon) = R) \\
&\le \mathbf{E}_{(x,y)}(H(X_{\tau_R}^\varepsilon) + K)/R \\
&\lesssim (H(x) + (\varepsilon + T)\sup\{|\nabla H(x)| : H(x) \le R\} + T + K)/R \\
&\lesssim (H(x) + (\varepsilon + T)\sup\{|x| : H(x) \le R\} + T + K)/R \to 0,
\end{aligned}
$$

as $R \to +\infty$, uniformly in $0 < \varepsilon < 1$. Thus, we have (2.89), and it also follows that

$$
\mathbf{P}_{(x,y)}(\sup_{0\le t\le T} |\nabla H(X_t^\varepsilon)| \ge R) \to 0, \tag{2.92}
$$

as $R \to +\infty$, uniformly in $0 < \varepsilon < 1$. To verify (2.90), we see that, for an arbitrary $\kappa > 0$ small,

$$
\begin{aligned}
\mathbf{P}_{(x,y)}\left(\sup_{\substack{0\le s<t\le T\\ |s-t|<\delta}} r(h(X_t^\varepsilon), h(X_s^\varepsilon)) > \kappa\right) &\le \sum_{k=0}^{[T/\delta]} \mathbf{P}_{(x,y)}\left(\sup_{t\le\delta} r(h(X_{k\delta+t}^\varepsilon), h(X_{k\delta}^\varepsilon)) > \kappa/4\right) \\
&\le \sum_{k=0}^{[T/\delta]} \mathbf{P}_{(x,y)}\left(\sup_{t\le\delta} |H(X_{k\delta+t}^\varepsilon) - H(X_{k\delta}^\varepsilon)| > \kappa/12\right).
\end{aligned}
$$

Since (2.92) holds, it is sufficient to prove, for each $R > 0$,

$$\sum_{k=0}^{[T/\delta]} \mathbf{P}_{(x,y)} \left(\sup_{t\leq\delta} |H(X^\varepsilon_{k\delta+t}) - H(X^\varepsilon_{k\delta})| > \kappa/12, \sup_{0\leq t\leq T} |\nabla H(X^\varepsilon_t)| \leq R \right) \to 0.$$

Let K' be the upper bound for $|b(x,y) - \nabla^\perp H(x)|$, $|u(x,y)|$, $|\nabla_x u(x,y)|$, and eigenvalues of $\nabla^2 H(x)$ and $(\nabla_y u\sigma\sigma^\mathsf{T}\nabla_y u^\mathsf{T})(x,y)$ over $\mathbb{R}^2 \times \mathbb{T}^m$. Here we need to deal with the cases where ε is small or large compared to κ separately. Namely, on the event $\{\sup_{0\leq t\leq T} |\nabla H(X^\varepsilon_t)| \leq R\}$:

(1) If $\varepsilon > \frac{\kappa}{48K'R}$, we have that, by (2.4), for $\delta < (\frac{\kappa}{24K'R})^2$ and all $0 \leq t \leq \delta$,

$$|H(X^\varepsilon_{k\delta+t}) - H(X^\varepsilon_{k\delta})| = |\frac{1}{\varepsilon}\int_{k\delta}^{k\delta+t} \nabla H(X^\varepsilon_s)\cdot b(X^\varepsilon_s, \xi^\varepsilon_s)ds| \leq \frac{\delta RK'}{\varepsilon} < \kappa/12,$$

and the probabilities are simply zero;

(2) If $\varepsilon < \frac{\kappa}{48K'R}$, by recalling (2.91), we have

$$\begin{aligned}
&\sup_{t\leq\delta} |H(X^\varepsilon_{k\delta+t}) - H(X^\varepsilon_{k\delta})| \\
&\leq \sup_{t\leq\delta} |\int_{k\delta}^{k\delta+t} \nabla_y u_h(X^\varepsilon_s, \xi^\varepsilon_s)^\mathsf{T}\sigma(\xi^\varepsilon_s)dW_s| + 2\varepsilon K'R + {K'}^2(K'+R)\delta \\
&\leq \sup_{0\leq t\leq\delta K'R^2} |\tilde{W}_t| + {K'}^2(K'+R)\delta + \kappa/24,
\end{aligned}$$

where $\tilde{W}_t$ is another Brownian motion. Hence, for δ sufficiently small, independently of ε,

$$\begin{aligned}
&\sum_{k=0}^{[T/\delta]} \mathbf{P}_{(x,y)} \left(\sup_{t\leq\delta} |H(X^\varepsilon_{k\delta+t}) - H(X^\varepsilon_{k\delta})| > \kappa/12, \sup_{0\leq t\leq T} |\nabla H(X^\varepsilon_t)| \leq R \right) \\
&\leq \left[\frac{T}{\delta} + 1\right] \cdot \mathbf{P}_{(x,y)} \left(\sup_{0\leq t\leq\delta K'R^2} |\tilde{W}_t| > \kappa/48 \right)
\end{aligned}$$

$$\to 0,$$

as $\delta \downarrow 0$, since each term is exponentially small as $\delta \downarrow 0$.

Chapter 3: Local limit theorem for time-inhomogeneous functions of Markov processes

3.1 Introduction

Suppose that $\{X_s^x\}_{s\geq 0}$ is a time-homogeneous Markov family on a metric space $(\boldsymbol{X}, \mathscr{B}(\boldsymbol{X}))$ with $P(s,x,\Gamma)$ being its transition function. We will drop the superscript x from the notation when it is clear what the initial point or distribution of the process is. Let $\mathcal{B}$ be the Banach space of bounded $\mathscr{B}(\boldsymbol{X})$-measurable complex functions on $\boldsymbol{X}$ equipped with the supremum norm: $\|f\| = \sup_{x\in\boldsymbol{X}} |f(x)|$. Then we have the corresponding semigroup $\{P_s\}_{s\geq 0}$ defined as

$$P_s f(x) = \mathbf{E}_x f(X_s) = \int_{\boldsymbol{X}} f(y) P(s,x,dy).$$

In order to state the local limit theorem, we make additional assumptions about $\{X_s\}_{s\geq 0}$. Among all, the most essential condition is the uniqueness of the invariant measure and the uniform convergence of X_s to this measure, which can be re-formulated in terms of the quasi-compactness ([33]) of the operators P_s, $s > 0$. To be more precise, an operator P acting on $\mathcal{B}$ is said to satisfy the property of quasi-compactness if $\mathcal{B}$ can be decomposed into two P-invariant closed subspaces:

$$\mathcal{B} = F \oplus H, \tag{3.1}$$

where $r(P|_H) < r(P)$, while $\dim(F) < \infty$, and each eigenvalue of $P_{|F}$ has modulus $r(P)$. Here, $r(\cdot)$ stands for the spectral radius of an operator. Suppose that ν is the unique invariant measure of $\{X_s\}_{s\geq 0}$, and the distribution of X_s converges to ν in total variation, uniformly in initial distribution. Then it is clear that, for every $s > 0$, P_s is quasi-compact with $F = \mathrm{span}\{1\}$ and $H = \{f \in \mathcal{B} : \langle \nu, f\rangle = 0\}$, where ν is understood as a functional on $\mathcal{B}$ via $\langle \nu, f\rangle = \int_{\boldsymbol{X}} f d\nu$. A more direct way to understand quasi-compactness is to give a decomposition of the operator itself. Define $\mathcal{L}_{\mathcal{B}}$ to be the space of all bounded linear operators from $\mathcal{B}$ to itself, and $\mathcal{B}'$ to be the space of all bounded linear operators from $\mathcal{B}$ to $\mathbb{C}$. For $\varphi \in \mathcal{B}'$ and $v \in \mathcal{B}$, we define $v \otimes \varphi \in \mathcal{L}_{\mathcal{B}}$ by $(v \otimes \varphi)f = \langle \varphi, f\rangle v$ for all $f \in \mathcal{B}$. Let us denote $Q = P_1$ be the operator corresponding to the Markov process $\{X_s\}_{s\geq 0}$ sampled at integer points. As we discussed, Q is quasi-compact with $F = \mathrm{span}\{1\}$ and $H = \{f \in \mathcal{B} : \langle \nu, f\rangle = 0\}$. So, we have the following decomposition for Q:

$$Q = 1 \otimes \nu + N, \tag{3.2}$$

where $N = Q\Pi_H$, Π_H is the projection onto H, and $r(N) < 1$.

In applications, one often considers limit theorems for additive functionals of Markov processes ([33],[34],[35]). For example, let $b(x)$ be a real-valued bounded continuous function on $\boldsymbol{X}$. One is interested in describing the distribution of $\int_0^T b(X_s)ds$ as $T \to \infty$. So, in addition to the process $\{X_s\}_{s\geq 0}$, assumptions also need to be made about the function $b(x)$. Similar questions have also been studied for discrete-time Markov chains, in which case, one is interested in the distribution of the sum $\sum_{k=1}^n b(X_k)$ as $n \to \infty$.

In this chapter, we focus on the integral formulation of the problem as described but allow the function b to vary in time. Namely, let $b : [0,1] \times \boldsymbol{X} \to \mathbb{R}^d$, $d \in \mathbb{N}$, be a real-valued bounded continuous function and $S_T = \int_0^T b(s/T, X_s)ds$. We will prove that a local limit theorem holds for S_T. The spectral method has been applied to prove various limit theorems for Markov chains (cf. [33], [36], [37], [35], [38], and the references therein). For example, in the case where the function b does not depend on time, the local limit theorem can be obtained by studying the characteristic function of the sum (or the integral in the continuous case) or, equivalently, the limiting behavior of the power of the corresponding Fourier kernels. Due to the quasi-compactness of the corresponding operator, the limiting behavior of its power is primarily determined by the dominant eigenvalues and the corresponding eigenfunctions. However, in this chapter, we need to deal with the additional dependence on another parameter that appears in the discretization of the time interval $[0, T]$ due to the time-inhomogeneity of the function b, and study the product of a sequence of different Fourier operators that depend on the parameter. This main technical step is given in Proposition 3.3.1.

The chapter is organized as follows: In the rest of Section 3.1, we introduce some notation, formulate the assumptions and the main result, and discuss some applications. In Section 3.2, we give preliminary results to motivate the ingredients of the proof. In Section 3.3, we prove the technical results. In Section 3.4, we give the proof of our main result, Theorem 3.1.4.

Throughout this chapter, $\mathbf{P}$ and $\mathbf{E}$ represent the probability and expectation, respectively, and the subscripts pertain to initial conditions. The Lebesgue measure is denoted by $\mathfrak{L}$. For each $k \geq 0$, a subset X of a Euclidean space, and a Banach space Y, $\mathcal{C}^k(X, Y)$ denotes the space of k times continuously differentiable functions from X to Y. For $C > 0$, let $\mathcal{B}_C = \{f \in \mathcal{B} : \|f\| \leq C\}$. Let $\mathcal{B}'_p$ be the space of probability measures on $\boldsymbol{X}$. Let $\lfloor T \rfloor$ denote the integer part of T.

Definition 3.1.1. *The operator* $Q(t,\alpha,\beta) \in \mathcal{L}_{\mathcal{B}}$*,* $t \in \mathbb{R}^d$*,* $\alpha, \beta \in [0,1]$*, is defined as*

$$Q(t,\alpha,\beta)f(x) = \mathbf{E}_x\left[\exp\left(it \cdot \int_0^1 b\left(\alpha + s(\beta-\alpha), X_s\right) ds\right) f(X_1)\right]. \tag{3.3}$$

In particular, for each α *and* β*, and* $t = 0$*, this operator coincides with* Q*.*

Assumptions:

(A1) $\{X_s\}_{s\geq 0}$ is a Markov process on the complete separable state space $(\boldsymbol{X}, \mathscr{B}(\boldsymbol{X}))$ with sample paths in the Skorokhod space $D([0,\infty), \boldsymbol{X})$.

(A2) ν is the unique invariant measure of $\{X_s\}_{s\geq 0}$ on $\boldsymbol{X}$. There exists $\mathcal{T} > 0$ such that

$$\sup_{x\in\boldsymbol{X}} \mathrm{TV}(P(\mathcal{T}, x, \cdot), \nu) < 1,$$

where TV is the total variation distance between two probability measures.

(A3) $b : [0,1] \times \boldsymbol{X} \to \mathbb{R}^d$ is continuous and bounded. For each $0 \leq \alpha \leq 1$, $\int_{\boldsymbol{X}} b(\alpha,\cdot)d\nu = 0$. For each $x \in \boldsymbol{X}$, $b(\alpha,x)$ is twice differentiable in α. Moreover, $\frac{\partial}{\partial\alpha}b(\alpha,x)$ and $\frac{\partial^2}{\partial\alpha^2}b(\alpha,x)$ are continuous and bounded on $[0,1]\times\boldsymbol{X}$.

(A4) Non-arithmetic condition on $(\{X_s\}_{s\geq 0}, b(\alpha,x))$ holds: For every $0 \leq \alpha \leq 1$ and every $t \neq 0$, $r(Q(t,\alpha,\alpha)) < 1$.

Remark 3.1.2. *It is not hard to verify that, under Assumption (A2),* Q *is quasi-compact as defined in* (3.1) *and has the decomposition* (3.2)*. In addition, under Assumption (A3),* $Q(t,\alpha,\beta) \in \mathcal{C}^2(\mathbb{R}^d \times [0,1]^2, \mathcal{L}_{\mathcal{B}})$.

Remark 3.1.3. *The non-arithmetic condition (A4) is required to prove the local limit theorem. However, the formulation is not intuitive and often not easy to verify. In fact, with Assumptions (A1)-(A3), it can be reduced to a more natural form - the non-lattice condition:*

(NL) *For each $\alpha \in [0,1]$, there is no $a \in \mathbb{R}^d$, a closed subgroup $H \subsetneq \mathbb{R}^d$, and bounded measurable function $u : \boldsymbol{X} \to \mathbb{R}^d$ such that $\int_0^1 b(\alpha, X_s)ds + u(X_1) - u(X_0) = a + H$, ν-a.s.*

In Section 3.5, we will provide a proof of this statement.

Theorem 3.1.4. *Let Assumptions (A1)-(A4) be satisfied and*

$$S_T = \int_0^T b(s/T, X_s)ds. \tag{3.4}$$

Then there exists a positive-definite matrix Σ such that, for each compactly supported function $g \in \mathcal{C}(\mathbb{R}^d, \mathbb{R})$, and $C > 0$, uniformly in real-valued $f \in \mathcal{B}_C$, initial distribution $\mu \in \mathcal{B}'_p$, and $u \in \mathbb{R}^d$,

$$\lim_{T\to\infty} |\det(\Sigma)(2\pi T)^{d/2}\mathbf{E}_\mu[f(X_T)g(S_T - u)] - e^{-\frac{1}{2T}u^*(\Sigma\Sigma^*)^{-1}u}\langle \nu, f\rangle\langle \mathfrak{L}, g\rangle| = 0.$$

Example 3.1.5. *The assumptions are satisfied if X_s is a non-degenerate diffusion process on $\mathbb{T}^m$, (A3) is satisfied, and $\{b(\alpha, x),\, x \in \boldsymbol{X}\}$ spans $\mathbb{R}^d$ for each $0 \leq \alpha \leq 1$.*

3.2 Preliminaries

Throughout the rest of the chapter, we assume that the conditions (A1)-(A4) are satisfied. In order to understand the limiting distribution of S_T, we study the limiting behavior of the characteristic function, which can be represented as a product of the operators $Q(t,\alpha,\beta)$. Since T is not necessarily an integer, we define a special operator $\tilde{Q}(t,T)$ that will serve as the last factor in the product. Namely, for each $f \in \mathcal{B}$, we define

$$\tilde{Q}(t,T)f = \mathbf{E}_x\left[\exp\left(it\cdot\int_0^{T-\lfloor T\rfloor} b\left(\frac{\lfloor T\rfloor + s}{T}, X_s\right)ds\right) f\left(X_{T-\lfloor T\rfloor}\right)\right]. \tag{3.5}$$

From the definition (3.4) of S_T, we immediately get the following lemma.

Lemma 3.2.1. *For each $f \in \mathcal{B}$ and each probability measure μ,*

$$\mathbf{E}_\mu[e^{it\cdot S_T} f(X_T)] = \langle \mu, \prod_{k=0}^{\lfloor T\rfloor-1} Q\left(t, \frac{k}{T}, \frac{k+1}{T}\right)\tilde{Q}(t,T)f\rangle, \tag{3.6}$$

where the product of the operators $\prod_{k=1}^n A_k$ means $A_1A_2...A_n$.

The operator in (3.5) does not play a role in the limiting distribution and the introduction of it is purely for technical reasons. We will not touch it until in the proof of the main result. It will be proved, using Assumption (A4), that the value of (3.6) decays exponentially fast as $T \to \infty$ if t is away from the origin. To deal with the situation where t is close to the origin, we need to study the behavior of $Q(t,\alpha,\beta)$ near the origin. In fact, since $Q(t,\alpha,\beta)$ can be treated as a perturbation of $Q = Q(0,\alpha,\beta)$, the decomposition in the form of (3.2) and the quasi-compactness can be extended locally to all t close to the origin. This result is given explicitly in the following perturbation theorem. The theorem is similar to Theorem 3.8 in [33]. The proof is also similar except that now we need to deal with the dependence of the operator $Q(t,\alpha,\beta)$ on the parameters α and β due to the time-inhomogeneity of the function b, and we need to establish additional differentiability of various terms in the decomposition and prove the existence of a neighborhood I that would work for all the values of the parameters. We include the proof for the sake of completeness.

Theorem 3.2.2 (Perturbation Theorem)**.** *There exists a neighborhood I of the origin, $C > 0$, and $r \in (0,1)$ such that, for $t \in I$, $\alpha, \beta \in [0,1]$, $Q(t,\alpha,\beta)$ has the following decomposition:*

$$Q(t,\alpha,\beta) = \lambda(t,\alpha,\beta)v(t,\alpha,\beta) \otimes \varphi(t,\alpha,\beta) + N(t,\alpha,\beta), \tag{3.7}$$

where

(i) $\lambda(t,\alpha,\beta) \in \mathcal{C}^2(I \times [0,1]^2, \mathbb{C})$*;*

(ii) $v(t,\alpha,\beta) \in \mathcal{C}^2(I \times [0,1]^2, \mathcal{B})$ *and* $\|v(t,\alpha,\beta)\| = 1$*;*

(iii) $\varphi(t,\alpha,\beta) \in \mathcal{C}^2(I \times [0,1]^2, \mathcal{B}')$ *and* $\langle \varphi(t,\alpha,\beta), v(t,\alpha,\beta)\rangle = 1$*;*

(iv) $\mathcal{B} = F(t,\alpha,\beta) \oplus H(t,\alpha,\beta)$*,* $F(t,\alpha,\beta) = \operatorname{span}\{v(t,\alpha,\beta)\}$*, and* $H(t,\alpha,\beta) = \{h : \langle \varphi(t,\alpha,\beta), h\rangle = 0\}$*;*

(v) $N(t,\alpha,\beta) = Q(t,\alpha,\beta)\Pi_{H(t,\alpha,\beta)} \in \mathcal{C}^2(I \times [0,1]^2, \mathcal{L}_{\mathcal{B}})$ *and* $r(N(t,\alpha,\beta))/|\lambda(t,\alpha,\beta)| < r$ *for all* $t \in I$ *and* $\alpha, \beta \in [0,1]$*;*

(vi) $\|\nabla_t^2(N(t,\alpha,\beta)^n)\| < C$ *for all* $\alpha, \beta \in [0,1]$ *and* $n \in \mathbb{N}$.

Proof. (1) Let us temporarily fix $\alpha_0, \beta_0 \in [0,1]$. Recall that $H = \{f \in \mathcal{B} : \langle \nu, f\rangle = 0\}$. Note that for $h \in H$, $\alpha_0, \beta_0 \in [0,1]$,

$$\langle \nu, Q(0,\alpha_0,\beta_0)(1+h)\rangle = \langle \nu, Q(1+h)\rangle = \langle Q^*\nu, 1+h\rangle = \langle \nu, 1+h\rangle = 1.$$

There exists a neighborhood $J_1(\alpha_0,\beta_0)$ in $\mathbb{R}^d \times [0,1]^2$ of $(0,\alpha_0,\beta_0)$ such that for all $(t,\alpha,\beta) \in J_1(\alpha_0,\beta_0)$ and $h \in H$ of norm no more than 1, we have $\langle \nu, Q(t,\alpha,\beta)(1+h)\rangle \neq 0$. Then we can define a function $F : J_1(\alpha_0,\beta_0) \times \{h \in H : \|h\| \leq 1\} \to H$ by

$$F(t,\alpha,\beta,h) = \frac{Q(t,\alpha,\beta)(1+h)}{\langle \nu, Q(t,\alpha,\beta)(1+h)\rangle} - (1+h).$$

Note that $F(0,\alpha_0,\beta_0,0) = 0$. By the fact that

$$F(0,\alpha_0,\beta_0,h) = \frac{Q(1+h)}{\langle \nu, Q(1+h)\rangle} - (1+h) = Qh - h,$$

we see that the partial differential of F at the point $(0,\alpha_0,\beta_0,0)$ w.r.t. h is $Q|_H - 1$ which is an isomorphism of H since 1 is not in the spectrum of $Q|_H$ due to (3.2). Hence, by the implicit function theorem, there exists a neighborhood $J_2(\alpha_0,\beta_0) \subset J_1(\alpha_0,\beta_0)$ of $(0,\alpha_0,\beta_0)$, a neighborhood $H_0 \subset H$ of 0, and a unique function $h(t,\alpha,\beta) \in \mathcal{C}^2(J_2(\alpha_0,\beta_0), H_0)$ such that for all $(t,\alpha,\beta) \in J_2(\alpha_0,\beta_0)$, $F(t,\alpha,\beta,h(t,\alpha,\beta)) = 0$. And we define

$$v(t,\alpha,\beta) = \frac{1+h(t,\alpha,\beta)}{\|1+h(t,\alpha,\beta)\|} \in \mathcal{C}^2(J_2(\alpha_0,\beta_0),\mathcal{B})$$

$$\lambda(t,\alpha,\beta) = \langle \nu, Q(t,\alpha,\beta)(1+h(t,\alpha,\beta))\rangle \in \mathcal{C}^2(J_2(\alpha_0,\beta_0),\mathbb{C}).$$

Note that $h(0,\alpha,\beta) = 0$, $v(0,\alpha,\beta) = 1$, and $\lambda(0,\alpha,\beta) = 1$ on $J_2(\alpha_0,\beta_0)$.

(2) For $f \in \mathcal{B}$, let $f^\perp = \{\psi \in \mathcal{B}' : \langle \psi, f\rangle = 0\}$. Let $H' = 1^\perp = \{\psi \in \mathcal{B}' : \langle \psi, 1\rangle = 0\}$.

Define the function $G_1 : J_2(\alpha_0, \beta_0) \times H' \to \mathcal{B}'$ by

$$G_1(t, \alpha, \beta, \psi) = Q(t, \alpha, \beta)^*(\nu + \psi) - \lambda(t, \alpha, \beta)(\nu + \psi).$$

Note that $\mathcal{B}' = H' \oplus \mathrm{span}\{\nu\}$. Define the projection π onto H' by $\pi\psi = \psi - \langle \psi, 1 \rangle \nu$ for each $\psi \in \mathcal{B}'$, and the function $G : J_2(\alpha_0, \beta_0) \times H' \to H'$ by the composition $G = \pi \circ G_1$. By the continuity of $v(t, \alpha, \beta)$, there exists a neighborhood $J_3(\alpha_0, \beta_0)$ smaller than $J_2(\alpha_0, \beta_0)$ such that $\langle \nu, v(t, \alpha, \beta) \rangle \neq 0$ for all $(t, \alpha, \beta) \in J_3(\alpha_0, \beta_0)$. So, for each $\psi \in \mathcal{B}'$,

$$\psi = \frac{\langle \psi, v(t, \alpha, \beta) \rangle}{\langle \nu, v(t, \alpha, \beta) \rangle} \nu + \psi - \frac{\langle \psi, v(t, \alpha, \beta) \rangle}{\langle \nu, v(t, \alpha, \beta) \rangle} \nu,$$

and $\mathcal{B}' = v(t, \alpha, \beta)^\perp \oplus \mathrm{span}\{\nu\}$. Therefore, for each $(t, \alpha, \beta) \in J_3(\alpha_0, \beta_0)$, π is bijective from $v(t, \alpha, \beta)^\perp$ to H'. On the other hand, since $G(0, \alpha_0, \beta_0, 0) = 0$ and the partial differential of G at point $(0, \alpha_0, \beta_0, 0)$ with respect to ψ is $Q^*|_{H'} - 1$, which is an isomorphism of H', we can find a neighborhood $J_4(\alpha_0, \beta_0) \subset J_3(\alpha_0, \beta_0))$ of $(0, \alpha_0, \beta_0)$, a neighborhood $H'_0 \subset H'$ of 0, and a unique function $\psi \in \mathcal{C}^2(J_4(\alpha_0, \beta_0), H'_0)$ such that, for all $(t, \alpha, \beta) \in J_4(\alpha_0, \beta_0)$, $G(t, \alpha, \beta, \psi(t, \alpha, \beta)) = 0$ and $\langle \nu + \psi(t, \alpha, \beta), v(t, \alpha, \beta) \rangle \neq 0$. By the fact that π is bijective from $v(t, \alpha, \beta)^\perp$ to H', $G_1(t, \alpha, \beta, \psi(t, \alpha, \beta)) = 0$. Then it is clear that $\nu + \psi(t, \alpha, \beta)$ is an eigenvector of $Q^*(t, \alpha, \beta)$. We define

$$\varphi(t, \alpha, \beta) = \frac{\nu + \psi(t, \alpha, \beta)}{\langle \nu + \psi(t, \alpha, \beta), v(t, \alpha, \beta) \rangle} \in \mathcal{C}^2(J_4(\alpha_0, \beta_0), \mathcal{B}'),$$

and define $N(t, \alpha, \beta) \in C^2(J_4(\alpha_0, \beta_0), \mathcal{L}_{\mathcal{B}})$ using (3.7). Again, it is clear that $\psi(0, \alpha, \beta) = 0$, $\varphi(0, \alpha, \beta) = \nu$, and $N(0, \alpha, \beta) = N$ on $J_4(\alpha_0, \beta_0)$.

(3) For each $f \in \mathcal{B}$, we have the equality

$$f = \langle \varphi(t,\alpha,\beta), f\rangle v(t,\alpha,\beta) + (f - \langle \varphi(t,\alpha,\beta), f\rangle v(t,\alpha,\beta)).$$

Hence, we have the decomposition of $\mathcal{B}$ as claimed, and it is not hard to verify that $N(t,\alpha,\beta) = Q(t,\alpha,\beta)\Pi_{H(t,\alpha,\beta)}$. Observe that $r(N) < 1$ (by the quasi-compactness of Q), $\lambda(t,\alpha,\beta)$ and $N(t,\alpha,\beta)$ are continuous, and spectral radius is an upper-semicontinuous function of the operator. Therefore, $r(N(t,\alpha,\beta))/|\lambda(t,\alpha,\beta)| < r := (r(N)+1)/2 < 1$ holds on a neighborhood $J_5(\alpha_0,\beta_0) \subset J_4(\alpha_0,\beta_0)$.

Finally, to get the bound in (vi), we express $\nabla_t^2(N(t,\alpha,\beta)^n)$ as a sum of n^2 terms of the forms

$$V = N(t,\alpha,\beta)^{n_1}\nabla_t N(t,\alpha,\beta) N(t,\alpha,\beta)^{n_2}\nabla_t N(t,\alpha,\beta)^* N(t,\alpha,\beta)^{n_3} \tag{3.8}$$

with $n_1+n_2+n_3 = n-2$, $n_1,n_2,n_3 \geq 0$, or

$$V = N(t,\alpha,\beta)^{n_1}\nabla_t^2 N(t,\alpha,\beta) N(t,\alpha,\beta)^{n_2} \tag{3.9}$$

with $n_1+n_2 = n-1$, $n_1,n_2 \geq 0$. Let us choose $m_0 \in \mathbb{N}$ such that $\|N^{m_0}\| < (\frac{2r(N)+1}{3})^{m_0}$. Since $N(t,\alpha,\beta) \in \mathcal{C}^2(I\times[0,1]^2, \mathcal{L}_{\mathcal{B}})$, there exist a neighborhood $J(\alpha_0,\beta_0) \subset J_5(\alpha_0,\beta_0)$ and a constant $C_1(\alpha_0,\beta_0) \geq 1$ such that $\|N(t,\alpha,\beta)^{m_0}\| < r^{m_0}$, and $N(t,\alpha,\beta)$, $\nabla_t N(t,\alpha,\beta)$, and $\nabla_t^2 N(t,\alpha,\beta)$ are bounded by $C_1(\alpha_0,\beta_0)$ on $J(\alpha_0,\beta_0)$. Let $C_2(\alpha_0,\beta_0) = (C_1(\alpha_0,\beta_0)/r)^{m_0}$. Then, for every $m \in \mathbb{N}$, with $m = km_0 + j$, $0 \leq j < m_0$, we have on $J(\alpha_0,\beta_0)$ that

$$\|N(t,\alpha,\beta)^m\| \leq \|N(t,\alpha,\beta)^{m_0}\|^k \cdot \|N(t,\alpha,\beta)\|^j \leq r^m \cdot \frac{C_1(\alpha_0,\beta_0)^j}{r^j} \leq C_2(\alpha_0,\beta_0) r^m.$$

Hence, we obtain that, for each term V in the form of (3.8) or (3.9), $\|V\| \le C_2(\alpha_0, \beta_0)^5 \cdot r^{n-2}$. Since there are n^2 terms in total, we have that

$$\|\nabla_t^2(N(t,\alpha,\beta)^n)\| \le C_2(\alpha_0,\beta_0)^5 n^2 \cdot r^{n-2} \to 0,$$

as $n \to \infty$. Therefore, the left-hand side is bounded uniformly in n and $(t,\alpha,\beta) \in J(\alpha_0,\beta_0)$.

(4) So far, we showed that $\lambda(t,\alpha,\beta), v(t,\alpha,\beta)$, $\varphi(t,\alpha,\beta)$, and $N(t,\alpha,\beta)$ that satisfy the decomposition (3.7) and the conditions (i)-(vi) stated in the theorem exist on $\bigcup_{\alpha_0,\beta_0\in[0,1]} J(\alpha_0,\beta_0)$. Then, by the compactness of $[0,1]^2$, we can find a neighborhood I of the origin in $\mathbb{R}^d$ such that the functions above exist on $I \times [0,1]^2$ and the conditions (v) and (vi) are satisfied with uniform constants. $\square$

3.3 Product of the operators

As suggested by Lemma 3.2.1, the main ingredient of the proof is to study the limiting behavior of the product of the operators. In this section, we deal with two situations where t is close to or away from the origin. The product contains the information about the characteristics of the limiting distribution in the first situation and vanishes as T tends to infinity in the second situation. Our arguments involve different products. To avoid confusion, we use the convention: for $p_k \in \mathbb{C}$, $\prod_{k=j}^{j-1} p_j = 1$; and, for $P_k \in \mathcal{L}_{\mathcal{B}}$, $\prod_{k=j}^{j-1} P_j = \mathrm{id}$ is the identity operator.

3.3.1 For t near 0

In this section, we apply the perturbation theorem to deal with the product where t is close to the origin. To present cleaner arguments, we assume that the operator N in (3.2) has its norm less than 1. This assumption is not restrictive since, by the uniform ergodicity assumption (A2), if we define $Q = P_n$ with n large enough instead of $Q = P_1$, then the total variation between the distribution of X_n and the invariant measure ν will be small enough, and the norm of N will be less than 1. The perturbation theorem still holds with the relation in *(v)* replaced by

$$\|N(t,\alpha,\beta)\|/|\lambda(t,\alpha,\beta)| < r, \tag{3.10}$$

and the left-hand side of (3.6) can still be represented as a product of those perturbed operators.

Proposition 3.3.1. *Let I be defined as in Theorem 3.2.2. There exist constants M and N large enough such that, for all $T > N$ and all $t \in I$, we have the formula for the product of the operators, for $f \in \mathcal{B}$ and $0 \le i \le \lfloor T \rfloor - 1$,*

$$\begin{aligned}
&\prod_{k=i}^{\lfloor T\rfloor-1} Q(t,\frac{k}{T},\frac{k+1}{T})f \\
&= \prod_{k=i}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T}) \prod_{k=i}^{\lfloor T\rfloor-2} \langle \varphi(t,\frac{k}{T},\frac{k+1}{T}), v(t,\frac{k+1}{T},\frac{k+2}{T})\rangle \\
&\qquad \cdot \langle \varphi(t,\frac{\lfloor T\rfloor-1}{T},\frac{\lfloor T\rfloor}{T}), f\rangle v(t,\frac{i}{T},\frac{i+1}{T}) \\
&+ \prod_{k=i}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T})\|f\| \left(p_{i,T}^t \cdot v(t,\frac{i}{T},\frac{i+1}{T}) + q_{i,T}^t \cdot h(t,\frac{i}{T},\frac{i+1}{T}) \right),
\end{aligned} \tag{3.11}$$

where $h(t,\alpha,\beta) \in H(t,\alpha,\beta)$, $\|h(t,\alpha,\beta)\| = 1$, $|p_{i,T}^t| < M/T$, and $|q_{i,T}^t| < Mr^{\lfloor T\rfloor - i} + M/2T$. In particular, $|p_{0,T}^t|, |q_{0,T}^t| < M/T$. The bounds on $p_{i,T}^t, q_{i,T}^t$ hold uniformly in $f \in \mathcal{B}$.

The proof of this result requires careful treatment of error terms induced by the shifts of the eigenspace corresponding to the top eigenvalue. Roughly speaking, the idea is to show that, compared with the first term in (3.11), which is obtained by exclusively looking at the vectors and projections on the eigenspace, error terms that emerge because of the shifts of the eigenspace and the presence of the operator $N(t, \frac{k}{T}, \frac{k+1}{T})$ are negligible.

Proof. We apply Theorem 3.2.2 to control the error brought by the shift of the eigenspace. Namely, there exists $K > 0$ such that the following relations hold for $t \in I$:

(i) For all $1 \leq k \leq \lfloor T \rfloor - 1$,

$$\begin{aligned}
&\left| \langle \varphi(t, \frac{k-1}{T}, \frac{k}{T}), v(t, \frac{k}{T}, \frac{k+1}{T}) \rangle - 1 \right| \\
&\leq \left| \langle \varphi(t, \frac{k-1}{T}, \frac{k}{T}), v(t, \frac{k}{T}, \frac{k+1}{T}) - v(t, \frac{k-1}{T}, \frac{k}{T}) \rangle \right| \\
&\leq K/T.
\end{aligned} \tag{3.12}$$

Consequently,

$$\left| \prod_{k=i}^{\lfloor T \rfloor - 1} \langle \varphi(t, \frac{k-1}{T}, \frac{k}{T}), v(t, \frac{k}{T}, \frac{k+1}{T}) \rangle \right| \leq e^K. \tag{3.13}$$

Moreover, by (3.10),

$$\begin{aligned}
&\| N(t, \frac{k-1}{T}, \frac{k}{T}) v(t, \frac{k}{T}, \frac{k+1}{T}) \| \\
&= \| N(t, \frac{k-1}{T}, \frac{k}{T}) \Pi_{H(t, \frac{k-1}{T}, \frac{k}{T})} v(t, \frac{k}{T}, \frac{k+1}{T}) \| \\
&\leq \| N(t, \frac{k-1}{T}, \frac{k}{T}) \| \| \Pi_{H(t, \frac{k-1}{T}, \frac{k}{T})} v(t, \frac{k}{T}, \frac{k+1}{T}) \| \\
&\leq r |\lambda(t, \frac{k-1}{T}, \frac{k}{T})| \| \Pi_{H(t, \frac{k-1}{T}, \frac{k}{T})} (v(t, \frac{k}{T}, \frac{k+1}{T}) - v(t, \frac{k-1}{T}, \frac{k}{T})) \| \\
&\leq r |\lambda(t, \frac{k-1}{T}, \frac{k}{T})| \cdot \frac{K}{T}.
\end{aligned} \tag{3.14}$$

(ii) For all $1 \leq k \leq \lfloor T \rfloor - 1$, $h \in H(t, \frac{k}{T}, \frac{k+1}{T})$, $\|h\| = 1$,

$$\left|\langle \varphi(t, \frac{k-1}{T}, \frac{k}{T}), h\rangle\right| \leq \left|\langle \varphi(t, \frac{k-1}{T}, \frac{k}{T}) - \varphi(t, \frac{k}{T}, \frac{k+1}{T}), h\rangle\right| \leq \frac{K}{T}, \tag{3.15}$$

$$\|N(t, \frac{k-1}{T}, \frac{k}{T})h\| \leq r|\lambda(t, \frac{k-1}{T}, \frac{k}{T})|. \tag{3.16}$$

Furthermore, we can choose K sufficiently large so that $\|\varphi(t, \alpha, \beta)\|$ is bounded by K for all $t \in I$, $\alpha, \beta \in [0, 1]$. Let a denote $2Ke^K/T$. We prove by backwards induction that, for all $t \in I$, $0 \leq i \leq \lfloor T \rfloor - 1$, and all T sufficiently large, (3.11) holds with

$$\begin{aligned} |p_{i,T}| &\leq \frac{ar(1+a)^{\lfloor T \rfloor - i - 2}(1 - r^{\lfloor T \rfloor - i - 1})}{1-r} + \frac{ar[(1+a)^{\lfloor T \rfloor - i - 1} - 1]}{1 - r - ar}, \\ |q_{i,T}| &\leq (1+a)^{\lfloor T \rfloor - i - 1} r^{\lfloor T \rfloor - i} + \frac{ar}{1 - r - ar}. \end{aligned} \tag{3.17}$$

In particular, it is not hard to see that this implies the desired result.

Let us verify that (3.17) holds if $i = \lfloor T \rfloor - 1$. By (3.7),

$$\begin{aligned} Q(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T})f = \lambda(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T})\langle \varphi(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}), f\rangle v(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}) \\ + N(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T})f. \end{aligned}$$

We have $p_{\lfloor T \rfloor - 1, T} = 0$ and, by (3.10),

$$|q_{\lfloor T \rfloor - 1, T}| \leq \|N(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T})\| / \lambda(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}) \leq r.$$

By assuming the validity of (3.17) at i, we obtain

$$\begin{aligned}
&\prod_{k=i-1}^{\lfloor T\rfloor-1} Q(t,\frac{k}{T},\frac{k+1}{T})f \\
&= Q(t,\frac{i-1}{T},\frac{i}{T}) \prod_{k=i}^{\lfloor T\rfloor-1} Q(t,\frac{k}{T},\frac{k+1}{T})f \\
&= \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T}) \prod_{k=i-1}^{\lfloor T\rfloor-2} \langle\varphi(t,\frac{k}{T},\frac{k+1}{T}),v(t,\frac{k+1}{T},\frac{k+2}{T})\rangle \\
&\qquad \cdot \langle\varphi(t,\frac{\lfloor T\rfloor-1}{T},\frac{\lfloor T\rfloor}{T}),f\rangle v(t,\frac{i-1}{T},\frac{i}{T}) \\
&\quad + \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T})\|f\|\langle\varphi(t,\frac{i-1}{T},\frac{i}{T}),v(t,\frac{i}{T},\frac{i+1}{T})\rangle p_{i,T} v(t,\frac{i-1}{T},\frac{i}{T}) && (3.18) \\
&\quad + \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T})\|f\|\langle\varphi(t,\frac{i-1}{T},\frac{i}{T}),h(t,\frac{i}{T},\frac{i+1}{T})\rangle q_{i,T} v(t,\frac{i-1}{T},\frac{i}{T}) && (3.19) \\
&\quad + \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T}) \prod_{k=i}^{\lfloor T\rfloor-2} \langle\varphi(t,\frac{k}{T},\frac{k+1}{T}),v(t,\frac{k+1}{T},\frac{k+2}{T})\rangle\langle\varphi(t,\frac{\lfloor T\rfloor-1}{T},\frac{\lfloor T\rfloor}{T}),f\rangle \\
&\qquad \cdot N(t,\frac{i-1}{T},\frac{i}{T})v(t,\frac{i}{T},\frac{i+1}{T})/\lambda(t,\frac{i-1}{T},\frac{i}{T}) && (3.20) \\
&\quad + \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T})\|f\| p_{i,T} N(t,\frac{i-1}{T},\frac{i}{T})v(t,\frac{i}{T},\frac{i+1}{T})/\lambda(t,\frac{i-1}{T},\frac{i}{T}) && (3.21) \\
&\quad + \prod_{k=i-1}^{\lfloor T\rfloor-1} \lambda(t,\frac{k}{T},\frac{k+1}{T})\|f\| q_{i,T} N(t,\frac{i-1}{T},\frac{i}{T})h(t,\frac{i}{T},\frac{i+1}{T})/\lambda(t,\frac{i-1}{T},\frac{i}{T}). && (3.22)
\end{aligned}$$

We deduce that, by (3.12) and (3.15),

$$|p_{i-1,T}| \le (1+a)|p_{i,T}| + a|q_{i,T}|, \tag{3.23}$$

where the two terms on the right-hand side correspond to (3.18) and (3.19), respectively; and, by

(3.13), (3.14), and (3.16),

$$|q_{i-1,T}| \le \frac{1}{2}ar + \frac{1}{2}ar|p_{i,T}| + (1+a)r|q_{i,T}|, \tag{3.24}$$

where the three terms on the right-hand side correspond to (3.20), (3.21), and (3.22), respectively.

Now using (3.23) and (3.24) and backwards induction on i, it is easy to see that (3.17) holds. Indeed, at each step of induction, we first show that $|p_{i,T}^t| < 1$. So (3.24) implies that

$$|q_{i-1,T}| \le ar + (1+a)r|q_{i,T}|, \tag{3.25}$$

when assuming (3.17) for $p_{i,T}^t$ and $q_{i,T}^t$. Then (3.17) for $i-1$ instead of i can be established using (3.23) and (3.25). As we noted, (3.17) implies the statement in the lemma. □

3.3.2 For t away from 0

As noted earlier, we expect that the product decays exponentially fast for all t that are at a uniformly positive distance from the origin, so it does not contribute to (3.6).

Lemma 3.3.2. *Suppose that $P(t,\alpha,\beta) \in \mathcal{C}(K \times [0,1]^2, \mathcal{L}_{\mathcal{B}})$ where K is a compact set in $\mathbb{R}^d$ and that $r(P(t,\alpha,\beta)) < r_0$. Then*

(i) There exists $r_1 < r_0$ and m_0 such that $\|P(t,\alpha,\beta)^m\| < r_1^m$ for all $t \in K$, $\alpha, \beta \in [0,1]$, and $m \ge m_0$.

(ii) There exist $r_2 < r_0$ and m_0 such that for each $m \ge m_0$ there exists $\delta > 0$ such that for all

$t \in K$,

$$\|\prod_{k=1}^{m} P(t, \alpha_k, \beta_k)\| < r_2^m$$

if $|\alpha_k - \alpha_1| < \delta$ and $|\beta_k - \beta_1| < \delta$ for all $1 \leq k \leq m$.

Proof. (i) This result follows from the proof of Corollary III.13 in [33].

(ii) Let m_0 be defined as in (i) and fix arbitrary $r_2 \in (r_1, r)$ and $m \geq m_0$. Define

$$P_m(t, \alpha_1, ..., \alpha_m, \beta_1, ..., \beta_m) = \prod_{k=1}^{m} P(t, \alpha_k, \beta_k).$$

Since $P(t, \alpha, \beta)$ is continuous in t, α, and β, P_m is continuous on $K \times [0,1]^{2m}$, hence is uniformly continuous. Then there exists $\delta > 0$ such that

$$\|P_m(t, \alpha_1, ..., \alpha_1, \beta_1, ..., \beta_1) - P_m(t, \alpha_1, ..., \alpha_m, \beta_1, ..., \beta_m)\| \leq r_2^m - r_1^m,$$

if $|\alpha_k - \alpha_1| < \delta$ and $|\beta_k - \beta_1| < \delta$ for all $1 \leq k \leq m$. Therefore, the result follows. □

Lemma 3.3.3. *For every compact set $K \subset \mathbb{R}^d \setminus \{0\}$, there exist constants $r_K \in (0,1)$ and N large enough such that, for all $T > N$ and all $t \in K$,*

$$\|\prod_{k=0}^{\lfloor T \rfloor - 1} Q(t, \frac{k}{T}, \frac{k+1}{T})\| < r_K^{\lfloor T \rfloor}.$$

Proof. By Lemma 3.3.2.(ii) and Assumption (A4), there exist $r_2 < 1$, $m \in \mathbb{N}$ large enough, and $\delta > 0$ such that

$$\|\prod_{k=1}^{m} Q(t, \alpha_k, \beta_k)\| < r_2^m$$

holds if $|\alpha_k - \alpha_1| < \delta$ and $|\beta_k - \beta_1| < \delta$ for all $1 \le k \le m$. Now fix arbitrary $r_K \in (r_2, 1)$. Choose $l_0 \in \mathbb{N}$ such that $(r_2/r_K)^{l_0} < r_K$ and choose $N > 3m(l_0 \vee 1/\delta)$. Let $\lfloor T \rfloor = \tilde{l}m + j$ with $0 \le j < m$. Then we have $\tilde{l} \ge l_0$ and

$$\begin{aligned}\|\prod_{k=0}^{\lfloor T\rfloor-1} Q(t,\frac{k}{T},\frac{k+1}{T})\| &\le \|\prod_{l=1}^{\tilde{l}}\prod_{k=(l-1)m}^{lm-1} Q(t,\frac{k}{T},\frac{k+1}{T})\|\cdot\|\prod_{k=\tilde{l}m}^{\lfloor T\rfloor-1} Q(t,\frac{k}{T},\frac{k+1}{T})\| \\ &\le \prod_{l=1}^{\tilde{l}}\|\prod_{k=(l-1)m}^{lm-1} Q(t,\frac{k}{T},\frac{k+1}{T})\| \le r_2^{\tilde{l}m} < r_K^{\tilde{l}m}\cdot r_K^m < r_K^{\lfloor T\rfloor}. \qquad \square\end{aligned}$$

3.3.3 Product of the top eigenvalues

In this section, we obtain the asymptotic value of the product of the top eigenvalues by the Taylor expansion near 0. The first result concerns the first and second derivatives of $\lambda(t,\alpha,\beta)$ w.r.t. t at 0.

Lemma 3.3.4. *(i) For all $\alpha, \beta \in [0,1]$, we have $\nabla_t\lambda(0,\alpha,\beta) = 0$.*

(ii) For all $\alpha \in [0,1]$, $\nabla_t^2\lambda(0,\alpha,\alpha)$ is real and negative-definite. Moreover,

$$\nabla_t^2\lambda(0,\alpha,\alpha) = -\lim_{T\to\infty}\frac{1}{T}\mathbf{E}_\nu(S_T^\alpha S_T^{\alpha *}), \tag{3.26}$$

where $S_T^\alpha = \int_0^T b(\alpha, X_s)ds$.

Proof. *(i).*With the same notation as in Theorem 3.2.2, we have the decomposition

$$Q(t,\alpha,\beta) = \lambda(t,\alpha,\beta)\tilde{v}(t,\alpha,\beta)\otimes\tilde{\varphi}(t,\alpha,\beta) + N(t,\alpha,\beta),$$

for $t \in I$ and $\alpha, \beta \in [0, 1]$, where

$$\tilde{v}(t, \alpha, \beta) = \frac{v(t, \alpha, \beta)}{\langle \nu, v(t, \alpha, \beta) \rangle} \quad \text{and} \quad \tilde{\varphi}(t, \alpha, \beta) = \langle \nu, v(t, \alpha, \beta) \rangle \varphi(t, \alpha, \beta).$$

Compared to the previous decomposition, we no longer have $\|\tilde{v}(t, \alpha, \beta)\| \equiv 1$, but instead have $\langle \tilde{\varphi}(0, \alpha, \beta), \tilde{v}(t, \alpha, \beta) \rangle \equiv 1$, and still have $Q(t, \alpha, \beta)\tilde{v}(t, \alpha, \beta) = \lambda(t, \alpha, \beta)\tilde{v}(t, \alpha, \beta)$. Differentiate the equations with respect to t. Then

$$\begin{aligned} \langle \tilde{\varphi}(0, \alpha, \beta), \nabla_t \tilde{v}(t, \alpha, \beta) \rangle &\equiv 0, \\ \nabla_t Q(t, \alpha, \beta)\tilde{v}(t, \alpha, \beta) + Q(t, \alpha, \beta)\nabla_t \tilde{v}(t, \alpha, \beta) & \\ = \nabla_t \lambda(t, \alpha, \beta)\tilde{v}(t, \alpha, \beta) &+ \lambda(t, \alpha, \beta)\nabla_t \tilde{v}(t, \alpha, \beta). \end{aligned} \tag{3.27}$$

Note that $\tilde{v}(0, \alpha, \beta) = 1$, $\tilde{\varphi}(0, \alpha, \beta) = \nu$, $\lambda(0, \alpha, \beta) = 1$, and $Q(0, \alpha, \beta) = Q$. Applying ν to (3.27) at $t = 0$, we obtain, by Assumption (A3),

$$\nabla_t \lambda(0, \alpha, \beta) = \langle \nu, \nabla_t Q(0, \alpha, \beta)1 \rangle = i\mathbf{E}_\nu (\int_0^1 b(\alpha + s(\beta - \alpha), X_s)ds) = 0. \tag{3.28}$$

*(ii).*Taking (3.28) into account, for $t = 0$ and $\alpha = \beta$, (3.27) reduces to

$$(1 - Q)\nabla_t \tilde{v}(0, \alpha, \alpha) = \nabla_t Q(0, \alpha, \alpha)1 = i\mathbf{E}_x (\int_0^1 b(\alpha, X_s)ds). \tag{3.29}$$

Since both $\nabla_t \tilde{v}(0, \alpha, \alpha)$ and $i\mathbf{E}_x(\int_0^1 b(\alpha, X_s)ds)$ have zero mean w.r.t ν, they are in the closed subspace H. Also recall that 1 is not in the spectrum of $Q_{|H}$. Therefore, $\nabla_t \tilde{v}(0, \alpha, \alpha)$ can be viewed as a solution to (3.29), and thus,

$$\nabla_t \tilde{v}(0, \alpha, \alpha) = iu(\alpha) \tag{3.30}$$

for a certain real-valued bounded function $u(\alpha)$ that depends on the parameter α. Take second derivatives on both sides of the equality $Q(t, \alpha, \alpha)\tilde{v}(t, \alpha, \alpha) = \lambda(t, \alpha, \alpha)\tilde{v}(t, \alpha, \alpha)$ and apply ν to the result at $t = 0$ (here $*$ stands for the transpose without taking complex conjugate):

$$\nabla_t^2 \lambda(0, \alpha, \alpha) = \langle \nu, \nabla_t^2 Q(0, \alpha, \alpha)1 + \nabla_t Q(0, \alpha, \alpha)\nabla_t \tilde{v}(0, \alpha, \alpha)^* + (\nabla_t Q(0, \alpha, \alpha)\nabla_t \tilde{v}(0, \alpha, \alpha)^*)^* \rangle. \tag{3.31}$$

Let us compute the terms on the right-hand side by recalling the definition in (3.3) and (3.30):

$$\begin{aligned} \langle \nu, \nabla_t^2 Q(0, \alpha, \alpha)1 \rangle &= -\mathbf{E}_\nu[(\int_0^1 b(\alpha, X_s)ds)(\int_0^1 b(\alpha, X_s)ds)^*], \\ \langle \nu, \nabla_t Q(0, \alpha, \alpha)\nabla_t \tilde{v}(0, \alpha, \alpha)^* \rangle &= -\mathbf{E}_\nu[(\int_0^1 b(\alpha, X_s)ds)u(\alpha)(X_1)^*]. \end{aligned} \tag{3.32}$$

Then using (3.31) and (3.32) for the first equality below and (3.29) and (3.30) for the second equality below, we obtain that

$$\begin{aligned} &\nabla_t^2 \lambda(0, \alpha, \alpha) \\ &= -\mathbf{E}_\nu \left[\left(\int_0^1 b(\alpha, X_s)ds + u(\alpha)(X_1) \right) \left(\int_0^1 b(\alpha, X_s)ds + u(\alpha)(X_1) \right)^* \right] + \mathbf{E}_\nu u(\alpha)u(\alpha)^* \\ &= -\mathbf{E}_\nu \left[\left(\int_0^1 b(\alpha, X_s)ds + u(\alpha)(X_1) - u(\alpha)(x) \right) \left(\int_0^1 b(\alpha, X_s)ds + u(\alpha)(X_1) - u(\alpha)(x) \right)^* \right]. \end{aligned}$$

Hence $\nabla_t^2 \lambda(0, \alpha, \alpha)$ is negative semi-definite. To see that it is negative definite, let us assume that there exists a non-zero vector $t \in \mathbb{R}^d$ such that $t^* \nabla_t^2 \lambda(0, \alpha, \alpha) t = 0$. Then we have ν-a.s. that

$$\mathbf{E}_x \left[t \cdot \left(\int_0^1 b(\alpha, X_s) ds + u(\alpha)(X_1) - u(\alpha)(x) \right) \right]^2 = 0,$$

which implies that $t \cdot \left(\int_0^1 b(\alpha, X_s) ds + u(\alpha)(X_1) - u(\alpha)(x) \right) = 0$, ν-a.s. However, this contradicts Assumption (A4). Indeed, let $f(x) = \exp(it \cdot u(\alpha)(x))$. Then we have

$$Q(t, \alpha, \alpha) f(x) = \mathbf{E}_x \left[\exp \left(it \cdot \left(\int_0^1 b(\alpha, X_s) ds + u(\alpha)(X_1) - u(\alpha)(x) \right) \right) \right] f(x).$$

Then $Q(t, \alpha, \alpha) f(x) = f(x)$, ν-a.s., which contradicts the fact that $r(Q(t, \alpha, \alpha)) < 1$.

To prove (3.26), we write, for $m \in \mathbb{N}$,

$$Q(t, \alpha, \alpha)^m = \lambda(t, \alpha, \alpha)^m v(t, \alpha, \alpha) \otimes \varphi(t, \alpha, \alpha) + N^m(t, \alpha, \alpha).$$

Taking the second derivatives, we obtain

$$\begin{aligned}
&\nabla_t^2 \left(Q(t, \alpha, \alpha)^m \right) \\
&= \nabla_t^2 (\lambda(t, \alpha, \alpha)^m) v(t, \alpha, \alpha) \otimes \varphi(t, \alpha, \alpha) + \nabla_t (v(t, \alpha, \alpha) \otimes \varphi(t, \alpha, \alpha)) \nabla_t (\lambda(t, \alpha, \alpha)^m)^* \\
&\quad + \nabla_t (\lambda(t, \alpha, \alpha)^m) \nabla_t (v(t, \alpha, \alpha) \otimes \varphi(t, \alpha, \alpha))^* + R_m(t, \alpha, \alpha),
\end{aligned}$$

where $R_m(t, \alpha, \alpha) = \lambda(t, \alpha, \alpha)^m \nabla_t^2 (v(t, \alpha, \alpha) \otimes \varphi(t, \alpha, \alpha)) + \nabla_t^2 (N^m(t, \alpha, \alpha))$. Note that

$$\begin{aligned}
\nabla_t^2 (\lambda(t, \alpha, \alpha)^m) = m(m-1) \lambda(t, \alpha, \alpha)^{m-2} \nabla_t \lambda(t, \alpha, \alpha) \nabla_t \lambda(t, \alpha, \alpha)^* \\
+ m \lambda(t, \alpha, \alpha)^{m-1} \nabla_t^2 \lambda(t, \alpha, \alpha).
\end{aligned}$$

Then it follows that $\nabla_t^2(\lambda(t,\alpha,\alpha)^m)(0,\alpha,\alpha) = m\nabla_t^2\lambda(0,\alpha,\alpha)$. By Theorem 3.2.2, $R_m(0,\alpha,\alpha)$ is uniformly bounded for all m. Thus, for all $\mu \in \mathcal{B}'_p$,

$$\sup_{m\geq 1} |\langle \mu, m\nabla_t^2\lambda(0,\alpha,\alpha)\cdot 1 - \nabla_t^2(Q^m(0,\alpha,\alpha))\cdot 1\rangle| = \sup_{m\geq 1} |\langle \mu, R_m(0,\alpha,\alpha)\cdot 1\rangle| < \infty.$$

Therefore, for all initial distribution $\mu \in \mathcal{B}'_p$,

$$\begin{aligned}
\nabla_t^2\lambda(0,\alpha,\alpha) &= \lim_{m\to\infty} \frac{1}{m}\langle \mu, \nabla_t^2(Q^m(t,\alpha,\alpha))(0,\alpha,\alpha)\cdot 1\rangle \\
&= -\lim_{m\to\infty} \frac{1}{m}\mathbf{E}_\mu(S_m^\alpha S_m^{\alpha\,*}),
\end{aligned}$$

which implies (3.26). □

We close this section by computing the product of the eigenvalues.

Lemma 3.3.5. *For $\tau \in \mathbb{R}^d$,*

$$\lim_{T\to\infty} \prod_{k=0}^{\lfloor T\rfloor -1} \lambda(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}) = \exp(-\frac{1}{2}\tau^*\Sigma\Sigma^*\tau),$$

where $\Sigma\Sigma^* := -\int_0^1 \nabla_t^2\lambda(0,\alpha,\alpha)d\alpha$ is positive-definite.

Proof. Since $\lambda(t,\alpha,\beta) \in \mathcal{C}^2(\mathbb{R}^d\times[0,1]^2,\mathbb{C})$, $\nabla_t^2\lambda(0,\alpha,\beta)$ is continuous on $[0,1]^2$ and hence uniformly continuous. Therefore,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{\lfloor T\rfloor -1}\nabla_t^2\lambda(0,\frac{k}{T},\frac{k+1}{T}) &= -\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{\lfloor T\rfloor -1}\nabla_t^2\lambda(0,\frac{k}{T},\frac{k}{T}) \\
&= -\int_0^1 \nabla_t^2\lambda(0,\alpha,\alpha)d\alpha \\
&= -\Sigma\Sigma^*.
\end{aligned}$$

Since, for each $\alpha \in [0,1]$, $\nabla_t^2 \lambda(0,\alpha,\alpha)$ is negative-definite, $\Sigma\Sigma^*$ is positive-definite. Finally, the product of the top eigenvalues at $t = \tau/\sqrt{T}$ is computed:

$$\begin{aligned}
\lim_{T\to\infty} \prod_{k=0}^{\lfloor T \rfloor -1} \lambda(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}) &= \lim_{T\to\infty} \prod_{k=0}^{\lfloor T \rfloor -1} (1 + \frac{1}{2T}\tau^* \nabla_t^2 \lambda(0, \frac{k}{T}, \frac{k+1}{T})\tau + o(\frac{1}{T})) \\
&= \exp\left(\lim_{T\to\infty} \frac{1}{2T} \sum_{k=0}^{\lfloor T \rfloor -1} \tau^* \nabla_t^2 \lambda(0, \frac{k}{T}, \frac{k+1}{T})\tau \right) \\
&= \exp\left(-\frac{1}{2}\tau^* \Sigma\Sigma^* \tau \right). \qquad \square
\end{aligned}$$

3.4 Proof of the main result

In this section, we prove the main result - Theorem 3.1.4. Let us first make some simple computations needed for the proof of the next lemma. Recall that I is the neighborhood of the origin defined in Theorem 3.2.2.

(i) Since $v(t,\alpha,\beta) \in \mathcal{C}^2(I \times [0,1]^2, \mathcal{B})$, $\nabla_t^2 v(0,\alpha,\beta)$ is uniformly continuous on $I \times [0,1]^2$. Thus, using the Taylor expansion for the first variable near zero, we have, for each $\tau \in \mathbb{R}^d$, all $0 \le k \le \lfloor T \rfloor - 2$, and all T sufficiently large,

$$\begin{aligned}
& v(\frac{\tau}{\sqrt{T}}, \frac{k+1}{T}, \frac{k+2}{T}) - v(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}) \\
&= 1 + \frac{\tau}{\sqrt{T}} \cdot \nabla_t v(0, \frac{k+1}{T}, \frac{k+2}{T}) + \frac{1}{2T}\tau^* \nabla_t^2 v(0, \frac{k+1}{T}, \frac{k+2}{T})\tau + o(\frac{1}{T}) \\
&\quad - 1 - \frac{\tau}{\sqrt{T}} \cdot \nabla_t v(0, \frac{k}{T}, \frac{k+1}{T}) - \frac{1}{2T}\tau^* \nabla_t^2 v(0, \frac{k}{T}, \frac{k+1}{T})\tau + o(\frac{1}{T}) \\
&= o(\frac{1}{T}).
\end{aligned} \tag{3.33}$$

(ii) Since $\varphi(t,\alpha,\beta)$ is bounded on $I \times [0,1]^2$, for each $\tau \in \mathbb{R}^d$ and all $0 \le k \le \lfloor T \rfloor - 2$, we have by part (iii) in Theorem 3.2.2 and (3.33),

$$\langle\varphi(\frac{\tau}{\sqrt{T}},\frac{k}{T},\frac{k+1}{T}),v(\frac{\tau}{\sqrt{T}},\frac{k+1}{T},\frac{k+2}{T})\rangle = 1+o(\frac{1}{T}).$$

Therefore, for each $\tau\in\mathbb{R}^d$, uniformly in $0\le i\le\lfloor T\rfloor-2$,

$$\lim_{T\to\infty}\prod_{k=i}^{\lfloor T\rfloor-2}\langle\varphi(\frac{\tau}{\sqrt{T}},\frac{k}{T},\frac{k+1}{T}),v(\frac{\tau}{\sqrt{T}},\frac{k+1}{T},\frac{k+2}{T})\rangle = 1. \tag{3.34}$$

Moreover, by arguments similar to those leading to (3.34), the product in (3.34) is uniformly bounded in $\tau/\sqrt{T}\in I$.

(iii) We showed in Lemma 3.3.4 that $\nabla_t^2\lambda(0,\alpha,\alpha)$ is negative-definite for all $\alpha\in[0,1]$. The absolute values of the eigenvalues of $\nabla_t^2\lambda(0,\alpha,\alpha)$ are bounded from below by a certain $c>0$ for all $\alpha\in[0,1]$. Then, by the continuity of $\nabla_t^2\lambda(0,\alpha,\beta)$, by making I smaller (if needed), we can make sure that for all T sufficiently large, all $\tau/\sqrt{T}\in I$, and all $0\le k\le\lfloor T\rfloor-2$,

$$\lambda(\frac{\tau}{\sqrt{T}},\frac{k}{T},\frac{k+1}{T})\le 1-\frac{c|\tau|^2}{3T}\le\exp(-\frac{c|\tau|^2}{3T}).$$

Then, for all $\tau/\sqrt{T}\in I$ and T sufficiently large,

$$\prod_{k=0}^{\lfloor T\rfloor-1}\lambda(\frac{\tau}{\sqrt{T}},\frac{k}{T},\frac{k+1}{T})\le\exp(-\frac{c|\tau|^2}{4}). \tag{3.35}$$

Combining this with the last statement in part (ii), we see that there exists $C_2>0$ such that for all $\tau/\sqrt{T} \in I$, $0 \le k \le \lfloor T \rfloor - 2$, and T sufficiently large,

$$\prod_{k=0}^{\lfloor T \rfloor -1} \lambda(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}) \prod_{k=0}^{\lfloor T \rfloor -2} \langle \varphi(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}), v(\frac{\tau}{\sqrt{T}}, \frac{k+1}{T}, \frac{k+2}{T}) \rangle \le C_2 \exp(-\frac{c|\tau|^2}{4}). \tag{3.36}$$

Define two (possibly complex) measures on $\mathbb{R}^d$ for $u \in \mathbb{R}^d$, $f \in \mathcal{B}$, and $\mu \in \mathcal{B}'_p$:

$$\begin{aligned} m_T^{u,f,\mu}(B) &= \det(\Sigma)(2\pi T)^{d/2} \mathbf{E}_\mu [f(X_T) 1_B(S_T - u)], \\ \tilde{m}_T^{u,f}(B) &= e^{-\frac{1}{2T} u^* (\Sigma\Sigma^*)^{-1} u} \langle \nu, f \rangle \mathfrak{L}(B), \end{aligned} \tag{3.37}$$

where Σ is defined in Lemma 3.3.5 and $\mathfrak{L}$ is the Lebesgue measure. We need to prove that they are sufficiently close as T tends to ∞. The following lemma is the main step towards this goal.

Lemma 3.4.1. *For each h that is a continuous integrable function on $\mathbb{R}^d$ with compactly supported Fourier transform and each constant $C > 0$, we have, uniformly in $u \in \mathbb{R}^d$, $f \in \mathcal{B}_C$, and initial distribution $\mu \in \mathcal{B}'_p$,*

$$|\langle m_T^{u,f,\mu}, h \rangle - \langle \tilde{m}_T^{u,f}, h \rangle| \to 0, \text{ as } T \to \infty.$$

Proof. Recall the expression in formula (3.6). By Proposition 3.3.1, there exists $M > 0$ such that, for all $t \in I$ and all T sufficiently large,

$$\begin{aligned} &\prod_{k=0}^{\lfloor T \rfloor -1} Q\left(t, \frac{k}{T}, \frac{k+1}{T}\right) \tilde{Q}(t,T) f \\ &= \prod_{k=0}^{\lfloor T \rfloor -1} \lambda(t, \frac{k}{T}, \frac{k+1}{T}) \prod_{k=0}^{\lfloor T \rfloor -2} \langle \mu(t, \frac{k}{T}, \frac{k+1}{T}), v(t, \frac{k+1}{T}, \frac{k+2}{T}) \rangle \\ &\quad \cdot \langle \varphi(t, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}), \tilde{Q}(t,T) f \rangle \cdot v(t, 0, \frac{1}{T}) \end{aligned}$$

$$+\prod_{k=0}^{\lfloor T\rfloor-1}\lambda(t,\frac{k}{T},\frac{k+1}{T})\|\tilde{Q}(t,T)f\|p_{0,T}^{t}\cdot h(t,0,\frac{1}{T})$$
$$+\prod_{k=0}^{\lfloor T\rfloor-1}\lambda(t,\frac{k}{T},\frac{k+1}{T})\|\tilde{Q}(t,T)f\|q_{0,T}^{t}\cdot v(t,0,\frac{1}{T}),$$

where $|p_{0,T}^t|, |q_{0,T}^t| < M/T$. By the formula of inverse Fourier transformation, with $\hat{h}$ denoting the Fourier transform of h,

$$\begin{aligned}\langle m_T^{u,f,\mu}, h\rangle &= \det(\Sigma)(2\pi T)^{d/2}\mathbf{E}_\mu[f(X_T)h(S_T-u)] \\ &= \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}\mathbf{E}_\mu\left[f(X_T)\int_{\mathbb{R}^d} e^{it\cdot(S_T-u)}\hat{h}(t)dt\right] \\ &= \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}\int_{\mathbb{R}^d}\hat{h}(t)e^{-it\cdot u}\langle\mu, \prod_{k=0}^{\lfloor T\rfloor-1} Q\left(t,\frac{k}{T},\frac{k+1}{T}\right)\tilde{Q}(t,T)f\rangle dt \\ &=: \mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3,\end{aligned}$$

where

$$\begin{aligned}\mathcal{A}_1 = \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}&\int_I \hat{h}(t)e^{-it\cdot u}\prod_{k=0}^{\lfloor T\rfloor-1}\lambda(t,\frac{k}{T},\frac{k+1}{T})\prod_{k=0}^{\lfloor T\rfloor-2}\langle\varphi(t,\frac{k}{T},\frac{k+1}{T}), v(t,\frac{k+1}{T},\frac{k+2}{T})\rangle \\ &\cdot\langle\varphi(t,\frac{\lfloor T\rfloor-1}{T},\frac{\lfloor T\rfloor}{T}),\tilde{Q}(t,T)f\rangle\langle\mu, v(t,0,\frac{1}{T})\rangle dt, \\ \mathcal{A}_2 = \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}&\int_I \hat{h}(t)e^{-it\cdot u}\prod_{k=0}^{\lfloor T\rfloor-1}\lambda(t,\frac{k}{T},\frac{k+1}{T})\|\tilde{Q}(t,T)f\| \\ &\cdot\left(p_{0,T}^t\langle\mu, h(t,0,\frac{1}{T})\rangle + q_{0,T}^t\langle\mu, v(t,0,\frac{1}{T})\rangle\right)dt, \\ \mathcal{A}_3 = \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}&\int_{I^c\cap K}\hat{h}(t)e^{-it\cdot u}\langle\mu, \prod_{k=0}^{\lfloor T\rfloor-1} Q\left(t,\frac{k}{T},\frac{k+1}{T}\right)\tilde{Q}(t,T)f\rangle dt,\end{aligned}$$

and K is the support of $\hat{h}$. By the change of variable with $t = \frac{\tau}{\sqrt{T}}$,

$$\mathcal{A}_1 = \int_{\mathbb{R}^d} \kappa_T^\mu(\tau) e^{-i\frac{\tau\cdot u}{\sqrt{T}}} \langle 1_I(\frac{\tau}{\sqrt{T}})\varphi(\frac{\tau}{\sqrt{T}}, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}), \tilde{Q}(\frac{\tau}{\sqrt{T}}, T) f\rangle d\tau,$$

where

$$\begin{aligned}\kappa_T^\mu(\tau) = \det(\Sigma)\left(\frac{1}{2\pi}\right)^{d/2} \cdot 1_I(\frac{\tau}{\sqrt{T}})\hat{h}(\frac{\tau}{\sqrt{T}}) \prod_{k=0}^{\lfloor T \rfloor -1} \lambda(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}) \\ \cdot \prod_{k=0}^{\lfloor T \rfloor -2} \langle \varphi(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T}), v(\frac{\tau}{\sqrt{T}}, \frac{k+1}{T}, \frac{k+2}{T})\rangle \cdot \langle \mu, v(\frac{\tau}{\sqrt{T}}, 0, \frac{1}{T})\rangle.\end{aligned}$$

Define

$$\kappa(\tau) = \det(\Sigma)(\frac{1}{2\pi})^{d/2}\hat{h}(0)\exp(-\tau^*\Sigma\Sigma^*\tau/2).$$

By Lemma 3.3.5 and (3.34), $\kappa_T^\mu(\tau) \to \kappa(\tau)$ as $T \to \infty$, for each τ, uniformly in $\mu \in \mathcal{B}_p'$. By (3.36), $|\kappa_T^\mu(\tau)| \leq C_3 \sup_{t\in I} |\hat{h}(t)| \exp(-\frac{c|\tau|^2}{4})$ with a certain constant $C_3 > 0$. Hence, by Lebesgue's dominated convergence theorem, uniformly in $\mu \in \mathcal{B}_p'$,

$$\int_{\mathbb{R}^d} |\kappa_T^\mu(\tau) - \kappa(\tau)| d\tau \to 0. \tag{3.38}$$

Note that $\tilde{Q}^*(\frac{\tau}{\sqrt{T}}, T)\varphi(\frac{\tau}{\sqrt{T}}, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}) \to \nu$ in $\mathcal{B}'$ as $T \to \infty$. Again, by Lebesgue's dominated convergence theorem, uniformly $\mu \in \mathcal{B}_p'$,

$$\int_{\mathbb{R}^d} \kappa_T^\mu(\tau) \| 1_I(\frac{\tau}{\sqrt{T}}) \tilde{Q}^*(\frac{\tau}{\sqrt{T}}, T)\varphi(\frac{\tau}{\sqrt{T}}, \frac{\lfloor T \rfloor - 1}{T}, \frac{\lfloor T \rfloor}{T}) - \nu \| d\tau \to 0. \tag{3.39}$$

From the definition of $\tilde{m}_T^{u,f}$ in (3.37) and the definition of $\kappa(\tau)$, it follows that

$$\langle \tilde{m}_T^{u,f}, h\rangle = \int_{\mathbb{R}^d} \kappa(\tau) e^{-i\frac{\tau\cdot u}{\sqrt{T}}}\langle \nu, f\rangle d\tau.$$

Therefore, for all $f \in \mathcal{B}_C$,

$$\begin{aligned}
&|\mathcal{A}_1 - \langle \tilde{m}_T^{u,f}, h\rangle| \\
&\leq |\int_{\mathbb{R}^d} (\kappa_T^{\mu}(\tau) - \kappa(\tau)) e^{-i\frac{\tau\cdot u}{\sqrt{T}}}\langle \nu, f\rangle d\tau| \\
&\quad + |\int_{\mathbb{R}^d} \kappa_T^{\mu}(\tau) e^{-i\frac{\tau\cdot u}{\sqrt{n}}} \left(1_I(\frac{\tau}{\sqrt{T}})\langle \varphi(\frac{\tau}{\sqrt{T}}, \frac{\lfloor T\rfloor - 1}{T}, \frac{\lfloor T\rfloor}{T}), \tilde{Q}(\frac{\tau}{\sqrt{T}}, T) f\rangle - \langle \nu, f\rangle \right) d\tau| \\
&\leq C \int_{\mathbb{R}^d} |\kappa_T^{\mu}(\tau) - \kappa(\tau)| d\tau \\
&\quad + C|\int_{\mathbb{R}^d} \kappa_T^{\mu}(\tau) \| 1_I(\frac{\tau}{\sqrt{T}}) \tilde{Q}^*(\frac{\tau}{\sqrt{T}}, T)\varphi(\frac{\tau}{\sqrt{T}}, \frac{\lfloor T\rfloor - 1}{T}, \frac{\lfloor T\rfloor}{T}) - \nu\| d\tau|.
\end{aligned}$$

So, by (3.38) and (3.39), $|\mathcal{A}_1 - \langle \tilde{m}_T^{u,f}, h\rangle| \to 0$, as $T \to \infty$, uniformly in $u \in \mathbb{R}^d$, $f \in \mathcal{B}_C$, and $\mu \in \mathcal{B}_p'$. Again, by Proposition 3.3.1,

$$\begin{aligned}
|\mathcal{A}_2| &\leq \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2} \int_I \frac{2M}{T} |\hat{h}(t)| \cdot |\prod_{k=0}^{\lfloor T\rfloor - 1} \lambda(t, \frac{k}{T}, \frac{k+1}{T})| dt \cdot \sup_{t\in I} \|\tilde{Q}(t,T)\| \|\mu\| \|f\| \\
&\leq C \int_{\mathbb{R}^d} \tilde{\kappa}_T(\tau) d\tau,
\end{aligned}$$

where

$$\tilde{\kappa}_T(\tau) = \frac{2M}{T}\det(\Sigma)\left(\frac{1}{2\pi}\right)^{d/2} \cdot 1_I(\frac{\tau}{\sqrt{T}}) |\hat{h}(\frac{\tau}{\sqrt{T}})| \cdot |\prod_{k=0}^{\lfloor T\rfloor - 1} \lambda(\frac{\tau}{\sqrt{T}}, \frac{k}{T}, \frac{k+1}{T})|,$$

since $\|\mu\|, \|\tilde{Q}(t,T)\| \leq 1$. By (3.35) and Lebesgue's dominated convergence theorem, we have that $\int_{\mathbb{R}^d} \tilde{\kappa}_T(\tau)d\tau \to 0$. Finally, by Lemma 3.3.3, uniformly in $u \in \mathbb{R}^d$, $f \in \mathcal{B}_C$, and $\mu \in \mathcal{B}'_p$, as $T \to \infty$,

$$
\begin{aligned}
|\mathcal{A}_3| &\leq \det(\Sigma)\left(\frac{T}{2\pi}\right)^{d/2}\int_{I^c\cap K}|\hat{h}(t)|\cdot|\langle\mu, \prod_{k=0}^{\lfloor T\rfloor-1} Q\left(t, \frac{k}{T}, \frac{k+1}{T}\right)\tilde{Q}(t,T)f\rangle|dt \\
&\leq \det(\Sigma)\cdot r_K^{\lfloor T\rfloor}\left(\frac{T}{2\pi}\right)^{d/2}\int_{I^c\cap K}|\hat{h}(t)|dt\cdot\|f\| \\
&\leq C\det(\Sigma)\cdot r_K^{\lfloor T\rfloor}\left(\frac{T}{2\pi}\right)^{d/2}\int_{I^c\cap K}|\hat{h}(t)|dt \to 0. \qquad \square
\end{aligned}
$$

The local limit theorem as in Theorem 3.1.4 with fixed (f, μ, u) follows from Lemma 3.4.1 by Theorem 10.7 in [39]. The following lemma enables us to state the uniform analogue of the result in Theorem 3.1.4. It is a multi-dimensional version of the Lemma IV.5 in [33].

Lemma 3.4.2. *Let χ be a parameter ranging over a set $\mathcal{X}$ and, for $T \geq 1$, let $m_T^\chi, \tilde{m}_T^\chi$ be positive measures on $\mathbb{R}^d$ with respect to which $\rho(x) = 1 \wedge 1/|x|^2$ is integrable. Suppose that $\sup_T \sup_{\chi\in\mathcal{X}} \tilde{m}_T^\chi(\rho(x))$ is finite, and, for each continuous integrable function h on $\mathbb{R}^d$ whose Fourier transform has compact support,*

$$\lim_{T\to\infty}\sup_{\chi\in\mathcal{X}}|\langle m_T^\chi, h\rangle - \langle\tilde{m}_T^\chi, h\rangle| = 0.$$

Then, for each compactly supported real-valued continuous function g, we have

$$\lim_{T\to\infty}\sup_{\chi\in\mathcal{X}}|\langle m_T^\chi, g\rangle - \langle\tilde{m}_T^\chi, g\rangle| = 0.$$

Proof of Theorem 3.1.4. Since, for each real-valued $f \in \mathcal{B}_C$, $f = f_+ - f_-$, where $f_+ = f \vee 0$ and $f_- = -(f \wedge 0)$, it suffices to consider all non-negative $f \in \mathcal{B}_C$. Lemma 3.4.1 verifies the assumptions in Lemma 3.4.2 with the measures defined as in (3.37) and the parameter set $\mathcal{X}$ being all $u \in \mathbb{R}^d$, non-negative $f \in \mathcal{B}_C$, and initial distribution $\mu \in \mathcal{B}_p'$. Therefore, the uniform convergence follows from the two preceding lemmas. $\square$

3.5 Non-arithmetic condition and non-lattice condition

It is clear that, under Assumptions (A1)-(A3), (A4) implies (NL). Here, we will show that they are actually equivalent under Assumptions (A1)-(A3). The proof is based on the results in [40] stated below. To apply the results, we consider a discrete Markov chain defined using X_t. Namely, we consider a Markov chain $\mathcal{X}_k$, $k \geq 1$, defined as the path of the Markov process X_t on the interval $[k-1, k]$, on the metric space $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}))$, where $\mathfrak{X}$ denotes $\mathcal{D}([0,1])$. Let $\mathfrak{P}$ denote the Markov kernel of $\mathcal{X}_k$, $\mathfrak{B}$ be the Banach space of bounded $\mathscr{B}(\mathfrak{X})$-measurable complex functions on $\mathfrak{X}$ equipped with the supremum norm: $\|F\| = \sup_{\mathfrak{x}\in\mathfrak{X}} |F(\mathfrak{x})|$. To introduce a counterpart of $Q(t,\alpha,\alpha)$, we define $I_\alpha \in \mathfrak{B}$, by $I_\alpha(\mathfrak{x}) = \int_0^1 b(\alpha, \mathfrak{x}_s)ds$, and the operator $\mathfrak{Q}(t,\alpha) : \mathfrak{B} \to \mathfrak{B}$ by

$$\mathfrak{Q}(t,\alpha)F(\mathfrak{x}) = \int_{\mathfrak{X}} \exp(it \cdot I_\alpha(\mathfrak{y}))F(\mathfrak{y})\mathfrak{P}(\mathfrak{x}, d\mathfrak{y}).$$

(i) We first prove that, for each $0 \leq \alpha \leq 1$ and $t \neq 0$, $r(\mathfrak{Q}(t,\alpha)) < 1$. The following lemma is a reformulation of the results in Theorem 3.1, Corollary 3.1, and Lemma 3.3 in [40], adapted to our situation:

Theorem 3.5.1. *If $\mathfrak{B}(\mathfrak{X})$ is countably generated and the Markov kernel $\mathfrak{P}$ satisfies the Doeblin condition, then either $r(\mathfrak{Q}(t,\alpha)) = r(\mathfrak{P}) = 1$ and $\mathfrak{Q}(t,\alpha)$ is quasi-compact, or $r(\mathfrak{Q}(t,\alpha)) < r(\mathfrak{P}) = 1$.*

It is not hard to verify $\mathfrak{B}(\mathfrak{X})$ is countably generated since $\mathfrak{X} = \mathcal{D}([0,1])$, and the Doeblin condition is satisfied for the operator $\mathfrak{P}$ given Assumptions (A1)-(A2). So, in order to prove (i), it remains to show that the first situation in the statement, i.e., where $r(\mathfrak{Q}(t,\alpha)) = r(\mathfrak{P}) = 1$ and $\mathfrak{Q}(t,\alpha)$ is quasi-compact, is not possible.

Suppose that $r(\mathfrak{Q}(t,\alpha)) = 1$ and $\mathfrak{Q}(t,\alpha)$ is quasi-compact. Then there exists $\hat{F} \in \mathfrak{B}$ with $\|\hat{F}\| = 1$ and an eigenvalue $\hat{\lambda}$ with $|\hat{\lambda}| = 1$ such that $\mathfrak{Q}(t,\alpha)\hat{F} = \hat{\lambda}\hat{F}$. Namely, for each $\mathfrak{x} \in \mathcal{D}([0,1])$,

$$\begin{aligned}\hat{\lambda}\hat{F}(\mathfrak{x}) = \mathfrak{Q}(t,\alpha)\hat{F} &= \int_{\mathfrak{X}} \exp(it \cdot I_\alpha(\mathfrak{y}))F(\mathfrak{y})\mathfrak{P}(\mathfrak{x}, d\mathfrak{y}) \\ &= \mathbf{E}_{\mathfrak{x}_1}\left[\exp\left(it \cdot \int_0^1 b(\alpha, X_s)ds\right)\hat{F}(X.)\right]\end{aligned}$$

by the Markov property of the process $\{X_s\}_{s\geq 0}$, where $X.$ stands for the path of X_s on $[0,1]$. Note that the right-hand side is a function of $\mathfrak{x}_1$. Hence, there exists $\hat{f} \in \mathcal{B}$ with $\|\hat{f}\| = 1$ such that $\hat{f}(\mathfrak{x}_1) = \hat{F}(\mathfrak{x})$, and, for each $x \in \boldsymbol{X}$,

$$\hat{\lambda}\hat{f}(x) = \mathbf{E}_x\left[\exp\left(it \cdot \int_0^1 b(\alpha, X_s)ds\right)\hat{f}(X_1)\right]. \tag{3.40}$$

For each $\varepsilon > 0$ small, we can find $x \in \boldsymbol{X}$ such that $|\hat{f}(x)| > 1-\varepsilon$ and, by Assumption (A2), we can find $N > 0$ such that $\mathrm{TV}(P(N,x,\cdot),\nu) < \varepsilon$. By applying (3.40) N times, we obtain:

$$\hat{\lambda}^N\hat{f}(x) = \mathbf{E}_x\left[\exp\left(it \cdot \int_0^N b(\alpha, X_s)ds\right)\hat{f}(X_N)\right].$$

It follows, from how we chose x and N, that

$$1-\varepsilon \leq |\hat{\lambda}^N \hat{f}(x)| \leq \mathbf{E}_x|\hat{f}(X_N)| \leq \langle \nu, |\hat{f}|\rangle + 2\varepsilon,$$

which implies that $\langle \nu, |\hat{f}|\rangle > 1-3\varepsilon$. Since this holds for arbitrary $\varepsilon > 0$ small, we conclude that $\langle \nu, |\hat{f}|\rangle = 1$, i.e., $|\hat{f}| = 1$, ν-a.s.

Let us come back to (3.40). Since $|\hat{f}| = 1$, ν-a.s., we can write $\hat{f} = \exp(iu)$, ν-a.s., where u is a real-valued bounded measurable function. Then, (3.40) implies that

$$\left|\mathbf{E}_\nu\left[\exp\left(i\left(t\cdot\int_0^1 b(\alpha, X_s)ds + u(X_1) - u(X_0)\right)\right)\right]\right| = 1,$$

which contradicts with the non-lattice condition (NL).

(ii) We now prove that, for each $0 \leq \alpha \leq 1$ and $t \neq 0$, $r(Q(t,\alpha,\alpha)) < 1$. There exists $C > 0$ such that, for each $f \in \mathcal{B}$ with $\|f\| = 1$ and $n > 0$,

$$\|Q(t,\alpha,\alpha)^n f\| = \|\mathfrak{Q}(t,\alpha)^n F\| < Cr(\mathfrak{Q}(t,\alpha))^n,$$

where F is defined by $F(\mathfrak{x}) = f(\mathfrak{x}_1)$. Therefore, $r(Q(t,\alpha,\alpha)) \leq r(\mathfrak{Q}(t,\alpha)) < 1$, and we conclude that (NL), together with (A1)-(A3), implies (A4).

3.6 Application to linear dynamical systems

In this section, let us discuss an application to the averaging principle for randomly perturbed linear dynamical systems.

Theorem 3.6.1. *Let Y_t^ε be defined by the equation:*

$$dY_t^\varepsilon = [A(t)Y_t^\varepsilon + v(t, X_{t/\varepsilon})]dt, \;\; Y_0^\varepsilon = y \in \mathbb{R}^d, \tag{3.41}$$

where $A(t)$ is a continuously differentiable $d-$dimensional square-matrix-valued function, $v : [0,\infty) \times \boldsymbol{X} \to \mathbb{R}^d$ is a continuous function, and $\{X_s\}_{s\geq 0}$ satisfies Assumptions (A1) and (A2). Let $\int_{\boldsymbol{X}} v(t,x)d\nu(x) = \bar{v}(t)$ and y_t be the averaged motion defined by:

$$dy_t = [A(t)y_t + \bar{v}(t)]dt, \;\; y_0 = y. \tag{3.42}$$

Fix $t > 0$ and suppose that $v(\alpha t, x) - \bar{v}(\alpha t)$ as a function of α and x satisfies (A3) and the non-arithmetic condition (A4) holds for $(\{X_s\}_{s\geq 0}, v(\alpha t, x) - \bar{v}(\alpha t))$. Then the local limit theorem holds for $\frac{1}{\varepsilon}(Y_t^\varepsilon - y_t)$ as $\varepsilon \to \infty$. Namely, there exists a positive-definite matrix Σ_t such that, for each compactly supported continuous real-valued function g, we have

$$\lim_{\varepsilon\to 0} |\det(\Sigma_t)(2\pi t/\varepsilon)^{d/2}\mathbf{E}_\mu[f(X_{t/\varepsilon})g(\frac{1}{\varepsilon}(Y_t^\varepsilon - y_t) - u)] - e^{-\frac{1}{2T}u^*(\Sigma_t\Sigma_t^*)^{-1}u}\langle \nu, f\rangle\langle \mathfrak{L}, g\rangle| = 0,$$

uniformly in real-valued measurable f satisfying $\|f\| \leq C$, $u \in \mathbb{R}^d$, and probability measure μ.

Proof. By taking the difference of (3.41) and (3.42) we obtain

$$\frac{1}{\varepsilon}(Y_t^\varepsilon - y_t) = \frac{1}{\varepsilon}\int_0^t \left(A(s)(Y_s^\varepsilon - y_s) + v(s, X_{s/\varepsilon}) - \bar{v}(s)\right) ds.$$

Let $U(t)$ solve the equation:

$$\begin{cases} dU(t) = A(t)U(t)dt \\ U(0) = I \end{cases}.$$

Then

$$\frac{1}{\varepsilon}(Y_t^\varepsilon - y_t) = \frac{1}{\varepsilon}\int_0^t U(t)U(s)^{-1}\left(v(s, X_{s/\varepsilon}) - \bar{v}(s)\right) ds.$$

After the change of variable, with $T = t/\varepsilon$, we obtain that

$$\frac{1}{\varepsilon}(Y_t^\varepsilon - y_t) = \int_0^T U(t)U(st/T)^{-1}\left(v(st/T, X_s) - \bar{v}(st/T)\right) ds.$$

The integral has the form of (3.4) with $b(\alpha, x) = U(t)U(\alpha t)^{-1}\left(v(\alpha t, x) - \bar{v}(\alpha t)\right)$. It is not hard to see that Assumption (A4) holds for $(\{X_s\}_{s\geq 0}, b(\alpha, x))$ since $U(\cdot)$ is non-degenerate. In addition, Assumption (A3) follows from the assumptions we have on functions A and v. Thus, the result follows from Theorem 3.1.4. □

Remark 3.6.2. *The functional central limit theorem is well-known for the fast-slow systems of the form:*

$$dZ_t^\varepsilon = c(Z_t^\varepsilon, X_{t/\varepsilon})dt. \tag{3.43}$$

Namely, under natural assumptions, Z_t^ε converges to a deterministic process z_t in probability as $\varepsilon \downarrow 0$ on each finite interval, and $\frac{1}{\sqrt{\varepsilon}}(Z_t^\varepsilon - z_t)$ converges to a Gaussian Markov process weakly ([5], [6]). The functional central limit theorem gives, in particular, the distribution of Z_t^ε in a neighborhood of z_t of size $O(\sqrt{\varepsilon})$. Theorem 3.6.1 refines this statement to spatial scales of order ε in the special case of linear systems.

In addition, it is not hard to verify that the arguments in this chapter can be applied to prove the following slightly stronger version of the local limit theorem:

Theorem 3.6.3. *Let assumptions (A1)-(A4) be satisfied. For $0 \le \rho < 1$, define $S(\rho,T) = \int_0^{(1-\rho)T} b(\rho + s/T, X_s)ds$. Then there exists a positive-definite matrix Σ_ρ that is continuous in ρ such that, for each compactly supported function $g \in \mathcal{C}(\mathbb{R}^d,\mathbb{R})$, $C > 0$, and $0 \le \rho_0 < 1$, uniformly in real-valued $f \in \mathcal{B}_C$, initial distribution $\mu \in \mathcal{B}'_p$, $0 \le \rho \le \rho_0$, and $u \in \mathbb{R}^d$,*

$$\lim_{T\to\infty} \left| \det(\Sigma_\rho)(2\pi T)^{d/2} \mathbf{E}_\mu[f(X_{T(1-\rho)})g(S(\rho,T)-u)] - e^{-\frac{1}{2T}u^*(\Sigma_\rho\Sigma_\rho^*)^{-1}u}\langle \nu, f\rangle\langle \mathfrak{L}, g\rangle \right| = 0.$$

Remark 3.6.4. *The formulation of the result in Theorem 3.6.3 might seem strange at first. However, it can be useful, for example, when considering the distribution of the process Y_t^ε in Theorem 3.6.1 starting from different points on the trajectory of y_t. Such result is used in [2] in order to establish the regularity of the density of fast-slow systems of the form* (3.43) *starting from different points on z_t (see also Theorem 2.5.8). Some of the steps in the proof of regularity in [2] involve a linearization of* (3.43)*, which yields a process of the form* (3.42)*, whose distribution can be described using Theorem 3.6.1 (see also **Step 2** in Section 2.5).*

Chapter 4: Large deviations for Hamiltonian systems on intermediate time scales

4.1 Introduction

Consider the Hamiltonian dynamical system in $\mathbb{R}^2$ defined by an ordinary differential equation:

$$dx_t = v(x_t)dt,\ x_0 \in \mathbb{R}^2, \tag{4.1}$$

where

$$v(x) = \nabla^{\perp} H(x) := \left(-\frac{\partial H(p,q)}{\partial q}, \frac{\partial H(p,q)}{\partial p}\right),$$

and Hamiltonian H is smooth enough.

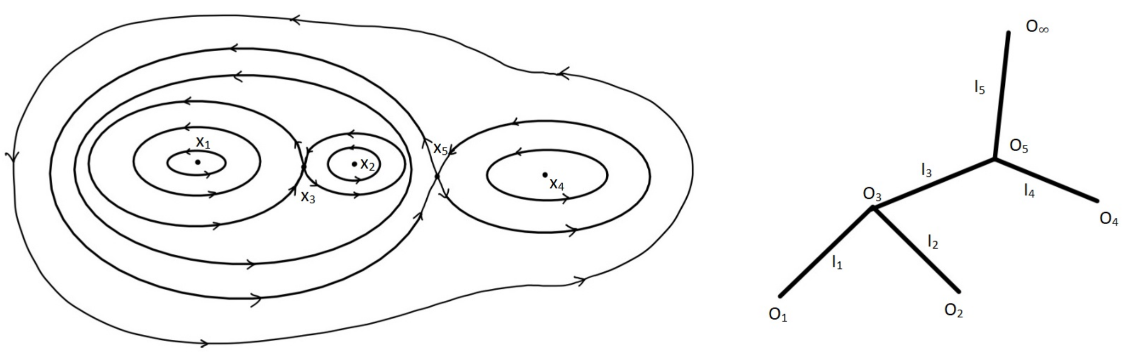


Figure 4.1: A typical example of a Hamiltonian system and the corresponding Reeb graph.

Small random perturbations of Hamiltonian systems have been studied extensively from different perspectives. In particular, let ε be a small positive parameter, and $\tilde{X}^\varepsilon$ be a diffusion process in $\mathbb{R}^2$ defined by the equation:

$$d\tilde{X}_t^\varepsilon = \nabla^\perp H(\tilde{X}_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(\tilde{X}_t^\varepsilon)d\tilde{W}_t, \qquad \tilde{X}_0^\varepsilon = x_0 \in \mathbb{R}^2, \tag{4.2}$$

where H is a smooth function from $\mathbb{R}^2$ to $\mathbb{R}$ with bounded second derivatives; $\sigma(x)$ is a smooth $2 \times l$ matrix-valued function such that $\sigma(x)\sigma^*(x)$ is uniformly positive-definite and bounded; $\tilde{W}_t$ is a standard $l-$dimensional Wiener process. On time scales of order one, the large deviation principle, known as Freidlin-Wentzell theory, established in [41], describes the probability of the event that a realization of $\tilde{X}_t^\varepsilon$ belongs to a neighborhood of a path that is different from the trajectory of the deterministic flow (4.1). On time scales of order $1/\varepsilon$, the averaging principle tells us that the projection of the process $\tilde{X}_t^\varepsilon$ onto the Reeb graph behaves as a strong Markov process. In particular, on each edge of the graph, $H(\tilde{X}_t^\varepsilon)$ behaves as a diffusion process ([42]). On intermediate time scales, i.e., when time is of order $|\log\varepsilon|\varepsilon^{\beta-1}$, a result of the averaging type can also be established ([16]). Namely the time-changed process converges to a motion on a rescaled graph, provided it starts on or near a level curve containing a saddle point of H. In this chapter, we also study the intermediate time scales, $t \sim \varepsilon^{\beta-1}$, however, without spatial rescaling. Our main result states that, after projection onto the graph, the process satisfies the large deviation principle on time scales of order $\varepsilon^{\beta-1}$. The action functional can be described in terms of the coefficients of the averaged process on the graph. In the particular case $\beta = 1/2$, estimates on transition probability for the process inside the edges and estimates for the exit time from a neighborhood of an interior vertex were given in [43]. (The main focus of the latter paper was on the propagation of the reaction front in the KPP equation with the underlying diffusion given by (4.2).) In this chapter, we allow arbitrary $\beta \in (0, 1)$, which requires a more complicated approach.

Since we are interested in the behavior of $\tilde{X}_t^\varepsilon$ with $t \sim \varepsilon^{\beta-1}$, it is convenient to rescale the time by defining $X_t^\varepsilon = \tilde{X}^\varepsilon_{t\varepsilon^{\beta-1}}$. Then we have the equation for X_t^ε:

$$dX_t^\varepsilon = \varepsilon^{\beta-1}\nabla^\perp H(X_t^\varepsilon)dt + \varepsilon^{\beta/2}\sigma(X_t^\varepsilon)dW_t, \qquad X_0^\varepsilon = x_0,$$

where W_t is an $l-$dimensional Wiener process. By Ito's formula applied to $H(X_t^\varepsilon)$,

$$H(X_t^\varepsilon) = H(x_0) + \varepsilon^\beta \int_0^t AH(X_s^\varepsilon)ds + \varepsilon^{\beta/2}\int_0^t \nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)dW_s, \tag{4.3}$$

where A is the operator $Au(x) = \frac{1}{2}\sum_{i,j}[\sigma(x)\sigma^*(x)]_{i,j}\cdot\frac{\partial^2}{\partial x_i \partial x_j}u(x)$. Note that $AH(x)$ is uniformly bounded, so, due to the ε^β factor, the second term on the right-hand side does not contribute to the action functional. However, this term is important to understand the behavior of the process near the critical points of H, where the integrand in the last term in (4.3) vanishes. In order to understand large deviations, we should study how the coefficients are averaged along the trajectories of the perturbed process.

Let Γ denote the Reeb graph corresponding to the Hamiltonian H, and let $Y : \mathbb{R}^2 \to \Gamma$ be the projection on the graph. Let $\mathbf{x}_1, ..., \mathbf{x}_N$ be all the critical points of H, and $O_k = Y(\mathbf{x}_k)$, $k = 1, ..., N$. Let the edges of Γ be labeled as $I_1, ..., I_{N'}$ and a symbol $\sim$ between a vertex and an edge means that the vertex is an endpoint of the edge. Then define:

- D_i is the set of all points $x \in \mathbb{R}^2$ such that $Y(x)$ belongs to the interior of I_i;
- $C_k = \{x : Y(x) = O_k\}$ is the extremum point $\mathbf{x}_k$ or the separatrix containing $\mathbf{x}_k$;
- $C_i(H) = \{x \in D_i : H(x) = H\}$ is one of the connected components of the level set of H;

- $D_i(H_1, H_2) = \{x \in D_i : H_1 < H(x) < H_2\}$, provided that $H_1 < H_2$, is the set between $C_i(H_1)$ and $C_i(H_2)$;

- $D_k(\pm\delta)$ is the connected component of the set $\{H(\mathbf{x}_k) - \delta < x < H(\mathbf{x}_k) + \delta\}$ containing C_k;

- $D(\pm\delta) = \bigcup_k D_k(\pm\delta)$;

- $C_{ki}(\delta) = \{x \in D_i : H(x) = H(\mathbf{x}_k) \pm \delta\}$, for $O_k \sim I_i$;

- $T_i(H) = \oint_{C_i(H)} \frac{1}{|\nabla H(x)|} dl$, for applicable H, is the rotation time of system (4.1);

- $B_i^2(H) = \frac{1}{T_i(H)} \oint_{C_i(H)} \frac{|\nabla H(x)^* \sigma(x)|^2}{|\nabla H(x)|} dl$ for H in $H(I_i)$; $B_i^2(H) = 0$ for $H = H(\mathbf{x}_k)$ and $O_k \sim I_i$, is the "averaged" diffusion coefficient w.r.t. the invariant measure on $C_i(H)$.

It's worth noting that $T_i(H)$ and $B_i(H)$ have nice regularity properties if H is smooth enough. The following lemma is a direct application of Lemma 8.1.1 from [42].

Lemma 4.1.1. *$T_i(H)$ and $B_i^2(H)$ are $k - 1$ times continuously differentiable at the interior points of the interval $H(I_i)$ if H is k times continuously differentiable. (This implies the Lipschitz continuity of $T_i(H)$, $B_i^2(H)$ and $T_i(H)B_i^2(H)$ on any closed interval $I \subset H(I_i)$ if $k \geq 2$.)*

In Section 2, we state the necessary assumptions and formulate the main result. The proof comes in the sections that follow. The lower and upper bounds of the local large deviation principle are proved in Sections 3 and 4, respectively. Then, exponential tightness is justified in Section 5. These arguments, together with Puhalskii's Theorem (cf. [44] and [45]), imply the main result. Some technical proofs are given in the Section 4.7.

4.2 Main results

Our main result concerns the large deviation principle for the process $Y(\tilde{X}^{\varepsilon}_{t\varepsilon^{\beta-1}})$. Large deviations from averaged behavior are described in terms of the action functional, which describes the probability that the trajectories of the process are close to a given deterministic trajectory. More precisely, we give the formal definition of an action functional ([42]):

Definition 4.2.1. *Let (X, ρ) be a metric space. Let μ^{ε}, $\varepsilon > 0$, be a family of probability measures on $\mathcal{B}(X)$, $\lambda(\varepsilon)$ be a positive real-valued function going to $+\infty$ as $\varepsilon \downarrow 0$, and $S(x)$ be a function on X with values in $[0, \infty]$. It is said that $\lambda(\varepsilon)S(x)$ is an action function (action functional if X is a function space) for μ^{ε} as $\varepsilon \downarrow 0$ if:*

(0) the set $\Phi(s) = \{x : S(x) \leq s\}$ is compact for every $s \geq 0$;

(1) for each $\delta > 0$, each $\gamma > 0$, and each $x \in X$, there exists $\varepsilon_0 > 0$ such that

$$\mu^{\varepsilon}\{y : \rho(x, y) < \delta\} \geq \exp\{-\lambda(\varepsilon)[S(x) + \gamma]\}$$

for all $\varepsilon \leq \varepsilon_0$;

(2) for each $\delta > 0$, each $\gamma > 0$, and each $s > 0$, there exists $\varepsilon_0 > 0$ such that

$$\mu^{\varepsilon}\{y : \rho(y, \Phi(s)) \geq \delta\} \leq \exp\{-\lambda(\varepsilon)(s - \gamma)\}$$

for all $\varepsilon \leq \varepsilon_0$.

Consider the following metric on Γ: $r(y_1, y_2)$ is the length of the shortest path connecting

y_1 and y_2. For example, if $y_1 = (1, H_1)$, $I_1 \sim O_1$, $O_1 \sim I_2$, $I_2 \sim O_2$, $O_2 \sim I_3$ and $y_2 = (3, H_2)$, then $r(y_1, y_2) = |H_1 - H(\mathbf{x}_1)| + |H(\mathbf{x}_1) - H(\mathbf{x}_2)| + |H(\mathbf{x}_2) - H_2|$. Based on this metric, we introduce the uniform metric on $\mathbf{C}([0,T], \Gamma)$: $\rho_{0,T}(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \sup_{0 \le t \le T} r(\boldsymbol{\varphi}(t), \boldsymbol{\psi}(t))$. In the metric space $(\mathbf{C}([0,T], \Gamma), \rho_{[0,T]})$, we consider the probability measures induced by the family of process $Y(\tilde{X}^{\varepsilon}_{t\varepsilon^{\beta-1}})$. We use the conventions $0/0 = 0$, $\sup\{\emptyset\} = -\infty$, and $\inf\{\emptyset\} = +\infty$. A trajectory $\boldsymbol{\varphi}$ on Γ can be written in terms of its components as (i_t, φ_t). Define the functional $S(\boldsymbol{\varphi})$ on $\mathbf{C}([0,T], \Gamma)$:

$$S(\boldsymbol{\varphi}) = \frac{1}{2} \int_0^T \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt$$

for $\varphi \in \mathbf{C}([0,T], \mathbb{R})$ that is absolutely continuous and satisfies $\dot{\varphi}_t / B_{i_t}(\varphi_t) \in L^2([0,T])$; otherwise $S(\boldsymbol{\varphi}) = \infty$. Note that, if $\boldsymbol{\varphi}_t = O_k$ for some interior vertex O_k, then we will have different representations of $\boldsymbol{\varphi}_t$ since it's on multiple edges at the same time. However, by the definition of B^2, there's no ambiguity because, according to our conventions, the integrand is defined as 0 at such vertices (see Lemma 4.7.1). The following conditions are assumed to hold throughout the chapter:

1. The Hamiltonian $H(x)$, $x \in \mathbb{R}^2$, is four times continuously differentiable with bounded second derivatives.

2. For sufficiently large $|x|$, there exist positive A_1, A_2 and A_3, such that $H(x) \ge A_1|x|^2$, $|\nabla H(x)| \ge A_2|x|$, and $\Delta H(x) \ge A_3$.

3. $H(x)$ has a finite number of critical points, and the Hessian matrix is non-degenerate at those points.

4. Each level curve corresponding to a vertex on the Reeb graph contains at most one critical

point.

5. $\sigma(x)$ is continuously differentiable with bounded derivatives and $\sigma(x)\sigma^*(x)$ is bounded and uniformly positive-definite.

Now we are ready to formulate the main result:

Theorem 4.2.2. *Let the assumptions above be satisfied and let the process $\tilde{X}_t^\varepsilon$ be defined as in* (4.2). *Then $\varepsilon^{-\beta}S(\boldsymbol{\varphi})$ is the action functional of the family of process $Y(\tilde{X}^\varepsilon_{t\varepsilon^{\beta-1}})$ in the sense of the uniform metric on $\boldsymbol{C}([0,T],\Gamma)$ restricted to the set of functions that start at $Y(x_0)$.*

4.3 Lower bound for the large deviation principle

Let us prove the lower bound for the large deviation principle. First, we look at a process that starts and evolves inside a single edge. Next, we deal with a process that starts at an exterior vertex but stays within an edge. Finally, we estimate the probability for a process to travel through an interior vertex.

4.3.1 The case where the process evolves inside one edge

In this section, the process evolves only inside one edge (for example I_i). Therefore, we simplify the notations when there is no ambiguity: $T_i(H) \to T(H)$; $B_i^2(H) \to B^2(H)$; $S(\boldsymbol{\varphi}) = S(i,\varphi) \to S(\varphi)$, etc. The following lemma tells us that if the process starts sufficiently near the initial point of the deterministic trajectory, it has a good chance to follow the trajectory and arrive in a tiny neighborhood of where the trajectory ends.

Lemma 4.3.1. *For every $\delta > 0$, $\gamma > 0$, and a trajectory $\boldsymbol{\varphi}_t = (i, \varphi_t) \in \boldsymbol{C}([0,T], I_i)$ such that $\inf_{t\in[0,T]} r(\varphi_t, H(O)) > 0$ for each $O \sim I_i$, there exists $h > 0$ such that for every $h' > 0$ there* *exists* $\varepsilon_0 > 0$ *such that for every initial point* $X_0^\varepsilon = x$ *satisfying* $|H(x) - \varphi_0| < h$,

$$\mathbf{P}(|H(X_T^\varepsilon) - \varphi_T| < h',\ \rho_{0,T}(H(X_t^\varepsilon), \varphi_t) < \delta) \geq \exp\left[-\varepsilon^{-\beta}(S(\varphi) + \gamma)\right],$$

for all $\varepsilon < \varepsilon_0$.

Proof. We present the proof in several steps. **i.** We can assume that $h' \leq h$ and that φ is absolutely continuous with $\varphi' \in L^2([0,T])$ because $S(\varphi) = \infty$ for any other φ. Also, without loss of generality, we assume that $T = 1$, $O_j \sim I$, $O_k \sim I$, $H(\mathbf{x}_j) < H(\mathbf{x}_k)$, and $H(\mathbf{x}_j) = 0$. We can assume that $0 < \tilde{m} \leq \varphi \leq \tilde{M}$ and $\delta < \frac{\tilde{m}}{2} \wedge \frac{H(\mathbf{x}_k) - \tilde{M}}{2}$. Let $C = \overline{D_i(\tilde{m} - \delta, \tilde{M} + \delta)}$, which is a compact set in $\mathbb{R}^2$. Then there exist $m > 0$, $M > 0$, $L > 0$ such that:

(1) $|AH(x)| + |\sigma(x)| < M,\ \forall x \in \mathbb{R}^2;\ |\nabla H(x)^* \sigma(x)|^2 < M,\ \forall x \in C$;

(2) $|\nabla H(x)^* \sigma(x)|^2 > m,\ \forall x \in C$;

(3) $m < T_i(H) < M$ and $m < B_i^2(H) < M,\ \forall H \in \{H : r(H, Ran(\varphi)) < \delta\}$;

(4) $|\nabla H(x) - \nabla H(y)| + ||\nabla H(x)^* \sigma(x)|^2 - |\nabla H(y)^* \sigma(y)|^2| < L|x - y|,\ \forall x, y \in C$;

(5) $|T_i(H_1) - T_i(H_2)| + |T_i(H_1) B_i^2(H_1) - T_i(H_2) B_i^2(H_2)| < L|H_1 - H_2|,\ \forall H_1, H_2 \in \{H : r(H, Ran(\varphi)) < \delta\}$ (see Lemma 4.1.1).

With m, M, L selected, we choose the parameters as follows:

(1) $\delta' > 0$ such that $\delta' < \delta$, $\alpha := \frac{8L(M+1)}{m^2}\delta' < 1/2$ and $\frac{\alpha}{1-\alpha}\int_0^1 \frac{|\varphi_t'|^2}{B^2(\varphi)} dt < \gamma$;

(2) $r > 0$ such that $r < \frac{m}{24M}$;

(3) $n \in \mathbb{N}$ such that $a, b \in [0, T]$ with $|a - b| < 1/n$ implies $|\varphi_a - \varphi_b| < r\delta' < \frac{1}{12}\delta'$;

(4) $h > 0$ such that $h < r\delta'$ and $\frac{\alpha}{1-\alpha}\int_0^1 \frac{|\varphi_t'|^2}{B^2(\varphi)}dt + \frac{1}{1-\alpha}\cdot\frac{n^2 h\delta'}{m^2} < \gamma$

ii. Now let us study how X_t^ε behaves on one interval $[\frac{k}{n}, \frac{k+1}{n}]$. To start with, X_t^ε satisfies the equation:

$$dX_t^\varepsilon = \varepsilon^{\beta-1}\nabla^\perp H(X_t^\varepsilon) + \varepsilon^{\beta/2}\sigma(X_t^\varepsilon)dW_t, \qquad X_0^\varepsilon = x.$$

For every $0 \le k < n$, define A_k as the event where the process is not perturbed significantly by the diffusion during the time required to make one rotation (which will allow us to use averaging) and does not travel far from the original level curve. More precisely,

$$A_k = \bigcap_{p=0}^{\lfloor \frac{1}{nT_k^\varepsilon} \rfloor} \left\{ \sup_{pT_k^\varepsilon + \frac{k}{n} \le t \le (p+1)T_k^\varepsilon + \frac{k}{n}} \varepsilon^{\beta/2} \left| \int_{k/n + pT_k^\varepsilon}^t \sigma(X_s^\varepsilon) dW_s \right| \le M \cdot \varepsilon^{\frac{1-\beta}{4}} \right\}$$
$$\bigcap \left\{ \sup_{0 \le t \le \frac{1}{n}} \left| H(X^\varepsilon_{\frac{k}{n}+t}) - H(X^\varepsilon_{\frac{k}{n}}) \right| \le \frac{\delta'}{2} \right\},$$

where $T_k^\varepsilon = \varepsilon^{1-\beta} T(H(X^\varepsilon_{\frac{k}{n}}))$. By Lemma 4.7.5, for the initial point $X_k^\varepsilon = x$ satisfying $|H(x) - \varphi_{\frac{k}{n}}| < h$,

$$\mathbf{P}\left(\Omega \setminus A_k\right) \le \exp\left(-\frac{n\delta'^2}{64M\varepsilon^\beta}\right) =: \lambda_3(\varepsilon).$$

For each $0 \le p \le \lfloor \frac{1}{nT_k^\varepsilon} \rfloor$ and $t \ge \frac{k}{n} + pT_k^\varepsilon$, define a deterministic process ξ_t^ε with random initial data by the equation:

$$d\xi_t^\varepsilon = \varepsilon^{\beta-1}\nabla^\perp H(\xi_t^\varepsilon)dt, \quad \xi^\varepsilon_{\frac{k}{n}+pT_k^\varepsilon} = X^\varepsilon_{\frac{k}{n}+pT_k^\varepsilon}.$$

On the event $A_k \cap \left\{ |\varphi_{\frac{k}{n}} - H(X^\varepsilon_{\frac{k}{n}})| < h \right\} \subset \{X_t^\varepsilon, \xi_t^\varepsilon \in C,\ t \in [\frac{k}{n}, \frac{k+1}{n}]\}$, estimate the difference

$|X_t^\varepsilon - \xi_t^\varepsilon|$ on the interval $[\frac{k}{n} + pT_k^\varepsilon, \frac{k}{n} + (p+1)T_k^\varepsilon]$:

$$\begin{aligned}
|X_t^\varepsilon - \xi_t^\varepsilon| &\leq \varepsilon^{\beta-1} \int_{k/n+pT_k^\varepsilon}^t \left|\nabla H^\perp(X_s^\varepsilon) - \nabla H^\perp(\xi_s^\varepsilon)\right| ds + \varepsilon^{\beta/2} \left| \int_{k/n+pT_k^\varepsilon}^t \sigma(X_s^\varepsilon) dW_s \right| \\
&\leq \varepsilon^{\beta-1} \int_{k/n+pT_k^\varepsilon}^t L\,|X_s^\varepsilon - \xi_s^\varepsilon|\, ds + M \cdot \varepsilon^{\frac{1-\beta}{4}}.
\end{aligned}$$

So, for ε small enough, by the Gronwall's inequality,

$$|X_t^\varepsilon - \xi_t^\varepsilon| \leq M \exp(L \cdot T(H(X_{\frac{k}{n}}^\varepsilon))) \cdot \varepsilon^{\frac{1-\beta}{4}} \leq M \exp(LM) \cdot \varepsilon^{\frac{1-\beta}{4}} \leq \frac{\delta'}{M}. \tag{4.4}$$

We approximate $|\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2$ with $B^2(\varphi_s)$ on the interval $[\frac{k}{n} + pT_k^\varepsilon, \frac{k}{n} + (p+1)T_k^\varepsilon]$, still on the event $A_k \cap \left\{ |\varphi_{\frac{k}{n}} - H(X_{\frac{k}{n}}^\varepsilon)| < h \right\}$:

$$\begin{aligned}
&\int_{\frac{k}{n}+pT_k^\varepsilon}^{\frac{k}{n}+(p+1)T_k^\varepsilon} |\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds \\
\leq &\int_{\frac{k}{n}+pT_k^\varepsilon}^{\frac{k}{n}+(p+1)T_k^\varepsilon} |\nabla H(\xi_s^\varepsilon)^* \sigma(\xi_s^\varepsilon)|^2 ds + LT_k^\varepsilon \cdot \frac{\delta'}{M} \\
\leq &\int_{\frac{k}{n}+pT_k^\varepsilon}^{\frac{k}{n}+pT_k^\varepsilon+\varepsilon^{1-\beta}T(H(X^\varepsilon_{\frac{k}{n}+pT_k^\varepsilon}))} |\nabla H(\xi_s^\varepsilon)^* \sigma(\xi_s^\varepsilon)|^2 ds + \varepsilon^{1-\beta} \cdot L\delta' M + \varepsilon^{1-\beta} T(H(X_{\frac{k}{n}}^\varepsilon)) L \cdot \frac{\delta'}{M} \\
\leq &\,\varepsilon^{1-\beta} T(H(X^\varepsilon_{\frac{k}{n}+pT_k^\varepsilon})) B^2(H(X^\varepsilon_{\frac{k}{n}+pT_k^\varepsilon})) + \varepsilon^{1-\beta} \cdot L\delta'(M+1) \\
\leq &\,\varepsilon^{1-\beta} T(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) B^2(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) + \varepsilon^{1-\beta} \cdot L\delta'(M+2) \\
\leq &\,\varepsilon^{1-\beta} T(H(X_{\frac{k}{n}}^\varepsilon)) B^2(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) + \varepsilon^{1-\beta} \cdot L\delta'(2M+2) \\
= &\,T_k^\varepsilon B^2(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) + 2\varepsilon^{1-\beta} \cdot L\delta'(M+1),
\end{aligned}$$

where the first inequality is by (4.4), the second inequality is by Lipschitz continuity of $T(H)$ and boundedness of $|\nabla H(\xi_s^\varepsilon)^* \sigma(\xi_s^\varepsilon)|$, and the fourth inequality is by the Lipschitz continuity of

$T(H)B^2(H)$.

With the same reasoning, we get an estimate in the other direction:

$$\int_{\frac{k}{n}+pT_k^\varepsilon}^{\frac{k}{n}+(p+1)T_k^\varepsilon} |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds \geq T_k^\varepsilon B^2(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) - 2\varepsilon^{1-\beta} \cdot L\delta'(M+1).$$

Note that these inequalities are valid for every $0 \leq p \leq \lfloor \frac{1}{nT_k^\varepsilon} \rfloor$. Therefore, on the event $A_k \cap \left\{|\varphi_{\frac{k}{n}} - H(X_{\frac{k}{n}}^\varepsilon)| < h\right\}$, $\int_{\frac{k}{n}}^{\frac{k+1}{n}} |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2$ can also be approximated by $\int_{\frac{k}{n}}^{\frac{k+1}{n}} B^2(\varphi_s)$ on the interval $[\frac{k}{n}, \frac{k+1}{n}]$:

$$\begin{aligned}
&\int_{\frac{k}{n}}^{\frac{k+1}{n}} |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds \\
\leq\ & \sum_{p=0}^{\lfloor \frac{1}{nT_k^\varepsilon} \rfloor - 1} \int_{\frac{k}{n}+pT_k^\varepsilon}^{\frac{k}{n}+(p+1)T_k^\varepsilon} |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds + T_k^\varepsilon M \\
\leq\ & \sum_{p=0}^{\lfloor \frac{1}{nT_k^\varepsilon} \rfloor - 1} T_k^\varepsilon B^2(\varphi_{\frac{k}{n}+pT_k^\varepsilon}) + \frac{1}{nT_k^\varepsilon} \cdot 2\varepsilon^{1-\beta} \cdot L\delta'(M+1) + T_k^\varepsilon M \\
\leq\ & (1+\alpha/2) \int_{\frac{k}{n}}^{\frac{k+1}{n}} B^2(\varphi_s) ds + \frac{4L(M+1)}{mn} \cdot \delta' \\
\leq\ & (1+\alpha) \int_{\frac{k}{n}}^{\frac{k+1}{n}} B^2(\varphi_s) ds,
\end{aligned}$$

for ε sufficiently small. And similarly,

$$\int_{\frac{k}{n}}^{\frac{k+1}{n}} |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds \geq (1-\alpha) \int_{\frac{k}{n}}^{\frac{k+1}{n}} B^2(\varphi_s) ds.$$

iii. We proceed to estimate the conditional probability that $H(X_t^\varepsilon)$ follows φ_t on the interval $[\frac{k}{n}, \frac{k+1}{n}]$, given that $H(X_{\frac{k}{n}}^\varepsilon)$ is close enough to $\varphi_{\frac{k}{n}}$. Let $c_k = \int_{\frac{k}{n}}^{\frac{k+1}{n}} B^2(\varphi_t) dt$, $s_0 = (1-\alpha)c_k\varepsilon^\beta$,

$s_1 = (1+\alpha)c_k\varepsilon^\beta$, $P_k = \varphi_{\frac{k+1}{n}} - H(X^\varepsilon_{\frac{k}{n}})$, and

$$Z_t = \varepsilon^{\beta/2}\int_{\frac{k}{n}}^{\frac{k}{n}+t} \nabla H(X^\varepsilon_s)^*\sigma(X^\varepsilon_s)dW_s = \tilde{W}(\langle Z\rangle_t),$$

where $\langle Z\rangle_t = \varepsilon^\beta\int_{\frac{k}{n}}^{\frac{k}{n}+t}|\nabla H(X^\varepsilon_s)^*\sigma(X^\varepsilon_s)|^2 ds$ and $\tilde{W}$ is a one-dimensional Wiener process. Earlier, we saw that $s_0 \le \langle Z\rangle_{\frac{1}{n}} \le s_1$ on the event $A_k \cap \left\{|\varphi_{\frac{k}{n}} - H(X^\varepsilon_{\frac{k}{n}})| < h\right\}$. On the other hand, the increment in $H(X^\varepsilon_t)$ is almost $Z_{1/n}$, since for ε small enough

$$\left|H(X^\varepsilon_{\frac{k+1}{n}}) - H(X^\varepsilon_{\frac{k}{n}}) - Z_{1/n}\right| = \varepsilon^\beta\left|\int_{\frac{k}{n}}^{\frac{k+1}{n}} AH(X^\varepsilon_s)ds\right| \le \varepsilon^\beta\frac{M}{n} < h'/2.$$

Then, for the initial point $X^\varepsilon_{\frac{k}{n}} = x$ satisfying $|H(x) - \varphi_{\frac{k}{n}}| < h$,

$$\begin{aligned}
&\mathbf{P}\left(\left\{|\varphi_{\frac{k+1}{n}} - H(X^\varepsilon_{\frac{k+1}{n}})| < h'\right\}\cap A_k\right)\\
&= \mathbf{P}\left(\left\{|H(X^\varepsilon_{\frac{k+1}{n}}) - H(X^\varepsilon_{\frac{k}{n}}) - P_k| < h'\right\}\cap A_k\right)\\
&\ge \mathbf{P}\left(\left\{\left|Z_{1/n} - P_k\right| < \frac{h'}{2}\right\}\cap A_k\right)\\
&= \mathbf{P}\left(\left\{\left|\tilde{W}(\langle Z\rangle_{\frac{1}{n}}) - P_k\right| < \frac{h'}{2}\right\}\cap\left\{s_0 \le \langle Z\rangle_{\frac{1}{n}} \le s_1\right\}\cap A_k\right)\\
&\ge \mathbf{P}\left(\left\{\sup_{s_0\le s\le s_1}|\tilde{W}_s - P_k| < \frac{h'}{2}\right\}\cap A_k\right)\\
&\ge \mathbf{P}\left(\sup_{s_0\le s\le s_1}|\tilde{W}_s - P_k| < \frac{h'}{2}\right) - \lambda_3(\varepsilon)\\
&\ge \mathbf{P}\left(|\tilde{W}_{s_0} - P_k| < \frac{h'}{4}, \sup_{s_0\le s\le s_1}|\tilde{W}_s - \tilde{W}_{s_0}| < \frac{h'}{4}\right) - \lambda_3(\varepsilon)\\
&= \frac{1}{\sqrt{2\pi s_0}}\int_{P_k - h'/4}^{P_k + h'/4}\exp(-\frac{y^2}{2s_0})dy\cdot\mathbf{P}\left(\sup_{s_0\le s\le s_1}|\tilde{W}_s - \tilde{W}_{s_0}| < \frac{h'}{4}\right) - \lambda_3(\varepsilon)\\
&\ge \frac{h'}{2}\frac{1}{\sqrt{2\pi s_0}}\exp\left(-\frac{1}{2s_0}\left(|\varphi_{\frac{k+1}{n}} - \varphi_{\frac{k}{n}}| + 2h\right)^2\right)\cdot\frac{1}{2} - \lambda_3(\varepsilon)
\end{aligned}$$

$$\geq 2\exp\left(-\frac{\left(|\varphi_{\frac{k+1}{n}}-\varphi_{\frac{k}{n}}|+2h\right)^2}{2(1-\alpha)\varepsilon^{\beta}\int_{\frac{k}{n}}^{\frac{k+1}{n}}B^2(\varphi_t)dt}\right)-\exp\left(-\frac{n\delta'^2}{64M\varepsilon^{\beta}}\right)$$

$$\geq \exp\left(-\frac{\left(|\varphi_{\frac{k+1}{n}}-\varphi_{\frac{k}{n}}|+2h\right)^2}{2(1-\alpha)\varepsilon^{\beta}\int_{\frac{k}{n}}^{\frac{k+1}{n}}B^2(\varphi_t)dt}\right).$$

iv. Since the previous result is valid for every $0 \leq k < n$,

$$\begin{aligned}
&\mathbf{P}\left(|H(X_1^{\varepsilon})-\varphi_1|<h',\ \rho_{0,1}(H(X_t^{\varepsilon}),\varphi_t)<\delta\right)\\
&\geq \mathbf{P}\left(|H(X_1^{\varepsilon})-\varphi_1|<h',\ \sup_{0\leq t\leq 1}|H(X_t^{\varepsilon})-\varphi_t|<\delta'\right)\\
&\geq \mathbf{P}\left(\bigcap_{k=0}^{n-1}A_k\cap\left\{|H(X_{\frac{k}{n}}^{\varepsilon})-\varphi_{\frac{k}{n}}|<h' \text{ for every } 1\leq k\leq n\right\}\right)\\
&\geq \prod_{k=0}^{n-1}\inf_{y:|H(y)-\varphi_{\frac{k}{n}}|<h}\mathbf{P}\left(\left\{|\varphi_{\frac{k+1}{n}}-H(X_{\frac{k+1}{n}}^{\varepsilon})|<h'\right\}\cap A_k|X_{\frac{k}{n}}^{\varepsilon}=y\right)\\
&\geq \prod_{k=0}^{n-1}\exp\left(-\frac{\left(|\varphi_{\frac{k+1}{n}}-\varphi_{\frac{k}{n}}|+2h\right)^2}{2(1-\alpha)\varepsilon^{\beta}\int_{\frac{k}{n}}^{\frac{k+1}{n}}B^2(\varphi_t)dt}\right)\\
&= \exp\left(-\frac{1}{2\varepsilon^{\beta}(1-\alpha)}\sum_{k=0}^{n-1}\left(\frac{\left|\int_{\frac{k}{n}}^{\frac{k+1}{n}}\dot{\varphi}_t dt\right|^2}{\int_{\frac{k}{n}}^{\frac{k+1}{n}}B^2(\varphi_t)dt}+4h\frac{h+\left|\int_{\frac{k}{n}}^{\frac{k+1}{n}}\dot{\varphi}_t dt\right|}{\int_{\frac{k}{n}}^{\frac{k+1}{n}}B^2(\varphi_t)dt}\right)\right)\\
&\geq \exp\left(-\frac{1}{2\varepsilon^{\beta}(1-\alpha)}\sum_{k=0}^{n-1}\left(\int_{\frac{k}{n}}^{\frac{k+1}{n}}\frac{|\dot{\varphi}_t|^2}{B^2(\varphi_t)}dt+\frac{nh\delta'}{m}\right)\right)\\
&\geq \exp\left(-\frac{1}{2\varepsilon^{\beta}}\left(\frac{1}{1-\alpha}\int_0^1\frac{|\dot{\varphi}_t|^2}{B^2(\varphi_t)}dt+\frac{1}{1-\alpha}\cdot\frac{n^2h\delta'}{m}\right)\right)\\
&\geq \exp\left(-\frac{1}{2\varepsilon^{\beta}}\left(\int_0^1\frac{|\dot{\varphi}_t|^2}{B^2(\varphi_t)}dt+\frac{\alpha}{1-\alpha}\int_0^1\frac{|\dot{\varphi}_t|^2}{B^2(\varphi_t)}dt+\frac{1}{1-\alpha}\cdot\frac{n^2h\delta'}{m}\right)\right)\\
&\geq \exp\left(-\varepsilon^{-\beta}\left(S(\varphi)+\gamma\right)\right). \qquad \square
\end{aligned}$$

4.3.2 The case where the process starts at an exterior vertex

In this section, we assume that $\mathbf{x}_0 = (0,0)$ (the origin in $\mathbb{R}^2$) is a local minimum point and $H(\mathbf{x}_0) = 0$, without loss of generality. We aim to estimate from below the probability that the random process starting at an exterior vertex escapes from a certain neighborhood of the vertex sufficiently fast. We start with the following lemma. Again, we simplify the notations as what we did in the preceding section.

Lemma 4.3.2. *Suppose that $X_0^\varepsilon = \mathbf{x}_0$. Then, for every $T > 0$, there exists a positive number k such that the stopping time $\tau_1 := \inf\{t : H(X_t^\varepsilon) = k\varepsilon^\beta\}$ satisfies $\mathbf{P}(\tau_1 < T) \geq \frac{1}{2}$ for all ε small enough.*

Proof. Let us again write down the equation for $H(X_t^\varepsilon)$,

$$H(X_t^\varepsilon) = \varepsilon^\beta \int_0^t AH(X_s^\varepsilon)ds + \varepsilon^{\beta/2} \int_0^t \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon) dW_s,$$

where A is the operator $Au(x) = \frac{1}{2}\sum_{i,j}[\sigma(x)\sigma^*(x)]_{i,j} \cdot \frac{\partial^2}{\partial x_i \partial x_j} u(x)$. First, note that both $\sigma\sigma^*$ and the Hessian matrix of H are positive-definite at $\mathbf{x}_0$, hence $d := AH(\mathbf{x}_0)/2 > 0$ and $AH(x) > d$ in a sufficiently small neighborhood of $\mathbf{x}_0$. Second, let $\lambda > 0$ be the smaller eigenvalue of the Hessian matrix of H at $\mathbf{x}_0$. Then, by Taylor's expansion at $\mathbf{x}_0$, we deduce that $|H(x)| \geq \frac{1}{4}\lambda|x|^2$ in a sufficiently small neighborhood of $\mathbf{x}_0$. Third, due to the boundedness of both the second derivatives of H and the norm of $\sigma\sigma^*$, we have $|\nabla H^*(x)\sigma(x)|^2 \leq \Lambda|x|^2$ for some Λ and all $x \in \mathbb{R}^2$. Let k be small enough so that $k \leq d/2$,

$$\sqrt{\frac{32k\Lambda T}{\pi\lambda d^2}} \exp(-\frac{\lambda d^2}{32k\Lambda T}) < \frac{1}{2},$$

and, in $U = \{x : |x| \leq \sqrt{\frac{4k}{\lambda}}\varepsilon^{\beta/2}\}$, we have $AH(x) > d$ and $|H(x)| \geq \frac{1}{4}\lambda|x|^2$. We define $\tau_2 = \inf\{t : H(X_t^\varepsilon) \in \partial U\}$, and it follows that $\tau_1 \leq \tau_2$. We are ready to estimate

$$\begin{aligned}
\mathbf{P}(\tau_1 \geq T) &= \mathbf{P}\left(\tau_2 \geq \tau_1 \geq T,\ \varepsilon^\beta \int_0^T AH(X_s^\varepsilon)ds + \varepsilon^{\beta/2}\int_0^T \nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)dW_s < k\varepsilon^\beta\right) \\
&\leq \mathbf{P}\left(\tau_2 \geq \tau_1 \geq T,\ \varepsilon^{\beta/2}\int_0^T \nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)dW_s < (k-d)\varepsilon^\beta\right) \\
&\leq \mathbf{P}\left(\tau_2 \geq \tau_1 \geq T,\ \int_0^T \nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)dW_s < -\frac{d\varepsilon^{\beta/2}}{2}\right) \\
&= \mathbf{P}\left(\tau_2 \geq \tau_1 \geq T,\ \tilde{W}(\int_0^T |\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds) < -\frac{d\varepsilon^{\beta/2}}{2}\right) \\
&\leq \mathbf{P}\left(\tau_2 \geq \tau_1 \geq T,\ \inf_{0\leq t\leq 4k\Lambda T\varepsilon^\beta/\lambda} \tilde{W}_t < -\frac{d\varepsilon^{\beta/2}}{2}\right) \\
&\leq 2\mathbf{P}\left(\tilde{W}_{4k\Lambda T\varepsilon^\beta/\lambda} > \frac{d\varepsilon^{\beta/2}}{2}\right) \\
&\leq \sqrt{\frac{32k\Lambda T}{\pi\lambda d^2}}\exp(-\frac{\lambda d^2}{32k\Lambda T}) \\
&< \frac{1}{2}.
\end{aligned}$$

□

Assuming that the process starts at x with $H(x) = k\varepsilon^\beta$, i.e. a certain distance away from the extremum point, we can apply Ito's formula with $f(x) = \sqrt{x}$ to the process $H(X_t^\varepsilon)$ in order to make the diffusion coefficient uniformly positive.

Lemma 4.3.3. *For a given non-constant φ with $\varphi_0 = H(x_0) = 0$ and $S(\varphi) < \infty$ and a positive constant c, there exists $\rho > 0$ such that the stopping time $\tau := \inf\{t : H(X_t^\varepsilon) = \rho\}$ satisfies $\mathbf{P}(\tau < T) \geq \exp(-c\varepsilon^{-\beta})$ for all ε sufficiently small, where $T = \inf\{t : |\varphi_t - \varphi_0| = \rho\}$.*

Proof. Let $\tau_1 = \inf\{t : H(X_t^\varepsilon) = k\varepsilon^\beta\}$ and let $\tau_2 = \tau - \tau_1$. By the strong Markov property of X_t^ε, we deduce that

$$\mathbf{P}(\tau < T) \geq \mathbf{P}(\tau_1 < \frac{T}{2},\ \tau - \tau_1 < \frac{T}{2}).$$

By Lemma 4.3.2, we can find such k that

$$\mathbf{P}(\tau_1 < \frac{T}{2}) \geq \frac{1}{2}.$$

Consider the process that starts at x satisfying $H(x) = k\varepsilon^\beta$. Since a two-dimensional nondegenerate diffusion does not reach the origin, we can apply Ito's formula to $\sqrt{H(X_t^\varepsilon)}$ and get

$$\begin{aligned}\sqrt{H(X_t^\varepsilon)} = \sqrt{k\varepsilon^\beta} + \varepsilon^\beta \int_0^t \left(\frac{AH(X_s^\varepsilon)}{2\sqrt{H(X_s^\varepsilon)}} - \frac{|\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2}{8\sqrt{H(X_s^\varepsilon)^3}} \right) ds \\ + \varepsilon^{\beta/2} \int_0^t \frac{\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)}{2\sqrt{H(X_s^\varepsilon)}} dW_s.\end{aligned}$$

Due to the positive-definiteness of the Hessian matrix at $\mathbf{x}_0 = (0,0)$, we can find $r > 0$ small enough that satisfies the conditions in Lemma 4.7.7, and there exist $\lambda_1, \lambda_2, \Lambda_1, \Lambda_2 > 0$ independent of r such that, for every $x \in B(\mathbf{x}_0, r)$, we have the estimates: $\lambda_1 |x|^2 \leq H(x) \leq \lambda_2 |x|^2$ and $\Lambda_1 |x| \leq |\nabla H(x)| \leq \Lambda_2 |x|$. Let ρ be sufficiently small so that $X_t^\varepsilon \in B(\mathbf{x}_0, r)$ for $t < \tau_2$. By Lemma 4.7.7,

$$\begin{aligned}
&\mathbf{P}\left(\tau_2 \geq T/2\right) \\
&= \mathbf{P}\left(\sqrt{H(X_{T/2}^\varepsilon)} \leq \sqrt{\rho},\ \tau_2 \geq T/2\right) \\
&\leq \mathbf{P}\left(\varepsilon^{\beta/2} \int_0^{T/2} \frac{\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)}{2\sqrt{H(X_s^\varepsilon)}} dW_s \leq \sqrt{\rho},\ \tau_2 \geq T/2\right) \\
&\leq \mathbf{P}\left(\tilde{W}\left(\int_0^{T/2} \left|\frac{\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)}{2\sqrt{H(X_s^\varepsilon)}}\right|^2 ds\right) \leq \sqrt{\rho}\varepsilon^{-\beta/2},\ \tau_2 \geq T/2\right) \\
&\leq \mathbf{P}\left(\tilde{W}_t \leq \sqrt{\rho}\varepsilon^{-\beta/2},\ \text{for some } \lambda T < t < 2\lambda T,\ \tau_2 \geq T/2\right) \\
&\leq \mathbf{P}\left(\tilde{W}_t \leq \sqrt{\rho}\varepsilon^{-\beta/2},\ \text{for some } \lambda T < t < 2\lambda T\right).
\end{aligned}$$

Here λ is a constant determined by $\lambda_1, \lambda_2, \Lambda_1, \Lambda_2$. Hence

$$\begin{aligned}\mathbf{P}\left(\tau_2 \leq T/2\right) &\geq \mathbf{P}\left(W_t > \sqrt{\rho}\varepsilon^{-\beta/2},\ \text{for all } \lambda T < t < 2\lambda T\right) \\ &\geq \frac{1}{2}\mathbf{P}\left(W_{\lambda T} > \sqrt{2\rho}\varepsilon^{-\beta/2}\right) \\ &\geq \frac{1}{\sqrt{2\pi\lambda T}}\exp(-\frac{3\rho}{\lambda T\varepsilon^{\beta}}).\end{aligned}$$

By Lemma 4.7.8, we can find ρ such that the value above is greater than $2\exp(-c\varepsilon^{-\beta})$. $\square$

4.3.3 Estimates near the separatrix

In this section, we prove that the probability that the process goes through an interior vertex sufficiently fast is not too small. For that purpose, we invoke the Morse-Palais Lemma to make concrete computations. Suppose that $\mathbf{x}_k$ is a nondegenerate saddle point of H and that $H(\mathbf{x}_k) = 0$. By the Morse-Palais Lemma, there exists a neighborhood U_0 (we may assume that U_0 is an $8l \times 4l$ rectangle with $l < 1$) of the origin and a diffeomorphism $\psi : U_0 \to V_0$, where V_0 is a neighborhood of $\mathbf{x}_k$, such that $H(\psi(\mu, \nu)) = \mu^2 - \nu^2 =: G(\mu, \nu)$, if $(\mu, \nu) \in U_0$. Since ψ is a diffeomorphism, there exists $\overline{M}$ such that $\|\psi\|$, $\|\psi^{-1}\|$, $\|J_\psi\|$, $\|J_{\psi^{-1}}\|$, $|\det(J_\psi)|$, $|\nabla\det(J_\psi)|$ are all bounded by $\overline{M}$, for all $(\mu, \nu) \in U_0$. Let us take $U \subset U_0$ to be the $4l \times 2l$ rectangle centered at the origin, and $V = \psi(U)$. Look at the deterministic slow motion

$$dx_t = \nabla^{\perp} H(x_t)dt.$$

Then we derive the equation for the process $\psi^{-1}(x_t)$ in U_0:

$$\frac{d\psi^{-1}(x_t)}{dt} = \frac{1}{\det(J_\psi(\psi^{-1}(x_t)))} \cdot \nabla^\perp G(\psi^{-1}(x_t)).$$

This means that $G(\psi^{-1}(x_t))$ remains constant as the process evolves, which leads to the fact that if x_0 is close to $\mathbf{x}_k$ and $H(x_0) > 0$, then the deterministic motion $\psi^{-1}(x_t)$ will exit from U across the upper or lower boundary, depending on which side of separatrix x_0 is on and the sign of $\det(J_\psi)$. We define, for $(\mu, \nu) \in U_0$,

$$T(\mu, \nu) = \frac{1}{2} \int_\nu^l \frac{\det(J_\psi(\sqrt{y^2 + \mu^2 - \nu^2}, y))}{\sqrt{y^2 + \mu^2 - \nu^2}} dy. \tag{4.5}$$

Without loss of generality, we assume $\det(J_\psi) > 0$. Then it's not hard to see that $T(\mu, \nu) = \inf\{t : \psi^{-1}(x_t) \notin U, \ \psi^{-1}(x_0) = (\mu, \nu)\}$ if $(\mu, \nu) \in U$, $\mu > 0$, and $0 < H(\psi((\mu, \nu))) < 3l^2$ (in this case, the process $\psi^{-1}(x_t)$ exits from U across the upper part of the boundary, and the integrand in (4.5) (multiplied by a half) is the reciprocal of the vertical speed); and that $T(\psi^{-1}(x_t)) - T(\psi^{-1}(x_{t+1})) = 1$ if $x_t, \ x_{t+1} \in V_0$ and $0 < H(x_t) < 3l^2$; and for $x_0 \in V$ and $0 < H(x_0) < 3l^2$,

$$T(\psi^{-1}(x_0)) \leq \overline{M} \left[\log\left(l + \sqrt{l^2 + H(x_0)}\right) - \frac{1}{2} \log(H(x_0)) \right]. \tag{4.6}$$

Furthermore, differentiating $T(\mu, \nu)$, we get the following estimates:

$$|\frac{\partial T}{\partial \mu}| \leq \overline{M} \left[\frac{|\pi\mu|}{\sqrt{\mu^2 - \nu^2}} + \frac{|\mu|}{(\mu^2 - \nu^2)\sqrt{1 + \mu^2 - \nu^2}} + \frac{|\nu|}{\mu^2 - \nu^2} \right] \leq \frac{C}{H(\psi(\mu, \nu))}, \tag{4.7}$$

$$|\frac{\partial T}{\partial \nu}| \leq \overline{M}\left[\frac{|\pi\nu|}{\sqrt{\mu^2-\nu^2}}+\frac{|\nu|}{(\mu^2-\nu^2)\sqrt{1+\mu^2-\nu^2}}+\frac{|\mu|}{\mu^2-\nu^2}\right] \leq \frac{C}{H(\psi(\mu,\nu))}, \quad (4.8)$$

for a certain positive constant C (details are given in the Section 4.7, Part 4). We will go through the following steps to get the desired estimates on the probability of going through an interior vertex sufficiently fast:

1. We first deal with the random process that starts on the separatrix. Then, by Theorem 3.2 in [16], for any $d > 0$, the process escapes to $C_{ki}(\varepsilon^{\frac{1+d}{2}\beta})$ in a fixed time with a probability bounded from below independently of ε.

2. If $d > 0$ is small enough, then the process that starts at $C_{ki}(\varepsilon^{\frac{1+d}{2}\beta})$ can reach $C_{ki}(\varepsilon^{\frac{1-d}{2}\beta})$ with at least an exponentially small probability, as is stated in Lemma 4.3.8.

3. After the process arrives at $C_{ki}(\varepsilon^{a\beta})$ with $a < 1/2$, it can further travel to $C_{ki}(\varepsilon^{(a-d)\beta})$ with d small but positive, with at least an exponentially small probability. This is the statement of Lemma 4.3.6. Hence, after finitely many steps, the process can reach $C_{ki}(\varepsilon^{d\beta})$.

4. The last step would be traveling from $C_{ki}(\varepsilon^{d\beta})$ to a constant distance from the separatrix. This is done in Lemma 4.3.10.

5. Similarly, the process that starts at a certain distance from the separatrix can hit it within a constant time with at least an exponentially small probability. Therefore, the probability that the process follows a certain trajectory going across the separatrix is not terribly small, which is the result of Lemma 4.3.14.

Let us define the curves:

$$\gamma_{in} = \{x \in \partial V \cap D_i : \nabla^{\perp} H(x) \text{ points inwards of } V\},$$

$$\gamma_{out} = \{x \in \partial V \cap D_i : \nabla^{\perp} H(x) \text{ points outwards of } V\}.$$

The following lemma contains simple estimates on the Brownian motion that will be needed later. The proof will be given in the Section 4.7, Part 4.

Lemma 4.3.4. *(i) For $0 < d < a < \frac{1}{2}$ and every $\kappa > 0$, we have the following estimate for ε small enough:*

$$\mathbf{P}\left(\varepsilon^{\beta/2} W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|,\ \varepsilon^{\beta/2} W_t < \frac{1}{4} a\beta|\log\varepsilon|\varepsilon^{a\beta},\ \forall\, 0 < t < T\right)$$
$$\geq 2\exp\left(-\varepsilon^{-(1-2(a-d))\beta-\kappa}\right).$$

(ii) For $0 < d < \frac{1/\beta - 1}{3} \wedge \frac{1}{2}$ and every $\kappa > 0$, we have the following estimate for ε small enough:

$$\mathbf{P}\begin{pmatrix} \varepsilon^{\beta/2} W_T < -2(1-d)\beta\varepsilon^{\frac{1-d}{2}\beta}|\log\varepsilon|; \\ \varepsilon^{\beta/2} W_t < \varepsilon^{\frac{1+d}{2}\beta},\ \forall\, 0 < t < \varepsilon^{d\beta}; \\ \varepsilon^{\beta/2} W_t < -\varepsilon^{\frac{1-d}{2}\beta},\ \forall\, \varepsilon^{d\beta} < t < T \end{pmatrix} \geq 2\exp(-\varepsilon^{-2d\beta-\kappa}).$$

(iii) For $0 < d < \frac{1/\beta - 1}{2}$ and every $A > 0$, we have the following estimate for ε small enough:

$$\mathbf{P}\left(\varepsilon^{\beta/2} W_T < -A,\ \varepsilon^{\beta/2} W_t < \frac{1}{4} d\beta|\log\varepsilon|\varepsilon^{d\beta},\ \forall\, 0 < t < T\right) \geq 2\exp\left(-\frac{A^2}{T}\varepsilon^{-\beta}\right).$$

As we've already seen from (4.6), the exit time for the deterministic process from the neighborhood of a saddle point is approximately $|\log H|$. Now we need a similar statement for X_t^ε.

Lemma 4.3.5. *Suppose that O_k is an interior vertex and an endpoint of I_i. For any given $0 < a < \frac{1}{2}$, $0 < d < \frac{1/\beta - 1}{2} \wedge a$, and for every $0 < \kappa < \frac{1}{8}(1-(1+2d)\beta)$, every $T > 0$, we have the following estimate for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{a\beta}, \frac{11}{10}\varepsilon^{a\beta})$ and ε small enough*

$$\mathbf{P}\left(\int_0^t |\log H(X_s^\varepsilon)\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds > t, \ \forall\, t \le \eta \wedge T\right) \ge 1 - \exp\left(-\varepsilon^{-(1-2(a-d))\beta - \kappa}\right),$$

where $\eta = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{(a-d)\beta}) \cup C_{ki}(\varepsilon^{a\beta}/2)\}$.

The idea behind this lemma is that the process starts on γ_{out} and moves away from the saddle point first, with the intervals of time away from and near the saddle point alternating repeatedly. The integrand is large when the process is away from the saddle point, which is sufficient to compensate for the relatively long time spent near the saddle point. A detailed proof will be given in the Section 4.7, Part 4. Now let us formulate the argument of Step 3 (above) in the following lemma.

Lemma 4.3.6. *Suppose that O_k is an interior vertex and an endpoint of I_i. For any given $0 < a < \frac{1}{2}$, $0 < d < \frac{1/\beta - 1}{2} \wedge a$, and for every $T > 0$, every $c > 0$, there exists ε_0 small enough such that, for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{a\beta}, \frac{11}{10}\varepsilon^{a\beta})$, we have*

$$\mathbf{P}(\tau \le T) \ge \exp\left(-c\varepsilon^{-\beta}\right), \ \forall \varepsilon < \varepsilon_0,$$

where $\tau = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{(a-d)\beta})\}$.

Proof. Define $\tau' = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{a\beta}/2)\}$, $\eta = \tau \wedge \tau'$, and $f(x) = x\log x - x$. Since

$f \in C^2([\varepsilon^{a\beta}/2, \varepsilon^{(a-d)\beta}])$, Ito's formula can be applied here:

$$
\begin{aligned}
f(H(X^\varepsilon_\eta)) &= f(H(x_0)) + \varepsilon^\beta \int_0^\eta \left(\log H(X^\varepsilon_s) A H(X^\varepsilon_s) + \frac{|\nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2}{2H(X^\varepsilon_s)} \right) ds \\
&\quad + \varepsilon^{\beta/2} \int_0^\eta \log H(X^\varepsilon_s) \nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s) dW_s \\
&= f(H(x_0)) + \varepsilon^\beta \int_0^\eta \left(\log H(X^\varepsilon_s) A H(X^\varepsilon_s) + \frac{|\nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2}{2H(X^\varepsilon_s)} \right) ds \\
&\quad + \varepsilon^{\beta/2} \tilde{W} \left(\int_0^\eta |\log H(X^\varepsilon_s) \nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2 ds \right),
\end{aligned}
$$

where $\tilde{W}$ is another Brownian motion. To simplify notations, define events R and S as follows:

$$
\begin{aligned}
R &= \left\{ \varepsilon^{\beta/2} \tilde{W}_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta |\log \varepsilon|,\ \varepsilon^{\beta/2} \tilde{W}_t < \frac{1}{4} a\beta |\log \varepsilon| \varepsilon^{a\beta} \text{ for every } 0 < t < T \right\}, \\
S &= \left\{ \int_0^t |\log H(X^\varepsilon_s) \nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2 ds > t,\ \forall\, t \le \eta \wedge T \right\}.
\end{aligned}
$$

By Lemma 4.3.4(i) and Lemma 4.3.5, we know that $\mathbf{P}(R \cap S) \ge \exp\left(-\varepsilon^{-(1-2(a-d))\beta - \kappa}\right)$ for any positive κ and all sufficiently small ε. Therefore, since κ can be taken arbitrarily small, it remains to show that $R \cap S$ implies $\tau \le T$. On this event, if we assume that $\eta > T$, then

$$
\int_0^T |\log H(X^\varepsilon_s) \nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2 ds > T.
$$

So, we may define $\eta' = \inf\{t : \int_0^t |\log H(X^\varepsilon_s) \nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2 ds = T\} < T < \eta$, and still write the equation

$$
f(H(X^\varepsilon_{\eta'})) = f(H(x_0)) + \varepsilon^\beta \int_0^{\eta'} \left(\log H(X^\varepsilon_s) A H(X^\varepsilon_s) + \frac{|\nabla H(X^\varepsilon_s)^* \sigma(X^\varepsilon_s)|^2}{2H(X^\varepsilon_s)} \right) ds
$$

$$+ \varepsilon^{\beta/2}\tilde{W}\left(\int_0^{\eta'} |\log H(X_s^\varepsilon)\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds\right)$$

$$= f(H(x_0)) + \varepsilon^\beta \int_0^{\eta'} \left(\log H(X_s^\varepsilon) A H(X_s^\varepsilon) + \frac{|\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2}{2H(X_s^\varepsilon)}\right) ds + \varepsilon^{\beta/2}\tilde{W}_T.$$

Since the Brownian motion here dominates the whole right-hand side due to the definition of R, we have $f(H(X_{\eta'}^\varepsilon)) < f(\varepsilon^{(a-d)\beta})$, which contradicts the fact that $\eta' < \eta$. Thus $\eta \le T$. Then it follows that $\tau \le T$ since $f(H(X_\eta^\varepsilon)) \ne f(\varepsilon^{a\beta}/2)$ by the definition of the set R. $\square$

We can complete the proofs corresponding to Steps 2 and 4 above in the same way.

Lemma 4.3.7. *Suppose that O_k is an interior vertex and an endpoint of I_i. For every $d < \frac{1/\beta - 1}{3} \wedge \frac{1}{2}$, and for every $T > 0$, we have the following estimate for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{\frac{1+d}{2}\beta}, \frac{11}{10}\varepsilon^{\frac{1+d}{2}\beta})$ and ε small enough*

$$\mathbf{P}\left(\int_0^t |\log H(X_s^\varepsilon)\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds > t,\ \forall\, t \le \eta \wedge T\right) \ge 1 - \exp\left(-\varepsilon^{-2d\beta-\kappa}\right),$$

where $\eta = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{\frac{1-d}{2}\beta}) \cup C_{ki}(\varepsilon^{\frac{1+d}{2}\beta}/2)\}$.

Since the proof is similar to that of Lemma 4.3.5, the details are omitted.

Lemma 4.3.8. *Suppose that O_k is an interior vertex and an endpoint of I_i. For every $d < \frac{1/\beta - 1}{3} \wedge \frac{1}{2}$, and for every $T > 0$, $c > 0$, there exists ε_0 small enough, such that for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{\frac{1+d}{2}\beta}, \frac{11}{10}\varepsilon^{\frac{1+d}{2}\beta})$, we have*

$$\mathbf{P}(\tau \le T) \ge \exp(-c\varepsilon^{-\beta}),\ \forall \varepsilon < \varepsilon_0,$$

where $\tau = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{\frac{1-d}{2}\beta})\}$.

Proof. The proof is similar to that of Lemma 4.3.6. Define $\tau' = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{\frac{1+d}{2}\beta}/2)\}$, $\eta = \tau \wedge \tau'$, and $f(x) = x \log x - x$. Since $f \in C^2([\frac{1}{2}\varepsilon^{\frac{1+d}{2}\beta}, \varepsilon^{\frac{1-d}{2}\beta}])$, Ito's formula can be applied here:

$$f(H(X_\eta^\varepsilon)) = f(H(x_0)) + \varepsilon^\beta \int_0^\eta \left(\log H(X_s^\varepsilon) A H(X_s^\varepsilon) + \frac{|\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2}{2H(X_s^\varepsilon)} \right) ds$$
$$+ \varepsilon^{\beta/2} \tilde{W} \left(\int_0^\eta |\log H(X_s^\varepsilon) \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds \right).$$

Let

$$R = \left\{ \begin{matrix} \varepsilon^{\beta/2} \tilde{W}_T < -2(1-d)\beta \varepsilon^{\frac{1-d}{2}\beta} |\log \varepsilon|;\ \varepsilon^{\beta/2} \tilde{W}_t < \varepsilon^{\frac{1+d}{2}\beta},\ \forall\, 0 < t < \varepsilon^{d\beta}; \\ \varepsilon^{\beta/2} \tilde{W}_t < -\varepsilon^{\frac{1-d}{2}\beta},\ \forall\, \varepsilon^{d\beta} < t < T \end{matrix} \right\},$$
$$S = \left\{ \int_0^t |\log H(X_s^\varepsilon) \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds > t,\ \forall\, t \le \eta \wedge T \right\}.$$

By Lemma 4.3.4(ii) and Lemma 4.3.7, we know that $\mathbf{P}(R \cap S) \ge \exp\left(-\varepsilon^{-2d\beta-\kappa}\right)$ for any positive κ, and $R \cap S$ implies $\tau \le T$, so we're done. $\square$

The proof of the following lemma is also similar to that of Lemma 4.3.5.

Lemma 4.3.9. *Suppose that O_k is an interior vertex and an endpoint of I_i. For every $d < \frac{1/\beta - 1}{2} \wedge 1$, and for every $T > 0$, we have the following estimate for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{d\beta}, \frac{11}{10}\varepsilon^{d\beta})$ and ε small enough*

$$\mathbf{P}\left(\int_0^t |\log H(X_s^\varepsilon) \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds > t,\ \forall\, t \le \eta \wedge T \right) \ge 1 - \exp\left(-\varepsilon^{-\beta-\kappa}\right),$$

where $\eta = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta) \cup C_{ki}(\varepsilon^{d\beta}/2)\}$.

Lemma 4.3.10. *Suppose that O_k is an interior vertex and an endpoint of I_i. For every $d < \frac{1/\beta-1}{2} \wedge 1$, and for every $T > 0$, $c > 0$, there exist $\delta > 0$ and $\varepsilon_0 > 0$ such that, for each $x_0 \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^{d\beta}, \frac{11}{10}\varepsilon^{d\beta})$, we have*

$$\mathbf{P}(\tau \leq T) \geq \exp\left(-c\varepsilon^{-\beta}\right),\ \forall \varepsilon < \varepsilon_0,$$

where $\tau = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta)\}$.

Proof. The proof is, again, similar to that of Lemma 4.3.6. Define $\tau' = \inf\{t : X_t^\varepsilon \in C_{ki}(\varepsilon^{d\beta}/2)\}$, $\eta = \tau \wedge \tau'$ and $f(x) = x\log x - x$. Since $f \in C^2([\frac{1}{2}\varepsilon^{d\beta}, \delta])$, Ito's formula can be applied here:

$$\begin{aligned} f(H(X_\eta^\varepsilon)) = f(H(x_0)) + \varepsilon^\beta \int_0^\eta \left(\log H(X_s^\varepsilon) A H(X_s^\varepsilon) + \frac{|\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2}{2H(X_s^\varepsilon)}\right) ds \\ + \varepsilon^{\beta/2}\tilde{W}\left(\int_0^\eta |\log H(X_s^\varepsilon) \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds\right). \end{aligned}$$

Let

$$\begin{aligned} R &= \left\{\varepsilon^{\beta/2}\tilde{W}_T < 2f(\delta),\ \varepsilon^{\beta/2}\tilde{W}_t < \frac{1}{4}d\beta|\log\varepsilon|\varepsilon^{d\beta} \text{ for every } 0 < t < T\right\}, \\ S &= \left\{\int_0^t |\log H(X_s^\varepsilon)\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds > t,\ \forall\, t \leq \eta \wedge T\right\}. \end{aligned}$$

By Lemma 4.3.4 (iii) (with $A = -2f(\delta)$) and Lemma 4.3.9, we know that $\mathbf{P}(R \cap S) \geq \exp\left(-\frac{4f(\delta)^2}{T}\varepsilon^{-\beta}\right)$, and $R \cap S$ implies $\tau \leq T$, so we're done if we choose δ so small that $4f(\delta)^2 < cT$. □

The following lemma ensures that the assumptions that the process starts on a piece of γ_{out} in the lemmas above make sense, since the process reaches the piece of the curve with a uniformly positive probability.

Lemma 4.3.11. *For $A < 1/2$ and every fixed $T > 0$, we have the following estimate for ε small enough:*

$$\mathbf{P}(\tau < T) \geq 1/2,$$

where $\tau = \inf\{t : X_t^\varepsilon \in \gamma_{out} \cap D_i(\frac{9}{10}\varepsilon^A, \frac{11}{10}\varepsilon^A)\}$, for each $x_0 \in C_{ki}(\varepsilon^A)$.

Proof. The proof is similar to that of Lemma 4.3.5. Let us look at the slow motion:

$$d\tilde{X}_t^\varepsilon = \nabla^\perp H(\tilde{X}_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(\tilde{X}_t^\varepsilon)d\tilde{W}_t, \qquad \tilde{X}_0^\varepsilon = x_0 \in \mathbb{R}^2,$$

and take $\alpha \in (A, 1/2)$. Then the idea is that $\tau < T$ on the event

$$E := \bigcap_{k=0}^{2A\overline{M}|\log\varepsilon|} \left\{ \sup_{\Delta\in[0,1]} |\sqrt{\varepsilon}\int_k^{k+\Delta} \sigma(\tilde{X}_s^\varepsilon)dW_s| < \varepsilon^\alpha \right\},$$

and $\mathbf{P}(E) \geq 1 - 2A\overline{M}|\log\varepsilon|\exp(-\varepsilon^{\alpha-1/2}) \geq 1/2$ for ε sufficiently small, since σ is bounded and the stochastic integral can be represented as a time changed Brownian motion. See the proof of Lemma 4.3.5 for details. □

Lemma 4.3.12. *Suppose that O_k is an interior vertex. For any $T > 0$, $c > 0$, there exist $\delta > 0$ and $\varepsilon_0 > 0$ such that, for every $\varepsilon < \varepsilon_0$ and each $x_0 \in C_k$,*

$$\mathbf{P}(\tau \leq T) \geq \exp(-c\varepsilon^{-\beta}),$$

where $\tau = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta)\}$.

Proof. Set $d = \frac{1/\beta - 1}{4} \wedge \frac{1}{4}$ and $n = [\frac{1}{d}] + 8$. By Theorem 3.2 in [16], for every $d > 0$, the process reaches $C_{ki}(\varepsilon^{\frac{1+d}{2}\beta})$ by time T/n with probability at least $p > 0$, where p is independent of ε. Then we apply Lemma 4.3.6, Lemma 4.3.8, Lemma 4.3.10, and Lemma 4.3.11 with our fixed d, $\frac{T}{2n}$, and $\frac{c}{2n}$. By the strong Markov property of X_t^ε, we get the desired result. □

As in Lemma 4.3.12, the process starting at $C_{kj}(\delta)$ arrives at the separatrix with probability at least $\exp(-c\varepsilon^{-\beta})$ prior to time T. Combining these two results, we get the following lemma.

Lemma 4.3.13. *Suppose that O_k is an interior vertex with $O_k \sim I_i$ and $O_k \sim I_j$. Then, for any $T > 0$, $c > 0$, there exists δ such that, for any $x_0 \in C_{kj}(\delta)$,*

$$\mathbf{P}(\tau \leq T) \geq \exp\left(-c\varepsilon^{-\beta}\right),$$

where $\tau = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta)\}$.

Next, we formulate the lower bound for the probability that the random process goes through an interior vertex, near a certain deterministic path.

Lemma 4.3.14. *Suppose that O_k is an interior vertex with $O_k \sim I_i$ and $O_k \sim I_j$. For every $\delta > 0$, $\gamma > 0$, $h > 0$, and a trajectory $\boldsymbol{\varphi}$ such that $\boldsymbol{\varphi}_0 = (j, H(\mathbf{x}_k) \pm \delta)$, $\boldsymbol{\varphi}_T = (i, H(\mathbf{x}_k) \pm \delta)$, and $|\varphi_t - H(\mathbf{x}_k)| \leq \delta$ for $t \in [0, T]$, there exist $h_0 > 0$ and $\varepsilon_0 > 0$ such that, for any x_0 satisfying $r(Y(x_0), \boldsymbol{\varphi}_0) < h_0$, we have the estimate for all $\varepsilon < \varepsilon_0$*

$$\mathbf{P}\left(r(Y(X_T^\varepsilon), \boldsymbol{\varphi}_T) < h, \rho_{0T}(Y(X_t^\varepsilon), \boldsymbol{\varphi}_t) < 2\delta\right) \geq \exp(-\varepsilon^{-\beta}(S_{0T}(\boldsymbol{\varphi}) + \gamma)).$$

Proof. Without loss of generality, we assume that $H(\mathbf{x}_k) = 0$ and $h < \delta$. Due to the absolute continuity of φ, we can fix $\hat{t}$ so small that $|a - b| < \hat{t}$ implies that $|\varphi_a - \varphi_b| < h/2$. Fix $\delta' < \delta/2$

positive and small, such that for any $x_0 \in C_{kj}(\delta')$,

$$\mathbf{P}(\tau_{\delta'} \leq \hat{t};\ X_t^\varepsilon \in D_k(\delta/2),\ t \in [0, \tau_{\delta'}]) \geq \exp\left(-\gamma\varepsilon^{-\beta}/4\right), \tag{4.9}$$

where $\tau_{\delta'} = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta')\}$. This is possible since, for δ' small enough, the process starting on $C_{kj}(\delta')$ escapes $D_k(\delta/2)$, within a fixed time, with a probability that is bounded from above by $\exp(-C\varepsilon^{-\beta})$ for certain constant C, while the constant c in Lemma 4.3.13 corresponding to δ' can be chosen such that $c < C$.

Let $t_1 = \inf\{t > 0 : \varphi_t = \pm\delta'/2\}$, $t_2 = \sup\{t > 0 : \varphi_t = \pm\delta'\}$, and $\tilde{\varphi}_t = \varphi_{t-\hat{t}}$ for $t \in [t_2 + \hat{t}, T]$. By Lemma 4.3.1, we can fix $h_0 < \delta$ such that, for any x_0 satisfying $|H(x_0) - \varphi_0| < h_0$,

$$\mathbf{P}(\rho_{0,t_1}(H(X_t^\varepsilon), \varphi_t) < \delta'/2) \geq \exp\left[-\varepsilon^{-\beta}(S_{0,t_1}(\varphi) + \gamma/4)\right], \tag{4.10}$$

and we can fix $\mu < \delta'$ such that, conditioned on $X_{t_2+\hat{t}}^\varepsilon = x$ satisfying $|H(x) - \tilde{\varphi}_{t_2+\hat{t}}| < \mu$,

$$\mathbf{P}(\rho_{t_2+\hat{t},T}(H(X_t^\varepsilon), \tilde{\varphi}_t) < h/2) \geq \exp\left[-\varepsilon^{-\beta}(S_{t_2,T-\hat{t}}(\varphi) + \gamma/4)\right], \tag{4.11}$$

where $\tilde{\varphi}_t = \varphi_{t-\hat{t}}$. Furthermore, let $\tau_1 = \inf\{t : X_t^\varepsilon \in C_{kj}(\delta')\}$ and $\tau_2 = \inf\{t : X_t^\varepsilon \in C_{ki}(\delta')\}$. Let's consider the event:

$$A = \left\{ \begin{array}{c} \rho_{0,t_1}(H(X_t^\varepsilon), \varphi_t) < \delta'/2; \tau_2 - \tau_1 < \hat{t},\ X_t^\varepsilon \in D_k(\delta),\ t \in [\tau_1, \tau_2]; \\ X_t^\varepsilon \in D_i(\delta' - \mu, \delta' + \mu), t \in [\tau_2, t_2 + \hat{t}];\ |H(X_t^\varepsilon) - \tilde{\varphi}_t| < h/2, t \in [t_2 + \hat{t}, T] \end{array} \right\}.$$

Note that A implies the desired event, hence it suffices to prove $\mathbf{P}(A) \geq \exp(-\varepsilon^{-\beta} S_{0T}(\boldsymbol{\varphi}) + \gamma)$.

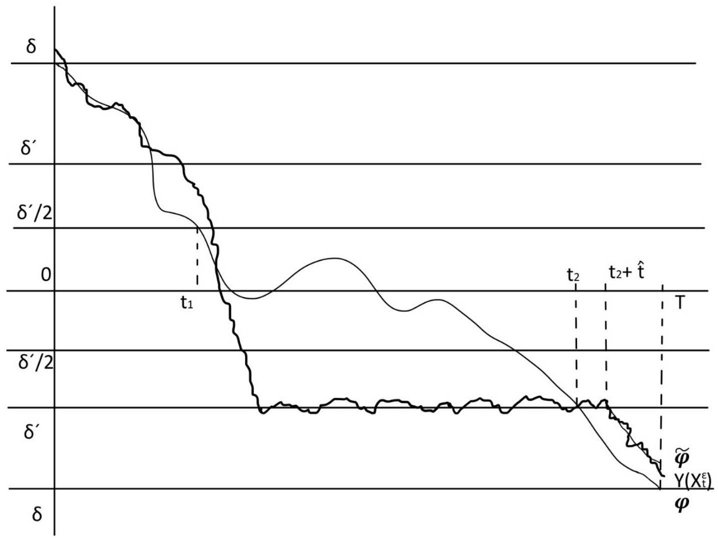


Figure 4.2: A typical graph where event A happens.

To see this, combine (4.9), (4.10) and (4.11) and note that (4.3) implies that

$$\mathbf{P}\left(X_t^\varepsilon \in D_i(\delta' - \mu, \delta' + \mu), t \in [\tau_2, t_2 + \hat{t}]\right) \geq \frac{1}{2},$$

for ε small enough. Thus, by the strong Markov property of X_t^ε,

$$\mathbf{P}(A) \geq \exp(-\varepsilon^{-\beta}(S_{0T}(\boldsymbol{\varphi}) + \gamma)). \qquad \square$$

4.3.4 Local large deviation principle: lower bound

With all the preliminary results obtained, we can get the lower bound for the large deviation principle.

Lemma 4.3.15. *For any $\delta > 0$, $\gamma > 0$, and $\boldsymbol{\varphi} \in \boldsymbol{C}([0,T],\Gamma)$ with $\boldsymbol{\varphi_0} = Y(x_0)$, there exists ε_0 such that*

$$\mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi}) < \delta\right) \geq \exp\left[-\varepsilon^{-\beta}(S(\boldsymbol{\varphi}) + \gamma)\right],$$

for any $\varepsilon < \varepsilon_0$.

Proof. We can assume that φ is absolutely continuous and $\varphi' \in L^2([0,T])$ for the rest of the proof, because $S(\boldsymbol{\varphi}) = \infty$ for any other $\boldsymbol{\varphi} = (i, \varphi)$. Consider the case where $\boldsymbol{\varphi}_0$ is at an exterior vertex O_k and $\boldsymbol{\varphi}_T$ is not at any vertex. Given $\delta > 0$ and $\gamma > 0$, by Lemma 4.3.3, we can find $\rho < \delta/4 \wedge \inf_k |\varphi_T - H(\mathbf{x}_k)|$ such that, for ε sufficiently small,

$$\mathbf{P}(\tau < q) \geq \exp(-\varepsilon^{-\beta}\gamma/4),$$

where $q = \inf\{t : r(\boldsymbol{\varphi}_t, O_k) = \rho\}$ and $\tau = \inf\{t : |H(X_t^\varepsilon) - H(\mathbf{x}_k)| = \rho\}$. Let $\delta' = \frac{1}{2}\rho$, and define the following three sequences inductively:

$$\begin{aligned}
q_0 &= q; \\
t_k &= \inf\left\{t > q_{k-1} : r(\boldsymbol{\varphi}_t, O) = \delta' \text{ for some vertex } O\right\}, \\
p_k &= \sup\left\{t < t_k : r(\boldsymbol{\varphi}_t, O) = 2\delta' \text{ for some vertex } O\right\}, \\
q_k &= \inf\left\{t > t_k : r(\boldsymbol{\varphi}_t, O) = 2\delta' \text{ for some vertex } O\right\},
\end{aligned}$$

for $k \leq n$, where n is such that the set $\{t > q_n : r(\boldsymbol{\varphi}_t, O) = \delta' \text{ for some vertex } O\}$ is empty. Note that $n \geq 0$ is finite due to the absolutely continuity of φ. We define $p_{n+1} = T$. By Lemma 4.3.1, we can find $h < \frac{1}{2}\delta'$ such that, for all $h' < h$, there exists ε_0 small enough such that, conditioned on $X_{q_{k-1}}^\varepsilon = x$ satisfying $|H(x) - \varphi_{q_{k-1}}| < h$ and for all $\varepsilon < \varepsilon_0$,

$$\mathbf{P}\left(|H(X_{p_k}^\varepsilon) - \varphi_{p_k}| < h',\ \rho_{q_{k-1},p_k}(Y(X_t^\varepsilon), \boldsymbol{\varphi}_t) < \delta\right) \geq \exp\left[-\varepsilon^{-\beta}(S_{q_{k-1},q_k}(\boldsymbol{\varphi}) + \frac{\gamma}{2^{k+2}})\right]$$

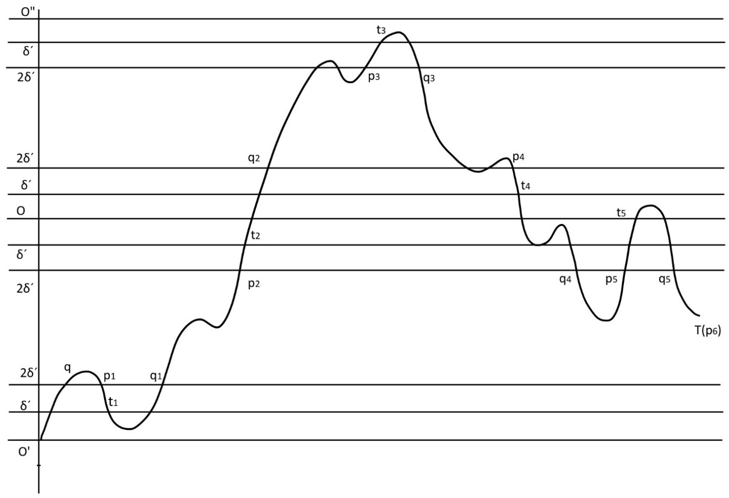


Figure 4.3: A typical partition of a trajectory.

holds for all $1 \leq k \leq n+1$. By the strong Markov property of X_t^ε (since the process can remain near the same level curve between time τ and q), for each $h' > 0$, there exists ε_0 such that, for all $\varepsilon < \varepsilon_0$,

$$\mathbf{P}(|H(X_q^\varepsilon) - \varphi_q| < h',\ \rho_{0,q}(Y(X_t^\varepsilon), \boldsymbol{\varphi}_t) < \delta) \geq \exp(-\varepsilon^{-\beta}\gamma/2).$$

On the other hand, for $1 \leq k \leq n$, on the interval $[p_k, q_k]$, we have two different situations:

(i) If $\boldsymbol{\varphi}_{p_k} = \boldsymbol{\varphi}_{q_k}$, then, conditioned on $X_{p_k}^\varepsilon = x$ satisfying $|H(x) - \varphi_{p_k}| < h/2$, for all ε small enough,

$$\begin{aligned} &\mathbf{P}\left(|H(X_{q_k}^\varepsilon) - \varphi_{q_k}| < h,\ \rho_{p_k,q_k}(Y(X_t^\varepsilon), \boldsymbol{\varphi}_t) < \delta\right) \\ \geq\, &\mathbf{P}\left(\sup_{p_k \leq t \leq q_k} |H(X_t^\varepsilon) - H(X_{p_k}^\varepsilon)| < h/2\right) > \frac{1}{2}. \end{aligned}$$

(ii) If $\boldsymbol{\varphi}_{p_k} \neq \boldsymbol{\varphi}_{q_k}$, then, for ε small enough, by Lemma 4.3.14, there exists $\mu < h/2$ such that, conditioned on $X^\varepsilon_{p_k} = x$ satisfying $|H(x) - \varphi_{p_k}| < \mu$,

$$\mathbf{P}\left(|H(X^\varepsilon_{q_k}) - \varphi_{q_k}| < h,\ \rho_{p_k,q_k}(Y(X^\varepsilon_t), \boldsymbol{\varphi}_t) < \delta\right) \geq \exp(-\varepsilon^{-\beta}(S_{p_k q_k}(\boldsymbol{\varphi}) + \frac{\gamma}{2^{k+2}})).$$

With the estimates above combined, we obtain that

$$\mathbf{P}\left(\rho_{0,T}(Y(X^\varepsilon_t), \boldsymbol{\varphi}) < \delta\right) \geq \exp\left[-\varepsilon^{-\beta}(S(\boldsymbol{\varphi}) + \gamma)\right]$$

for any ε small enough. In the case where $\boldsymbol{\varphi}_0$ is at an interior vertex or inside an edge, or $\boldsymbol{\varphi}_T$ is at a vertex, the proof is similar. $\square$

Theorem 4.3.16. *Given $\boldsymbol{\varphi} \in \boldsymbol{C}\left([0,T], \Gamma\right)$ with $\boldsymbol{\varphi_0} = Y(x_0)$,*

$$\varliminf_{\delta \to 0} \varliminf_{\varepsilon \to 0} \varepsilon^\beta \log \mathbf{P}\left(\rho_{0,T}(Y(X^\varepsilon_t), \boldsymbol{\varphi})\right) < \delta) \geq -S(\boldsymbol{\varphi}).$$

Proof. This is an immediate consequence of the previous lemma. $\square$

4.4 Upper bound for the large deviation principle

In this section, we deal with the upper bound for the local large deviation principle. We start with the discussion about what happens inside a single edge (Lemma 4.4.1 and Lemma 4.4.2). Again, we simplify the notations when there's no ambiguity, as we did in Section 3.1.

Lemma 4.4.1. *Suppose that $C = \overline{D_i(H_1, H_2)} \subset D_i$ and $X^\varepsilon_0 = x \in C$. Then, for every $\delta > 0$,* *every $\gamma > 0$, and every $s > 0$, there exists $\varepsilon_0 > 0$ such that*

$$\mathbf{P}(X_t^\varepsilon \in C, \rho_{0,T}(Y(X_t^\varepsilon), \Phi(s)) \geq \delta) \leq \exp(-\varepsilon^{-\beta}(s-\gamma))$$

for all $\varepsilon \leq \varepsilon_0$ and every x, where $\Phi(s) = \{\boldsymbol{\varphi} \in \boldsymbol{C}([0,T], I_i) : S(\boldsymbol{\varphi}) \leq s,\ \varphi_0 \in [H_1, H_2]\}$.

Proof. Without loss of generality, we let $T = 1$. Since $H(C)$ is compact, and C is away from the critical points, there exist $m > 0$, $M > 0$, and $L > 0$, such that:

(1) $|AH(x)| + |\nabla H(x)^* \sigma(x)|^2 < M,\ \forall x \in C$;

(2) $|\nabla H(x)^* \sigma(x)|^2 > m,\ \forall x \in C$;

(3) $m < T(H) < M$ and $m < B^2(H) < M,\ \forall H \in H(C)$;

(4) $|\nabla H(x) - \nabla H(y)| + ||\nabla H(x)^* \sigma(x)|^2 - |\nabla H(y)^* \sigma(y)|^2| < L|x-y|,\ \forall x, y \in C$;

(5) $|T(H_1) - T(H_2)| + |T(H_1)B^2(H_1) - T(H_2)B^2(H_2)| < L|H_1 - H_2|,\ \forall H_1, H_2 \in H(C)$.

For $s > 0$, $\gamma > 0$, and $\delta > 0$, we choose the following parameters:

(1) $0 < \mu < 1/3$ such that $s(1-2\mu) > s - \gamma$;

(2) $\delta' > 0$ such that $\delta' < \delta \wedge \frac{m^2\mu^2}{4L(2M+1)}$, which particularly implies that

$$L\delta'(2M+1) < \frac{\mu m^2}{3},$$

$$\frac{1-\mu^2}{2} \cdot \frac{2m + L\delta'}{2m - L\delta'} < \frac{1-\mu^2/2}{2};$$

(3) $n \in \mathbb{N}$ such that $n{\delta'}^2 > 64M(s-\gamma)$.

For every $0 \leq k < n$, define the random variable $T_k^\varepsilon = \varepsilon^{1-\beta} T(H(X_{\frac{k}{n}}^\varepsilon))$, the event A_k as in Lemma 4.3.1, and the event B_k where the process stays within the compact set, in which $|\nabla H(X_t^\varepsilon)|$ is bounded, so that A_k happens with overwhelming probability:

$$
\begin{aligned}
A_k = & \bigcap_{p=0}^{\lfloor \frac{1}{nT_k^\varepsilon} \rfloor} \left\{ \sup_{pT_k^\varepsilon + \frac{k}{n} \leq t \leq (p+1)T_k^\varepsilon + \frac{k}{n}} \varepsilon^{\beta/2} \left| \int_{k/n + pT_k^\varepsilon}^{t} \sigma(X_s^\varepsilon) dW_s \right| \leq M \cdot \varepsilon^{\frac{1-\beta}{4}} \right\} \\
& \bigcap \left\{ \sup_{0 \leq t \leq \frac{1}{n}} \left| H(X_{\frac{k}{n}+t}^\varepsilon) - H(X_{\frac{k}{n}}^\varepsilon) \right| \leq \frac{\delta'}{2} \right\}; \\
B_k = & \left\{ X_t^\varepsilon \in C, \forall t \in [\frac{k}{n}, \frac{k+1}{n}] \right\}.
\end{aligned}
$$

If we denote

$$
Y_k = \frac{1}{2\varepsilon^\beta} \cdot \frac{n}{B^2 \left(H(X_{\frac{k}{n}}^\varepsilon) \right) - \frac{L\delta'}{2}} \left| H(X_{\frac{k+1}{n}}^\varepsilon) - H(X_{\frac{k}{n}}^\varepsilon) \right|^2,
$$

then, by Lemma 4.7.9, we have $\mathbf{E}(\chi_{B_k \cap A_k} \cdot \exp((1-\mu)Y_k) \mid \mathcal{F}_{\frac{k}{n}}) < c_\mu$ for some $0 < c_\mu < \infty$.

Let l^ε be the random polygon with vertices at $\left(\frac{k}{n}, H(X_{\frac{k}{n}}^\varepsilon) \right)$ for $0 \leq k \leq n$, and let $B = \bigcap_{k=1}^{n-1} B_k$. Then

$$
\begin{aligned}
& \mathbf{P}\left(X_t^\varepsilon \in C, \rho_{0,1}(Y(X_t^\varepsilon), \Phi(s)) \geq \delta \right) \\
= & \mathbf{P}\left(X_t^\varepsilon \in C, \rho_{0,1}(Y(X_t^\varepsilon), \Phi(s)) \geq \delta, \rho_{0,1}(Y(X_t^\varepsilon), l_t^\varepsilon) < \delta) \right) \\
& + \mathbf{P}\left(X_t^\varepsilon \in C, \rho_{0,1}(Y(X_t^\varepsilon), \Phi(s)) \geq \delta, \rho_{0,1}(Y(X_t^\varepsilon), l_t^\varepsilon) \geq \delta) \right) \\
\leq & \mathbf{P}\left(X_t^\varepsilon \in C, S(l^\varepsilon) > s \right) + \mathbf{P}\left(\left\{ \sup_{0 \leq t \leq 1} |H(X_t^\varepsilon) - l_t^\varepsilon| \geq \delta \right\} \cap B \right) \\
\leq & \mathbf{P}\left(\bigcap_{k=0}^{n-1} A_k \bigcap \{S(l^\varepsilon) > s\} \bigcap B \right) + \mathbf{P}\left(\bigcup_{k=0}^{n-1} (\Omega \setminus A_k) \bigcap B \right) \\
& + \mathbf{P}\left(\left\{ \sup_{0 \leq t \leq 1} |H(X_t^\varepsilon) - l_t^\varepsilon| \geq \delta \right\} \bigcap B \right).
\end{aligned}
$$

We estimate the first term:

$$
\begin{aligned}
&\mathbf{P}\left(\bigcap_{k=0}^{n-1} A_k \bigcap \{S(l^\varepsilon) > s\} \bigcap B\right) \\
&= \mathbf{P}\left(\bigcap_{k=0}^{n-1} (A_k \cap B_k) \bigcap \left\{\frac{1}{2}\int_0^1 \frac{\left|{l_t^\varepsilon}'\right|^2}{B^2(l_t^\varepsilon)} dt > s\right\}\right) \\
&= \mathbf{P}\left(\bigcap_{k=0}^{n-1} (A_k \cap B_k) \bigcap \left\{\frac{n^2}{2}\sum_{k=0}^{n-1}\int_{\frac{k}{n}}^{\frac{k+1}{n}} \frac{\left|H(X^\varepsilon_{\frac{k+1}{n}}) - H(X^\varepsilon_{\frac{k}{n}})\right|^2}{B^2(l_t^\varepsilon)} dt > s\right\}\right) \\
&\leq \mathbf{P}\left(\bigcap_{k=0}^{n-1} (A_k \cap B_k) \bigcap \left\{\varepsilon^\beta \sum_{k=0}^{n-1} Y_k > s\right\}\right) \\
&\leq \frac{\mathbf{E}\left[\prod_{k=0}^{n-1} \chi_{A_k \cap B_k} \cdot \exp\left((1-\mu)Y_k\right)\right]}{\exp\left((1-\mu)s \cdot \varepsilon^{-\beta}\right)} \qquad (4.12) \\
&= \frac{\mathbf{E}\left[\prod_{k=0}^{n-2} \chi_{A_k \cap B_k} \cdot \exp\left((1-\mu)Y_k\right) \cdot \mathbf{E}\left(\chi_{B_{n-1}\cap A_{n-1}} \cdot \exp((1-\mu)Y_{n-1}) | \mathcal{F}_{\frac{n-1}{n}}\right)\right]}{\exp\left((1-\mu)s \cdot \varepsilon^{-\beta}\right)} \\
&\leq c_\mu \cdot \frac{\mathbf{E}\left[\prod_{k=0}^{n-2} \chi_{A_k \cap B_k} \cdot \exp\left((1-\mu)Y_k\right)\right]}{\exp\left((1-\mu)s \cdot \varepsilon^{-\beta}\right)} \qquad (4.13) \\
&\leq \frac{c_\mu^n}{\exp\left((1-\mu)s \cdot \varepsilon^{-\beta}\right)} \qquad (4.14) \\
&\leq \frac{1}{3}\exp(-(1-2\mu)s \cdot \varepsilon^{-\beta}) \\
&\leq \frac{1}{3}\exp(-\varepsilon^{-\beta}(s-\gamma)),
\end{aligned}
$$

where (4.12) follows from the exponential Chebyshev's inequality, and (4.13) follows from Lemma 4.7.9, and (4.14) can be obtained by applying Lemma 4.7.9 repeatedly. We estimate the second term using Lemma 4.7.6:

$$
\mathbf{P}\left(\bigcup_{k=0}^{n-1} (\Omega \setminus A_k) \bigcap B\right) \leq n \exp\left(-\frac{n{\delta'}^2}{64M\varepsilon^\beta}\right) \leq \frac{1}{3}\exp(-\varepsilon^{-\beta}(s-\gamma)),
$$

by recalling that $n\delta'^2 > 64M(s-\gamma)$. Similarly, we estimate the third term:

$$\begin{aligned}
\mathbf{P}\left(\left\{\sup_{0\le t\le 1}|H(X_t^\varepsilon)-l_t^\varepsilon|\ge\delta\right\}\bigcap B\right) &\le \sum_{k=0}^{n-1}\mathbf{P}\left(\left\{\sup_{\frac{k}{n}\le t\le\frac{k+1}{n}}|H(X_t^\varepsilon)-l_t^\varepsilon|\ge\delta\right\}\bigcap B\right)\\
&\le \sum_{k=0}^{n-1}\mathbf{P}\left(\left\{\sup_{0\le t\le\frac{1}{n}}\left|H(X^\varepsilon_{\frac{k}{n}+t})-H(X^\varepsilon_{\frac{k}{n}})\right|>\frac{\delta'}{2}\right\}\bigcap B\right)\\
&\le n\exp\left(-\frac{n\delta'^2}{64M\varepsilon^\beta}\right)\\
&\le \frac{1}{3}\exp(-\varepsilon^{-\beta}(s-\gamma)).
\end{aligned}$$

Thus,

$$\mathbf{P}(X_t^\varepsilon\in C,\rho_{0,1}(Y(X_t^\varepsilon),\Phi(s))\ge\delta)\le\exp(-\varepsilon^{-\beta}(s-\gamma)). \qquad \square$$

The next result, which relies on Lemma 4.4.1 and Theorem 4.7.4, gives an upper bound on probability that the process stays close to a given trajectory on the graph.

Lemma 4.4.2. *Suppose that* $\boldsymbol{\varphi}=(i,\varphi)\in\boldsymbol{C}([0,T],I_i)$ *is such that* $\inf_{t\in[0,T]}r(\varphi_t,H(O))>0$ *for each* $O\sim I_i$ *and* $S(\boldsymbol{\varphi})<\infty$*. Then, for every* $\gamma>0$*, there exist* $\delta>0$ *and* $\varepsilon_0>0$ *such that, for each* $\varepsilon\le\varepsilon_0$ *and each initial point* $x_0\in D_i(\varphi_0-\delta,\varphi_0+\delta)$*,*

$$\mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon),\boldsymbol{\varphi})\le\delta\right)\le\exp\left(-\varepsilon^{-\beta}(S(\boldsymbol{\varphi})-\gamma)\right).$$

Proof. Without loss of generality, we assume that $O_j\sim I_i$, $O_k\sim I_i$, and $H(\mathbf{x}_k)>H(\mathbf{x}_j)=0$. Suppose that $0<m'\le\varphi\le M'<H(\mathbf{x}_k)$. We define the compact set $C=\overline{D_i(m'/2,\frac{M'+H(\mathbf{x}_k)}{2})}\subset D_i$ and $s=S(\boldsymbol{\varphi})-\gamma/2$. Then, by Theorem 4.7.4, $\rho_{0,T}(\boldsymbol{\varphi},\Phi(s))>0$, where $\Phi(s)=\{\tilde{\boldsymbol{\varphi}}:S(\tilde{\boldsymbol{\varphi}})\le s,\ \tilde{\varphi}_0\in[m'/2,\frac{M'+H(\mathbf{x}_k)}{2}]\}$ (if $\Phi(s)$ is empty, then $\rho_{0,T}(\boldsymbol{\varphi},\Phi(s))=\infty$). We choose

$\delta > 0$ such that $\delta < \rho_{0,T}(\boldsymbol{\varphi}, \Phi(s))/2$ and $D_i(m' - \delta, M' + \delta) \subset C$. If $\rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi}) < \delta$, then

$$\rho_{0,T}(Y(X_t^\varepsilon), \Phi(s)) \geq \rho_{0,T}(\boldsymbol{\varphi}, \Phi(s)) - \rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi}) > \delta.$$

So, by Lemma 4.4.1, for ε small enough,

$$\begin{aligned}
\mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi})) < \delta\right) &= \mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi}) < \delta, X_t^\varepsilon \in C\right) \\
&\leq \mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon), \Phi(s)) > \delta, X_t^\varepsilon \in C\right) \\
&\leq \exp\left(-\varepsilon^{-\beta}(s - \gamma/2)\right) \\
&= \exp\left(-\varepsilon^{-\beta}(S(\boldsymbol{\varphi}) - \gamma)\right). \qquad \square
\end{aligned}$$

Theorem 4.4.3. *Given* $\boldsymbol{\varphi} \in \boldsymbol{C}([0,T], \Gamma)$ *with* $\boldsymbol{\varphi_0} = Y(x_0)$,

$$\overline{\lim_{\delta \to 0}}\, \overline{\lim_{\varepsilon \to 0}}\, \varepsilon^\beta \log \mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon), \boldsymbol{\varphi})) < \delta\right) \leq -S(\boldsymbol{\varphi}).$$

Proof. There exists a compact set C such that $\{x : \rho_{0,T}(Y(x), \boldsymbol{\varphi}) \leq \delta\} \subset C$. Let $\overline{M}$ be such that

$$|AH(x)| + |\nabla H(x)^* \sigma(x)|^2 < \overline{M}, \ \forall x \in C.$$

Let us consider the case where neither $\boldsymbol{\varphi}_0$ nor $\boldsymbol{\varphi}_T$ is at a vertex. Fix $\gamma > 0$. If $S(\boldsymbol{\varphi}) = \frac{1}{2}\int_0^T \frac{|\dot{\varphi}_t|^2}{B_{i_t}^2(\varphi_t)} dt < \infty$, then there exists $0 < \delta' < \inf_k |\varphi_0 - H(\mathbf{x}_k)| \wedge \inf_k |\varphi_T - H(\mathbf{x}_k)|$ so small that

$$\int_0^T \frac{|\dot{\varphi}_t|^2}{B_{i_t}^2(\varphi_t)} \cdot \chi_{\{\boldsymbol{\varphi}_t \in D(\pm\delta')\}} dt < \gamma.$$

Now let us define three sequences in $[0, T]$ inductively:

$$
\begin{aligned}
&q_0 = 0;\\
&t_k = \inf\left\{t > q_{k-1} : r(\boldsymbol{\varphi}_t, O) = \delta' \text{ for some vertex } O\right\},\\
&p_k = \sup\left\{t < t_k : r(\boldsymbol{\varphi}_t, O) = 2\delta' \text{ for some vertex } O\right\},\\
&q_k = \inf\left\{t > t_k : r(\boldsymbol{\varphi}_t, O) = 2\delta' \text{ for some vertex } O\right\},
\end{aligned}
$$

for $1 \le k \le n$, and n is such that the set $\{t > q_n : \boldsymbol{\varphi}_t \in D(\delta')\}$ is empty. Define $p_{n+1} = T$. The sequences are all finite due to the absolutely continuity of φ. Hence, by Lemma 4.4.2, there exist $\delta > 0$ and $\varepsilon_0 > 0$ such that, conditioned on $X^{\varepsilon}_{q_k} = x$, for all $\varepsilon \le \varepsilon_0$,

$$
\mathbf{P}\left(\rho_{q_k, p_{k+1}}(Y(X^{\varepsilon}_t), \boldsymbol{\varphi}) \le \delta\right) \le \exp\left(-\varepsilon^{-\beta}\left(\frac{1}{2}\int_{q_k}^{p_{k+1}} \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt - 2^{-k}\cdot\gamma\right)\right),
$$

since $\{t : r(\boldsymbol{\varphi}_t, O) \ge \delta' \text{ for every vertex } O \text{ on } \Gamma\} \supset \bigcup_{k=0}^{n}(q_k, p_{k+1})$. Thus, we obtain

$$
\begin{aligned}
\mathbf{P}\left(\rho_{0,T}(Y(X^{\varepsilon}_t), \boldsymbol{\varphi})) < \delta\right) &\le \mathbf{P}\left(\rho_{q_k, p_{k+1}}(H(X^{\varepsilon}_t), \varphi) < \delta,\ 0 \le k \le n\right)\\
&\le \exp\left(-\varepsilon^{-\beta}\left(\frac{1}{2}\int_{\bigcup_{k=0}^{n}(q_k, p_{k+1})} \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt - 2\gamma\right)\right)\\
&\le \exp\left(-\varepsilon^{-\beta}(S(\boldsymbol{\varphi}) - \gamma - 2\gamma)\right)\\
&= \exp\left(-\varepsilon^{-\beta}(S(\boldsymbol{\varphi}) - 3\gamma)\right)
\end{aligned}
$$

by the strong Markov property. Since we chose $\gamma > 0$ arbitrarily,

$$
\overline{\lim_{\delta \to 0}}\, \overline{\lim_{\varepsilon \to 0}}\, \varepsilon^{\beta} \log \mathbf{P}\left(\rho_{0,T}(Y(X^{\varepsilon}_t), \boldsymbol{\varphi})) < \delta\right) \le -S(\boldsymbol{\varphi}).
$$

In the case where $S(\boldsymbol{\varphi}) = \infty$, it is easy to show that, for every $s > 0$, there exist finitely many disjoint closed intervals $\{J_k\}$ such that $\boldsymbol{\varphi}$ stays away from all vertices on all J_k's and $\sum_k S_{J_k}(\boldsymbol{\varphi}) > s$. The result then follows in a similar way as in the finite action functional case. □

4.5 Exponential tightness

In this section, we establish the exponential tightness of the process $Y(X_t^\varepsilon)$ in $\mathbf{C}([0,T],\Gamma)$. We start with the definition before heading to the proof (cf. [46]).

Definition 4.5.1. *The family $Y(X_t^\varepsilon)$ is exponentially tight (with rate ε^β) in $\boldsymbol{C}([0,T],\Gamma)$ if there exists an increasing sequence of compact sets $(K_j)_{\{j\geq 1\}}$ in $\boldsymbol{C}([0,T],\Gamma)$ such that*

$$\lim_{j\to\infty}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \mathbf{P}(Y(X_t^\varepsilon)\in \boldsymbol{C}([0,T],\Gamma)\setminus K_j) = -\infty.$$

Our approach is based on Theorem 3.1 in [44]. It suffices to check the following conditions:

$$(i)\ \lim_{C\to\infty}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \mathbf{P}\left(\sup_{[0,T]}|H(X_t^\varepsilon)| > C\right) = -\infty,$$

$$(ii)\ \lim_{\delta\to 0}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \sup_{\tau\leq T-\delta}\mathbf{P}\left(\sup_{t\in[0,\delta]} r(Y(X_{\tau+t}^\varepsilon), Y(X_\tau^\varepsilon)) > \eta\right) = -\infty, \quad \forall \eta > 0,$$

where τ is a stopping time w.r.t. the filtration $\mathcal{F}$.

Theorem 4.5.2. *Under our assumptions, conditions (i) and (ii) are both valid, hence the family $Y(X_t^\varepsilon)$ is exponentially tight (with rate ε^β) in $\boldsymbol{C}([0,T],\Gamma)$.*

Proof. We present the proof in two steps. **i.** Let us check condition (i) first. By assumptions, there exists M large such that $H(x) \geq A_1|x|^2$ for $|x| \geq M$, $|AH(x)| + \|\sigma(x)\| \leq M$ for $x \in \mathbb{R}^2$, and $|\nabla H|$ is Lipschitz with constant M. For any large $C > A_1 M^2 + 2|H(x_0)| + A_1|x_0|^2 + 1$, define the stopping time $\tau = \inf\{t : H(X_t^\varepsilon) = C\} \wedge T$. Then we have

$$H(X_\tau^\varepsilon) = H(x_0) + \varepsilon^\beta \int_0^\tau AH(X_s^\varepsilon)ds + \varepsilon^{\beta/2} \int_0^\tau \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon) dW_s.$$

Let $F = \{x : H(x) \leq C\}$. Then, for $x \in F$, $|x| \leq \sqrt{\frac{C}{A_1}}$. In light of the boundedness of σ and of the second derivatives of H,

$$|\nabla H^*(x)\sigma(x)| \leq M^2|x - x_0| < M^2(\sqrt{\frac{C}{A_1}} + |x_0|) < 2M^2\sqrt{\frac{C}{A_1}}, \tag{4.15}$$

for each $x \in F$. From the fact $H(X_\tau^\varepsilon) = C$ we deduce that, on the event $\{\tau < T\}$,

$$\begin{aligned}\varepsilon^{\beta/2} \int_0^\tau \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon) dW_s &= C - H(x_0) - \varepsilon^\beta \int_0^\tau AH(X_s^\varepsilon)ds \\ &\geq C - H(x_0) - MT\varepsilon^\beta \\ &\geq \frac{1}{2}C,\end{aligned}$$

for ε sufficiently small. Thus, on the event $\{\tau < T\}$,

$$\begin{aligned}\int_0^T \chi_{\{\tau \geq s\}} \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon) dW_s &\geq \frac{1}{2}C\varepsilon^{-\beta/2} \\ \tilde{W}(\int_0^T |\chi_{\{\tau \geq s\}} \nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds) &\geq \frac{1}{2}C\varepsilon^{-\beta/2} \\ \sup_{[0, \frac{4M^4TC}{A_1}]} \tilde{W}_t &\geq \frac{1}{2}C\varepsilon^{-\beta/2},\end{aligned}$$

where $\tilde{W}$ is another Brownian motion. Hence $\{\tau < T\}$ implies that $\{\sup_{t\in[0,\frac{4M^4TC}{A_1}]} \tilde{W}_t \geq \frac{1}{2}C\varepsilon^{-\beta/2}\}$. As for the probability $\mathbf{P}\left(\sup_{[0,T]}|H(X_t^\varepsilon)| > C\right)$ for fixed C large enough,

$$
\begin{aligned}
\mathbf{P}\left(\sup_{[0,T]}|H(X_t^\varepsilon)| > C\right) &= \mathbf{P}\left(\tau < T\right) \\
&\leq \mathbf{P}\left(\sup_{[0,\frac{4M^4TC}{A}]} \tilde{W}_t \geq \frac{1}{2}C\varepsilon^{-\beta/2}\right) \\
&= 2\mathbf{P}\left(\tilde{W}_{\frac{4M^4TC}{A}} \geq \frac{1}{2}C\varepsilon^{-\beta/2}\right) \\
&\leq \sqrt{\frac{32M^4T\varepsilon^\beta}{\pi AC}}\exp\left(-\frac{AC}{32M^4T\varepsilon^\beta}\right).
\end{aligned}
$$

Therefore,

$$
\lim_{C\to\infty}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \mathbf{P}\left(\sup_{[0,T]}|H(X_t^\varepsilon)| > C\right) = -\infty
$$

ii. Since we've already checked the validity of the first condition, it suffices to instead prove that for every $\eta > 0$, $C > 0$,

$$
\lim_{\delta\to 0}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \sup_{\tau\leq T-\delta}\mathbf{P}\left(\sup_{t\in[0,\delta]} r(Y(X_{\tau+t}^\varepsilon), Y(X_\tau^\varepsilon)) > \eta, \sup_{[0,T]}|H(X_t^\varepsilon)| \leq C\right) = -\infty.
$$

Without loss of generality, we assume that η is so small that $\eta < |H(\mathbf{x}_k) - H(\mathbf{x}_j)|$ for any critical points $\mathbf{x}_j \neq \mathbf{x}_k$. Note that

$$
\left\{\sup_{t\in[0,\delta]} r(Y(X_{\tau+t}^\varepsilon), Y(X_\tau^\varepsilon)) > \eta\right\} \subset \left\{\sup_{t\in[0,\delta]} |H(X_{\tau+t}^\varepsilon) - H(X_\tau^\varepsilon)| > \eta/3\right\}.
$$

Thus, it suffices that show that, for any positive η and C,

$$\lim_{\delta\to 0}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^{\beta}\log\sup_{\tau\le T-\delta}\mathbf{P}\left(\sup_{t\in[0,\delta]}|H(X^{\varepsilon}_{\tau+t})-H(X^{\varepsilon}_{\tau})|>\frac{\eta}{3},\sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\right)=-\infty.$$

Since $AH(x)$ is bounded, this is implied by

$$\lim_{\delta\to 0}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^{\beta}\log\sup_{\tau\le T-\delta}\mathbf{P}\left(\sup_{t\in[0,\delta]}|\int_{\tau}^{\tau+t}\nabla H(X^{\varepsilon}_s)^*\sigma(X^{\varepsilon}_s)dW_s|>\frac{\eta}{4},\sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\right)=-\infty.$$

Define the process $Y^{\varepsilon}_t=\varepsilon^{\beta/2}\int_0^t\nabla H(X^{\varepsilon}_s)^*\sigma(X^{\varepsilon}_s)dW_s$, which is a martingale. Fix $\delta>0$. Then, for any stopping time $\tau\le T-\delta$, define a random change of time $\hat{\tau}_t:=\tau+t$ and a new process $\hat{Y}^{\varepsilon}_t:=Y_{\hat{\tau}_t}$ with a new filtration $\hat{\mathcal{F}}_t:=\mathcal{F}_{\hat{\tau}_t}$, $\forall t\ge 0$. By Theorem 1 from Chap.4, Sect.7 in [47], $\hat{Y}^{\varepsilon}_t$ is also a martingale, hence $Z^{\varepsilon}_t=\hat{Y}^{\varepsilon}_t-\hat{Y}^{\varepsilon}_0$ is martingale with filtration $\hat{\mathcal{F}}_t$. Furthermore, by Problem 2.28 from Chap.3 in [48], $\zeta^{\varepsilon}_t:=\exp(\lambda Z^{\varepsilon}_t-\frac{1}{2}\lambda^2\langle Z^{\varepsilon}\rangle_t)$ is supermartingale for each $\lambda\in\mathbb{R}$. Take a stopping time $\sigma:=\inf\{t\le\delta:|Z^{\varepsilon}_t|\ge\eta/4\}\wedge\delta$ w.r.t. $\hat{\mathcal{F}}_t$. By the Optional Sampling Theorem, $E\zeta^{\varepsilon}_{\sigma}\chi_{\{Z^{\varepsilon}_{\sigma}\ge\eta/4,\ \sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\}}\le E\zeta^{\varepsilon}_0=1$. Keeping (4.15) in mind, we conclude that

$$\mathbf{P}\left(Z^{\varepsilon}_{\sigma}\ge\eta/4,\ \sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\right)\le\exp\left(-\frac{\lambda\eta}{4}+\frac{\lambda^2}{2}\cdot\frac{4M^4C}{A_1}\cdot\delta\varepsilon^{\beta}\right)$$

holds for each $\lambda\in\mathbb{R}$. Thus, $\mathbf{P}\left(Z^{\varepsilon}_{\sigma}\ge\eta/4,\ \sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\right)\le\exp\left(-\frac{A_1\eta^2}{128M^4C\delta\varepsilon^{\beta}}\right)$. Similarly, one can prove $\mathbf{P}\left(Z^{\varepsilon}_{\sigma}\le-\eta/4,\ \sup_{[0,T]}|H(X^{\varepsilon}_t)|\le C\right)\le\exp\left(-\frac{A_1\eta^2}{128M^4C\delta\varepsilon^{\beta}}\right)$. So the second condition is also valid. $\square$

4.6 Proof of the main result

The following theorem is a result by Puhalskii adapted to our particular case (cf. Theorem 1.2 in [44], and also Corollary 3.4 in [45]).

Theorem 4.6.1. *If the family of process $Y(X_t^\varepsilon)$ is exponentially tight (with rate ε^β) and functional S satisfies the following condition for every $\boldsymbol{\varphi}$ with $\boldsymbol{\varphi}_0 = Y(x_0)$:*

$$\begin{aligned} S(\boldsymbol{\varphi}) &= -\lim_{\delta\to 0}\overline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon),\boldsymbol{\varphi})) < \delta\right) \\ &= -\lim_{\delta\to 0}\underline{\lim_{\varepsilon\to 0}}\,\varepsilon^\beta \log \mathbf{P}\left(\rho_{0,T}(Y(X_t^\varepsilon),\boldsymbol{\varphi})) < \delta\right), \end{aligned}$$

*then $\varepsilon^{-\beta}S(\boldsymbol{\varphi})$ is the action functional of the family of process $Y(X_t^\varepsilon)$ in **C** $([0,T],\Gamma)$ restricted to the set of functions that start at $Y(x_0)$.*

Proof of Theorem 4.2.2. The validity of the conclusion is due to Theorem 4.3.16, Theorem 4.4.3, Theorem 4.5.2, and Theorem 4.6.1. □

4.7 Technical proofs

Part 1. We prove the level set $\Phi_\delta(s) = \{r(\boldsymbol{\varphi_0}, Y(x_0)) \le \delta : S(\varphi) \le s\}$ is compact in $\mathbf{C}([0,T],\mathbb{R})$ for any $\delta \ge 0$, $s \ge 0$.

Lemma 4.7.1. *Suppose that $\varphi : [0,T] \to \mathbb{R}$ is absolutely continuous, and let $E = \{t : \varphi_t = c\}$, where c is a fixed constant. Then, for almost every point t in E (with respect to the Lebesgue measure), we have $\dot{\varphi}_t = 0$.*

Proof. Since φ is absolutely continuous on $[0,T]$, it's differentiable almost everywhere. Let F

be the subset of E where φ is not differentiable, so $\lambda(F) = 0$. For each $t_0 \in E \setminus F$, the limit

$$\lim_{h \to 0} \frac{\varphi_{t_0+h} - \varphi_{t_0}}{h}$$

exists. If $\dot{\varphi}_{t_0} \neq 0$, then $\varphi_t \neq \varphi_{t_0} = 0$ on $[t_0 - h, t_0 + h]$ for some positive h, which means that t_0 is an isolated point in E. There can be at most countably many isolated points inside a set of finite measure, so $G := \{t : \dot{\varphi}_t \text{ exists and is nonzero}\}$ has measure 0. Thus $E \setminus (F \cup G)$ is the set where $\dot{\varphi}_t = 0$, and $\lambda(F \cup G) = 0$. $\square$

Lemma 4.7.2. *The mapping $S : \boldsymbol{C}([0, T], \Gamma) \to \mathbb{R}$ is lower semicontinuous.*

Proof. Suppose that $\boldsymbol{\varphi}^{(l)} \to \boldsymbol{\varphi}$ in $\mathbf{C}([0, T], \Gamma)$. It is sufficient to consider the case where $\liminf_{l \to \infty} S(\boldsymbol{\varphi}^{(l)})$ is finite. For l large enough, $\varphi^{(l)}$ is uniformly bounded by some $M_1 \in \mathbb{R}$. Consequently, $B^2_{i^{(l)}}(\varphi^{(l)})$ is uniformly bounded by some $M_2 \in \mathbb{R}$. Using the Lemma from [49] on page 75, we evaluate

$$\begin{aligned}
&\sup_{0 \le t_0 < \ldots < t_N \le T} \sum_{i=1}^{N} \frac{|\varphi_{t_i} - \varphi_{t_{i-1}}|^2}{t_i - t_{i-1}} \\
&= \sup_{0 \le t_0 < \ldots < t_N \le T} \lim_{l \to \infty} \sum_{i=1}^{N} \frac{|\varphi^{(l)}_{t_i} - \varphi^{(l)}_{t_{i-1}}|^2}{t_i - t_{i-1}} \\
&\le \varliminf_{l \to \infty} \sup_{0 \le t_0 < \ldots < t_N \le T} \sum_{i=1}^{N} \frac{|\varphi^{(l)}_{t_i} - \varphi^{(l)}_{t_{i-1}}|^2}{t_i - t_{i-1}} \\
&= \varliminf_{l \to \infty} \int_0^T |\dot{\varphi}^{(l)}|^2 \\
&\le 2M_2 \varliminf_{l \to \infty} S(\boldsymbol{\varphi}^{(l)}) < \infty,
\end{aligned}$$

which implies that φ is absolutely continuous and has square integrable derivative. Now fix $\gamma > 0$, and define $U_\gamma = \{t : B^2_{i_t}(\varphi_t) > \gamma\}$. This is an open set that is a union of at most countably many open intervals $\{\tilde{I}_j : j \geq 1\}$ inside $[0,T]$. Take L such that $B_i^2(H)$ is Lipschitz continuous on $[\gamma, M_2]$ with constant L uniformly in i. Then, by the uniform continuity of φ, for each $\varepsilon > 0$, there exists $\delta > 0$ such that $|\varphi_t - \varphi_s| < \gamma\varepsilon/L$ whenever $t, s \in \tilde{I}_j$ and $|t-s| < \delta$. Notice that on a single interval $\tilde{I}_j$, i_t remains the same. Hence $t, s \in \tilde{I}_j$ and $|t-s| < \delta$ implies

$$\frac{B_{i_s}^2(\varphi_s)}{B_{i_t}^2(\varphi_t)} \leq 1 + \varepsilon.$$

Take n large enough so that $T/n < \delta$, and, for each $\tilde{I}_j = (a,b)$, let $a_k = a + \frac{k}{n}(b-a)$. Again, i_t remains the same on $\tilde{I}_j$, so

$$\begin{aligned}
\int_a^b \frac{|\dot{\varphi}_t|^2}{B_i^2(\varphi_t)} dt &\leq (1+\varepsilon) \sum_{k=0}^{n-1} \frac{1}{B_i^2(\varphi_{a_k})} \int_{a_k}^{a_{k+1}} |\dot{\varphi}_t|^2 dt \\
&= (1+\varepsilon) \sum_{k=0}^{n-1} \frac{1}{B_i^2(\varphi_{a_k})} \sup_{a_k \leq t_k^0 < \ldots < t_k^{M_k} \leq a_{k+1}} \sum_{i=1}^{M_k} \frac{|\varphi_{t_k^i} - \varphi_{t_k^{i-1}}|^2}{t_k^i - t_k^{i-1}} \\
&= (1+\varepsilon) \sum_{k=0}^{n-1} \frac{1}{B_i^2(\varphi_{a_k})} \sup_{a_k \leq t_k^0 < \ldots < t_k^{M_k} \leq a_{k+1}} \lim_{l\to\infty} \sum_{i=1}^{M_k} \frac{|\varphi^{(l)}_{t_k^i} - \varphi^{(l)}_{t_k^{i-1}}|^2}{t_k^i - t_k^{i-1}} \\
&\leq (1+\varepsilon) \sum_{k=0}^{n-1} \frac{1}{B_i^2(\varphi_{a_k})} \varliminf_{l\to\infty} \sup_{a_k \leq t_k^0 < \ldots < t_k^{M_k} \leq a_{k+1}} \sum_{i=1}^{M_k} \frac{|\varphi^{(l)}_{t_k^i} - \varphi^{(l)}_{t_k^{i-1}}|^2}{t_k^i - t_k^{i-1}} \\
&= (1+\varepsilon) \sum_{k=0}^{n-1} \frac{1}{B_i^2(\varphi_{a_k})} \varliminf_{l\to\infty} \int_{a_k}^{a_{k+1}} |\dot{\varphi}^{(l)}_t|^2 \\
&\leq (1+\varepsilon)^2 \sum_{k=0}^{n-1} \varliminf_{l\to\infty} \int_{a_k}^{a_{k+1}} \frac{|\dot{\varphi}^{(l)}_t|^2}{B_i^2(\varphi^{(l)}_t)} dt \\
&\leq (1+\varepsilon)^2 \varliminf_{l\to\infty} \int_a^b \frac{|\dot{\varphi}^{(l)}_t|^2}{B_i^2(\varphi^{(l)}_t)} dt.
\end{aligned}$$

Since ε was chosen arbitrarily, we obtain $\int_a^b \frac{|\dot{\varphi}_t|^2}{B_i^2(\varphi_t)} dt \leq \liminf_{l\to\infty} \int_a^b \frac{|\dot{\varphi}^{(l)}_t|^2}{B_i^2(\varphi^{(l)}_t)} dt$, which implies that $\int_{U_\gamma} \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt \leq \liminf_{l\to\infty} \int_0^T \frac{|\dot{\varphi}_t^{(l)}|^2}{B^2_{i_t}(\varphi_t^{(l)})} dt$. Since this is true for every $\gamma > 0$, we have

$$\int_U \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt \leq \varliminf_{l\to\infty} \int_0^T \frac{|\dot{\varphi}_t^{(l)}|^2}{B^2_{i_t}(\varphi_t^{(l)})} dt,$$

where $U = \{t : B^2_{i_t}(\varphi_t) > 0\}$. By convention $0/0 = 0$ and Lemma 4.7.1, we have

$$\int_{[0,T]\setminus U} \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt = 0,$$

which implies that

$$S(\boldsymbol{\varphi}) = \frac{1}{2} \int_0^T \frac{|\dot{\varphi}_t|^2}{B^2_{i_t}(\varphi_t)} dt \leq \varliminf_{l\to\infty} \frac{1}{2} \int_0^T \frac{|\dot{\varphi}_t^{(l)}|^2}{B^2_{i_t}(\varphi_t^{(l)})} dt = \varliminf_{l\to\infty} S(\boldsymbol{\varphi}^{(l)}).$$

So S is lower semicontinuous. $\square$

Lemma 4.7.3. *The set $\Phi_\delta(s) = \{r(\boldsymbol{\varphi}_0, Y(x_0)) \leq \delta : S(\boldsymbol{\varphi}) \leq s\}$ is precompact in $\boldsymbol{C}([0,T],\Gamma)$ for any $\delta \geq 0$, $s \geq 0$.*

Proof. We first prove the uniform boundedness of $\Phi_\delta(s)$. Since H has bounded second derivatives, $|\nabla H(x)|$ is Lipschitz continuous with coefficient $L > 0$. By assumption 5, we can assume that $\|\sigma\| < M$. And due to assumption 2, we can find $M_1 > |x_0|$ such that $H(x) \geq A_1|x|^2$ for all $|x| \geq M_1 - |x_0|$, $M_1 > \frac{2|\nabla H(x_0)|}{L}$, and $L(M_1 - |x_0|) > |\nabla H(0)|$. Notice that

$$|H(x)| \leq H(x_0) + \frac{1}{2}L|x - x_0|^2 + |\nabla H(x_0)| \cdot |x - x_0|.$$

Hence, $|H(x)| \geq |H(x_0)| + M_1^2 L$ implies that $|x - x_0| \geq M_1$, and therefore $|H(x)| \geq A_1|x|^2$.

Let $M_2 = |H(x_0)| + \delta + 1 + M_1^2 L + \frac{16sM^2TL^2}{A_1}$. Suppose that $\Phi_\delta(s)$ is not uniformly bounded, so there exists $\boldsymbol{\varphi} = (i_t, \varphi_t) \in \Phi_\delta(s)$ such that $\sup_{[0,T]} |\varphi| > 2M_2$. Let $\tau = \inf\{t : |\varphi_t| = 2M_2\} > 0$ and $\tau_1 = \sup\{t < \tau, |\varphi_t| = M_2\} > 0$. For $t \in [\tau_1, \tau]$, $|\varphi_t| \geq M_2 \geq |H(x_0)| + M_1^2 L$, which implies that $|x - x_0| \geq M_1$ for all $x \in C_{i_t}(\varphi_t)$. Thus, we obtain

$$B_{i_t}^2(\varphi_t) \leq \sup_{x \in C_{i_t}(\varphi_t)} |\nabla H^*(x)\sigma(x)|^2 \leq M^2 \sup_{x \in C_{i_t}(\varphi_t)} (L|x| + |\nabla H(0)|)^2 \leq \frac{4M^2L^2}{A_1}|\varphi_t|.$$

Thus

$$\begin{aligned}
M_2^2 &= |\varphi_\tau - \varphi_{\tau_1}|^2 \\
&= \left| \int_{\tau_1}^{\tau} \dot{\varphi}_t dt \right|^2 \\
&\leq \int_{\tau_1}^{\tau} \frac{|\dot{\varphi}_t|^2}{B_{i_t}^2(\varphi_t)} dt \cdot \int_{\tau_1}^{\tau} B_{i_t}^2(\varphi_t) dt \\
&\leq 2s \cdot T \cdot \frac{4M^2L^2}{A_1} \cdot 2M_2 \\
&= \frac{16M^2sTL^2}{A_1} M_2.
\end{aligned}$$

However, by definition, $M_2 > \frac{16M^2sTL^2}{A_1}$, which leads to a contradiction. Next, we prove the equicontinuity of $\Phi_\delta(s)$. Since $\Phi_\delta(s)$ is uniformly bounded, we assume that $B_{i_t}^2(\varphi_t) \leq M_3$ for each $\varphi \in \Phi_\delta(s)$. Instantly, we obtain

$$r(\boldsymbol{\varphi}_{t+h}, \boldsymbol{\varphi}_t)^2 \leq \left(\int_t^{t+h} |\dot{\varphi}_t| dt \right)^2 \leq \int_t^{t+h} \frac{|\dot{\varphi}_t|^2}{B_{i_t}^2(\varphi_t)} dt \cdot \int_t^{t+h} B_{i_t}^2(\varphi_t) dt \leq h \cdot 2sM_3.$$

Thus

$$r(\boldsymbol{\varphi}_{t+h}, \boldsymbol{\varphi}_t) \leq \sqrt{2sM_3} \cdot \sqrt{h}.$$

The precompactness of $\Phi_\delta(s)$ now follows from the Arzelà–Ascoli theorem. □

Theorem 4.7.4. *The set $\Phi_\delta(s) = \{r(\boldsymbol{\varphi_0}, Y(x_0)) \leq \delta : S(\varphi) \leq s\}$ is compact in $\boldsymbol{C}([0,T],\mathbb{R})$ for any $\delta \geq 0$, $s \geq 0$.*

Proof. This is a direct consequence of Lemmas 4.7.2 and 4.7.3. □

Part 2. We make two claims about the behavior of the random process during a short period of time.

Lemma 4.7.5. *Let the assumptions of Lemma 4.3.1 be satisfied and event A_k be defined as in the proof of Lemma 4.3.1. Then, for ε sufficiently small,*

$$\mathbf{P}\left(\Omega \setminus A_k\right) \leq \exp\left(-\frac{n\delta'^2}{64M\varepsilon^\beta}\right),$$

for every initial point $X^\varepsilon_{\frac{k}{n}} = x$ that satisfies $|H(x) - \varphi_{\frac{k}{n}}| < h$.

Proof. Without loss of generality, we assume that $k = 0$ in the proof. On the interval $[0, \frac{1}{n}]$, we have $|AH(X^\varepsilon_t)| < M$, so the increment in $H(X^\varepsilon_t)$ due to the ordinary integral is negligible in the sense that

$$\begin{aligned}|H(X^\varepsilon_t) - H(X^\varepsilon_0)| &\leq \varepsilon^\beta \int_0^t |AH(X^\varepsilon_s)|ds + \varepsilon^{\beta/2}\left|\int_0^t \nabla H(X^\varepsilon_s)^*\sigma(X^\varepsilon_s)dW_s\right| \\ &\leq \varepsilon^\beta \cdot \frac{M}{n} + \varepsilon^{\beta/2}\left|\int_0^t \nabla H(X^\varepsilon_s)^*\sigma(X^\varepsilon_s)dW_s\right|.\end{aligned}$$

Hence by recalling that, in Lemma 4.3.1, C was defined as a compact set contained in D_i, for ε

small enough, we have that

$$
\begin{aligned}
&\mathbf{P}(\sup_{0\le t\le \frac{1}{n}} |H(X_t^\varepsilon) - H(X_0^\varepsilon)| \le \frac{\delta'}{2}) \\
&\ge \mathbf{P}(\sup_{0\le t\le \frac{1}{n}} \varepsilon^{\beta/2} \left| W^{(1)}(\int_0^t |\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 \, ds\right|) \le \frac{\delta'}{4}, X_t^\varepsilon \in C) \\
&\ge \mathbf{P}(\sup_{0\le t\le \frac{M}{n}} |W_t^{(1)}| \le \frac{\delta'}{4\varepsilon^{\beta/2}}, X_t^\varepsilon \in C) \\
&= \mathbf{P}(\sup_{0\le t\le \frac{M}{n}} |W_t^{(1)}| \le \frac{\delta'}{4\varepsilon^{\beta/2}}) \\
&\ge 1 - \frac{8\sqrt{2M\varepsilon^\beta}}{\sqrt{\pi n}\cdot \delta'} \exp(-\frac{n\delta'^2}{32M\varepsilon^\beta}) =: 1 - \lambda_1(\varepsilon),
\end{aligned}
$$

where $W_t^{(1)}$ is a one-dimensional Wiener process. Equivalently, we have

$$
\mathbf{P}\left(\sup_{0\le t\le \frac{1}{n}} |H(X_t^\varepsilon) - H(X_0^\varepsilon)| > \frac{\delta'}{2}\right) \le \lambda_1(\varepsilon).
$$

Let σ_i denote the i-th row of σ. For ε small enough,

$$
\begin{aligned}
&\mathbf{P}\left(\sup_{pT_0^\varepsilon \le t\le (p+1)T_0^\varepsilon} \varepsilon^{\beta/2} \left|\int_{pT_0^\varepsilon}^t \sigma(X_s^\varepsilon) dW_s\right| > M\cdot \varepsilon^{\frac{1-\beta}{4}}\right) \\
&\le \mathbf{P}\left(\bigcup_{j=1}^2 \{\sup_{pT_0^\varepsilon \le t\le (p+1)T_0^\varepsilon} \varepsilon^{\beta/2} \left|\int_{pT_0^\varepsilon}^t \sigma_j(X_s^\varepsilon) dW_s\right| > \frac{M\cdot \varepsilon^{\frac{1-\beta}{4}}}{\sqrt{2}}\}\right) \\
&\le \sum_{j=1}^2 \mathbf{P}\left(\sup_{pT_0^\varepsilon \le t\le (p+1)T_0^\varepsilon} \varepsilon^{\beta/2} \left|\int_{pT_0^\varepsilon}^t \sigma_j(X_s^\varepsilon) dW_s\right| > \frac{M\cdot \varepsilon^{\frac{1-\beta}{4}}}{\sqrt{2}}\right) \\
&\le \sum_{j=1}^2 \mathbf{P}\left(\sup_{pT_0^\varepsilon \le t\le (p+1)T_0^\varepsilon} \left|W_j^{(2)}(\varepsilon^\beta \int_{pT_0^\varepsilon}^t |\sigma_j(X_s^\varepsilon)|^2 ds)\right| > \frac{M\cdot \varepsilon^{\frac{1-\beta}{4}}}{\sqrt{2}}\right) \\
&\le \sum_{j=1}^2 \mathbf{P}\left(\sup_{0\le t\le M^2 T_0^\varepsilon \varepsilon^\beta} \left|W_j^{(2)}(t)\right| > \frac{M\cdot \varepsilon^{\frac{1-\beta}{4}}}{\sqrt{2}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\le 4\sum_{j=1}^{2}\mathbf{P}\left(W_j^{(2)}(M^2T_0^\varepsilon\varepsilon^\beta) > \frac{M\cdot\varepsilon^{\frac{1-\beta}{4}}}{\sqrt{2}}\right)\\
&\le 4\sum_{j=1}^{2}\frac{\sqrt{T(H(X_0^\varepsilon))}\cdot\varepsilon^{\frac{1+\beta}{4}}}{\sqrt{\pi}}\exp\left(-\frac{1}{4T(H(X_0^\varepsilon))\varepsilon^{\frac{1+\beta}{2}}}\right)\\
&= \frac{8\sqrt{T(H(X_0^\varepsilon))}\cdot\varepsilon^{\frac{1+\beta}{4}}}{\sqrt{\pi}}\exp\left(-\frac{1}{4T(H(X_0^\varepsilon))\varepsilon^{\frac{1+\beta}{2}}}\right),
\end{aligned}
$$

where $W_j^{(2)}$ is a one-dimensional Wiener process for each j. Particularly, we conclude that

$$
\begin{aligned}
&\mathbf{P}\left(\sup_{pT_0^\varepsilon\le t\le (p+1)T_0^\varepsilon}\varepsilon^{\beta/2}\left|\int_{pT_0^\varepsilon}^{t}\sigma(X_s^\varepsilon)dW_s\right| > M\cdot\varepsilon^{\frac{1-\beta}{4}}\right)\\
&\le \frac{8\sqrt{T(H(X_0^\varepsilon))}\cdot\varepsilon^{\frac{1+\beta}{4}}}{\sqrt{\pi}}\exp\left(-\frac{1}{4T(H(X_0^\varepsilon))\varepsilon^{\frac{1+\beta}{2}}}\right)\\
&\le \frac{8\sqrt{M}\cdot\varepsilon^{\frac{1+\beta}{4}}}{\sqrt{\pi}}\exp\left(-\frac{1}{4m\varepsilon^{\frac{1+\beta}{2}}}\right) =: \lambda_2(\varepsilon) < \lambda_1(\varepsilon).
\end{aligned}
$$

Since $|H(x)-\varphi_0| < h$, $T_0^\varepsilon \ge m\varepsilon^{1-\beta}$, by the discussions above, we know that the event A_0 happens with overwhelming probability when ε is small enough, in the sense that

$$
\begin{aligned}
&\mathbf{P}\left(\Omega\setminus A_0\right)\\
&\le \lambda_1(\varepsilon) + \mathbf{P}\left(\bigcup_{p=0}^{\lfloor\frac{1}{nT_0^\varepsilon}\rfloor}\sup_{pT_0^\varepsilon\le t\le (p+1)T_0^\varepsilon}\varepsilon^{\beta/2}\left|\int_{pT_0^\varepsilon}^{t}\sigma(X_s^\varepsilon)dW_s\right| > M\cdot\varepsilon^{\frac{1-\beta}{4}}\right)\\
&\le \lambda_1(\varepsilon) + \mathbf{P}\left(\bigcup_{p=0}^{\lfloor\frac{1}{mn\varepsilon^{1-\beta}}\rfloor}\sup_{pT_0^\varepsilon\le t\le (p+1)T_0^\varepsilon}\varepsilon^{\beta/2}\left|\int_{pT_0^\varepsilon}^{t}\sigma(X_s^\varepsilon)dW_s\right| > M\cdot\varepsilon^{\frac{1-\beta}{4}}\right)\\
&\le \frac{2}{mn\varepsilon^{1-\beta}}\cdot\lambda_1(\varepsilon)\\
&= \frac{2}{mn\varepsilon^{1-\beta}}\cdot\frac{8\sqrt{2M\varepsilon^\beta}}{\sqrt{\pi n}\cdot\delta'}\exp\left(-\frac{n\delta'^2}{32M\varepsilon^\beta}\right)\\
&\le \exp\left(-\frac{n\delta'^2}{64M\varepsilon^\beta}\right). \qquad \square
\end{aligned}
$$

Lemma 4.7.6. *Let the assumptions in Lemma 4.4.1 be satisfied and events A_k, B_k be defined as in the proof of Lemma 4.4.1. Then, for ε sufficiently small,*

$$\mathbf{P}\left((\Omega \setminus A_k) \cap B_k\right) \leq \exp\left(-\frac{n\delta'^2}{64M\varepsilon^\beta}\right).$$

Proof. The proof is almost the same as that of Lemma 4.7.5. □

Part 3. We formulate the next two lemmas to prepare for the proof of Lemma 4.3.3. Lemma 4.7.7 shows that the random process that starts at an exterior vertex does not get stuck there; and Lemma 4.7.8 shows that the trajectory that starts at an exterior vertex does not escape too fast. Recall the assumptions of Section 3.2: $\mathbf{x}_0 = (0, 0)$ (the origin on $\mathbb{R}^2$) is a local minimum point and $H(\mathbf{x}_0) = 0$.

Lemma 4.7.7. *With the same assumptions as in Lemma 4.3.3, in a small neighborhood $B(\mathbf{x}_0, r)$, we have that, for every $x \in B(\mathbf{x}_0, r) \setminus \{\mathbf{x}_0\}$,*

$$\frac{AH(x)}{2\sqrt{H(x)}} - \frac{|\nabla H^*(x)\sigma(x)|^2}{8\sqrt{H(x)^3}} > 0.$$

Proof. It suffices to prove that

$$4H(x)AH(x) - |\nabla H^*(x)\sigma(x)|^2 > 0$$

in $B(\mathbf{x}_0, r)$. By Taylor's expansion, in $B(\mathbf{x}_0, r)$ we have $H(x) = a\hat{x}^2 + b\hat{y}^2 + 2c\hat{x}\hat{y} + O(|x|^3)$, where $x = (\hat{x}, \hat{y})$. The Hessian matrix of H at $\mathbf{x}_0$ is positive-definite, hence $a, b > 0$, and $ab - c^2 > 0$. Since H is four times continuously differentiable, the derivatives can also be approximated in the following way:

$$\frac{\partial H}{\partial \hat{x}} = 2a\hat{x} + 2c\hat{y} + O(|x|^2), \qquad \frac{\partial H}{\partial \hat{y}} = 2b\hat{y} + 2c\hat{x} + O(|x|^2),$$
$$\frac{\partial^2 H}{\partial \hat{x}^2} = 2a + O(|x|), \quad \frac{\partial^2 H}{\partial \hat{y}^2} = 2b + O(|x|), \quad \frac{\partial^2 H}{\partial \hat{x}\partial \hat{y}} = 2c + O(|x|).$$

Define the matrix-valued function

$$b(x) = \begin{bmatrix} 2H\frac{\partial^2 H}{\partial \hat{x}^2} - \left(\frac{\partial H}{\partial \hat{x}}\right)^2 & 2H\frac{\partial^2 H}{\partial \hat{x}\partial \hat{y}} - \frac{\partial H}{\partial \hat{x}}\frac{\partial H}{\partial \hat{y}} \\ 2H\frac{\partial^2 H}{\partial \hat{x}\partial \hat{y}} - \frac{\partial H}{\partial \hat{x}}\frac{\partial H}{\partial \hat{y}} & 2H\frac{\partial^2 H}{\partial \hat{y}^2} - \left(\frac{\partial H}{\partial \hat{y}}\right)^2 \end{bmatrix}.$$

With the Taylor's expansions, we can compute, for $x \in B(\mathbf{x}_0, r)$ with r sufficiently small,

$$2H\frac{\partial^2 H}{\partial \hat{x}^2} - \left(\frac{\partial H}{\partial \hat{x}}\right)^2 = 4\hat{y}^2(ab - c^2) + O(|x|^3) > 0,$$
$$2H\frac{\partial^2 H}{\partial \hat{y}^2} - \left(\frac{\partial H}{\partial \hat{y}}\right)^2 = 4\hat{x}^2(ab - c^2) + O(|x|^3) > 0,$$
$$2H\frac{\partial^2 H}{\partial \hat{x}\partial \hat{y}} - \frac{\partial H}{\partial \hat{x}}\frac{\partial H}{\partial \hat{y}} = 4\hat{x}\hat{y}(c^2 - ab) + O(|x|^3).$$

It follows that $\det(b(x)) = O(|x|^5)$. Since $a(x) = \sigma(x)\sigma^*(x)$ is positive-definite and uniformly nondegenerate, there exists m such that $a_{11}a_{22} - a_{12}^2 > m$. Due to the continuity of $a(x)$, it's bounded on any compact set, so there exists $\gamma \in (0, 1)$ such that $|a_{12}(x)| < \sqrt{a_{11}(x)a_{22}(x) - m} < (1 - \gamma)\sqrt{a_{11}(x)a_{22}(x)}$ for every $x \in B(\mathbf{x}_0, 1)$. Let r be smaller than 1 and such that $b_{12}(x)^2 < (1 + \gamma)^2 b_{11}(x)b_{22}(x)$ for every $x \in B(\mathbf{x}_0, r)$. Therefore,

$$|a_{12}(x)b_{12}(x)| < (1 - \gamma^2)\sqrt{a_{11}(x)a_{22}(x)b_{11}(x)b_{22}(x)}.$$

It follows that

$$\sum_{i,j=1}^{2} a_{ij}(x)b_{ij}(x) > 0,$$

which completes the proof. $\square$

Lemma 4.7.8. *Suppose that $\boldsymbol{\varphi}_t = (i, \varphi_t) \in \boldsymbol{C}([0,T], I_i)$ is non-constant with $S(\boldsymbol{\varphi})$ finite, and $\boldsymbol{\varphi}_0$ is at an exterior vertex O. If $t_0 = \inf\{t : \boldsymbol{\varphi}_t \neq O\}$, then $\dot{\varphi}_{t_0+} = 0$.*

Proof. Without loss of generality, we assume that $t_0 = 0$ and $\varphi_0 = H(\mathbf{x}_0) = 0$, where $\mathbf{x}_0$ is the origin, and a local minimum is achieved there. Due to the positive-definiteness of the Hessian matrix at $\mathbf{x}_0$, we have $r, \lambda, \Lambda > 0$ such that, for every $x \in B(\mathbf{x}_0, r)$, we have the estimates: $H(x) \geq \lambda|x|^2$ and $|\nabla H(x)| \leq \Lambda|x|$. It follows that there exists $H_0 > 0$ such that, for every $H < H_0$, $C_i(H) \subseteq B(\mathbf{x}_0, r)$. Furthermore, let the eigenvalues of $\sigma\sigma^*$ be bounded from above by M. Thus, for $H < H_0$,

$$B_i^2(H) \leq \max_{x \in C_i(H)} |\nabla H(x)^* \sigma(x)|^2 \leq \max_{x \in C_i(H)} M|\nabla H(x)|^2 \leq \max_{x \in C_i(H)} M\Lambda^2|x|^2,$$

$$H \geq \max_{x \in C_i(H)} \lambda|x|^2$$

$$\Rightarrow \frac{B_i^2(H)}{H} \leq \frac{M\Lambda^2}{\lambda}.$$

A direct consequence of this estimate is that if t is so small that for every $s \leq t$, $\varphi_s < H_0$, then

$$\int_0^t B_i^2(\varphi_s)ds \leq \int_0^t \frac{B_i^2(\varphi_s)}{\varphi_s} \cdot \varphi_s ds \leq \frac{M\Lambda^2}{\lambda} \cdot \int_0^t \varphi_s ds.$$

By the Cauchy-Schwarz Inequality, when t is sufficiently small,

$$\begin{aligned}\left(\int_0^t |\dot{\varphi}_s| ds\right)^2 &\leq \int_0^t \frac{\dot{\varphi}_s^2}{B_i^2(\varphi_s)} ds \cdot \int_0^t B_i^2(\varphi_s) ds \\ &\leq \int_0^t \frac{\dot{\varphi}_s^2}{B_i^2(\varphi_s)} ds \cdot \frac{M\Lambda^2}{\lambda} \cdot \int_0^t \varphi_s ds \\ &\leq \int_0^t \frac{\dot{\varphi}_s^2}{B_i^2(\varphi_s)} ds \cdot \frac{M\Lambda^2}{\lambda} \cdot \int_0^t \int_0^s |\dot{\varphi}_u| du ds \\ &\leq \int_0^t \frac{\dot{\varphi}_s^2}{B_i^2(\varphi_s)} ds \cdot \frac{M\Lambda^2}{\lambda} \cdot t \int_0^t |\dot{\varphi}_s| ds.\end{aligned}$$

Thus, we can conclude that

$$\frac{\varphi_t}{t} \leq \frac{\int_0^t |\dot{\varphi}_s| ds}{t} \leq \frac{M\Lambda^2}{\lambda} \int_0^t \frac{\dot{\varphi}_s^2}{B_i^2(\varphi_s)} ds \to 0,$$

as $t \to 0$, which completes the proof. □

Part 4. Now we turn to the technical preparation for Section 3.3.

Verification of (4.7) *and* (4.8).

$$\begin{aligned}\left|\frac{\partial T}{\partial \mu}(\mu,\nu)\right| &= \left|\frac{1}{2}\int_\nu^l \frac{\partial}{\partial \mu}\left(\frac{\det(J_\psi(\sqrt{y^2+\mu^2-\nu^2}, y))}{\sqrt{y^2+\mu^2-\nu^2}}\right) dy\right| \\ &\leq \frac{\overline{M}}{2}\int_\nu^l \left(\frac{\mu}{y^2+\mu^2-\nu^2} + \frac{\mu}{(y^2+\mu^2-\nu^2)^{3/2}}\right) dy \\ &= \frac{\overline{M}}{2}\left(\left[\arctan\frac{y}{\sqrt{\mu^2-\nu^2}} \cdot \frac{\mu}{\sqrt{\mu^2-\nu^2}}\right]_\nu^l + \left[\frac{y}{\sqrt{y^2+\mu^2-\nu^2}} \cdot \frac{\mu}{\mu^2-\nu^2}\right]_\nu^l\right) \\ &\leq \overline{M}\left(\frac{\pi\mu}{\sqrt{\mu^2-\nu^2}} + \frac{\mu}{(\mu^2-\nu^2)\sqrt{1+\mu^2-\nu^2}} + \frac{|\nu|}{\mu^2-\nu^2}\right) \\ &\leq \frac{C}{H(\psi(\mu,\nu))}.\end{aligned}$$

$$
\begin{aligned}
|\frac{\partial T}{\partial \nu}(\mu,\nu)| &= \left| \frac{1}{2}\int_\nu^l \frac{\partial}{\partial \nu}\left(\frac{\det(J_\psi(\sqrt{y^2+\mu^2-\nu^2},y))}{\sqrt{y^2+\mu^2-\nu^2}}\right) dy - \frac{\det(J_\psi(\mu,\nu))}{2|\mu|}\right| \\
&\le \frac{\overline{M}}{2}\int_\nu^l \left(\frac{|\nu|}{y^2+\mu^2-\nu^2} + \frac{|\nu|}{(y^2+\mu^2-\nu^2)^{3/2}}\right) dy + \frac{\overline{M}}{2\mu} \\
&= \frac{\overline{M}}{2}\left(\left[\arctan\frac{y}{\sqrt{\mu^2-\nu^2}}\cdot\frac{|\nu|}{\sqrt{\mu^2-\nu^2}}\right]_\nu^l + \left[\frac{y}{\sqrt{y^2+\mu^2-\nu^2}}\cdot\frac{|\nu|}{\mu^2-\nu^2}\right]_\nu^l + \frac{1}{\mu}\right) \\
&\le \overline{M}\left(\frac{|\pi\nu|}{\sqrt{\mu^2-\nu^2}} + \frac{|\nu|}{(\mu^2-\nu^2)\sqrt{1+\mu^2-\nu^2}} + \frac{\nu^2}{\mu(\mu^2-\nu^2)} + \frac{1}{\mu}\right) \\
&= \overline{M}\left(\frac{|\pi\nu|}{\sqrt{\mu^2-\nu^2}} + \frac{|\nu|}{(\mu^2-\nu^2)\sqrt{1+\mu^2-\nu^2}} + \frac{\mu}{\mu^2-\nu^2}\right) \\
&\le \frac{C}{H(\psi(\mu,\nu))}.
\end{aligned}
$$
□

Proof of Lemma 4.3.4. We only give the proof of the first statement here, and the others can be proved similarly. Our approach here is based on the reflection principle. Let us start with the probability of a larger event,

$$
\begin{aligned}
&\mathbf{P}\left(\varepsilon^{\beta/2}W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|\right) \\
\ge\ &\mathbf{P}\left(\varepsilon^{\beta/2}W_T \in (-\varepsilon^{a\beta} - 4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|, -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|)\right) \\
\ge\ &\varepsilon^{a\beta}\frac{1}{\sqrt{2T\pi\varepsilon^\beta}}\exp\left(-\frac{(4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon| + \varepsilon^{a\beta})^2}{2T\varepsilon^\beta}\right),
\end{aligned}
$$

while the probability of the difference is estimated, with $\tilde\varepsilon := 4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon| + \frac{1}{2}a\beta|\log\varepsilon|\varepsilon^{a\beta}$, as follows:

$$
\begin{aligned}
&\mathbf{P}\left(\varepsilon^{\beta/2}W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|,\ \varepsilon^{\beta/2}W_t \ge \frac{1}{4}a\beta|\log\varepsilon|\varepsilon^{a\beta},\ \text{for some } t\in(0,T)\right) \\
=\ &\mathbf{P}\left(\varepsilon^{\beta/2}W_T > 4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon| + \frac{1}{2}a\beta|\log\varepsilon|\varepsilon^{a\beta}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi \cdot \varepsilon^{\beta} T}} \int_{\tilde{\varepsilon}}^{\infty} \exp(-\frac{x^2}{2\varepsilon^{\beta} T}) dx \\
&\leq \frac{1}{\sqrt{2\pi \cdot \varepsilon^{\beta} T}} \cdot \frac{\varepsilon^{\beta} T}{\tilde{\varepsilon}} \cdot \exp\left(-\frac{\tilde{\varepsilon}^2}{2T\varepsilon^{\beta}}\right),
\end{aligned}
$$

where the first equality follows from the reflection principle and the inequality follows from the relation: $\int_y^{\infty} \exp(-cx^2)dx \leq \frac{1}{2cy}\exp(-cy^2)$, for all $y, c > 0$. Consider the ratio of these two probabilities:

$$
\begin{aligned}
&\frac{\mathbf{P}\left(\varepsilon^{\beta/2} W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|,\ \varepsilon^{\beta/2} W_t \geq \frac{1}{4}a\beta|\log\varepsilon|\varepsilon^{a\beta},\ \text{for some } t \in (0,T)\right)}{\mathbf{P}\left(\varepsilon^{\beta/2} W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|\right)} \\
&\leq \frac{\varepsilon^{\beta} T}{\varepsilon^{a\beta}\tilde{\varepsilon}} \exp\left(\frac{(4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon| + \varepsilon^{a\beta})^2}{2T\varepsilon^{\beta}} - \frac{\tilde{\varepsilon}^2}{2T\varepsilon^{\beta}}\right) \to 0,\ \text{as } \varepsilon \to 0.
\end{aligned}
$$

Combining the estimates above, we obtain

$$
\begin{aligned}
&\mathbf{P}\left(\varepsilon^{\beta/2} W_T < -4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon|,\ \varepsilon^{\beta/2} W_t < \frac{1}{4}a\beta|\log\varepsilon|\varepsilon^{a\beta},\ \forall\, 0 < t < T\right) \\
&\geq \frac{1}{2}\varepsilon^{a\beta}\frac{1}{\sqrt{2T\pi\varepsilon^{\beta}}}\exp\left(-\frac{(4\varepsilon^{(a-d)\beta}(a-d)\beta|\log\varepsilon| + \varepsilon^{a\beta})^2}{2T\varepsilon^{\beta}}\right) \\
&\geq 2\exp\left(-\varepsilon^{-(1-2(a-d))\beta-\kappa}\right),
\end{aligned}
$$

for ε small enough. □

Proof of Lemma 4.3.5. Let us look at the slow motion for small ε:

$$
dx_t = \nabla^{\perp} H(x_t)dt;
$$

$$
d\tilde{X}_t^{\varepsilon} = \nabla^{\perp} H(\tilde{X}_t^{\varepsilon})dt + \sqrt{\varepsilon}\sigma(\tilde{X}_t^{\varepsilon})d\tilde{W}_t.
$$

Let F be the region between the level curves $\{x : H(x) = \frac{1}{2}\varepsilon^{a\beta}\}$ and $\{x : H(x) = \varepsilon^{(a-d)\beta}\}$, and between γ_{out} and γ_{in}, where x_t moves from γ_{in} to γ_{out}. The integrand is at least $c|\log\varepsilon|^2$ large (for some constant c) if the process is away from the critical point, and we would like to show that, with high probability, the process does not spend more than $C|\log\varepsilon|$ time (for some constant C) in F during a typical rotation. By (4.6), the deterministic process starting at any point in F spends no more than time $C|\log\varepsilon|$ in F before exit. We'll bound the exit time for the random process by approximating it with piecewise deterministic motion and will see that the small diffusion added does not significantly delay the time for the process to exit from F.

Define $T_1(x) = \inf\{t : x_t \notin F,\ x_0 = x\}$ and $T_2(x) = \inf\{t : x_t \in \gamma_{in},\ x_0 = x\}$. For the process starting at x, define stopping times $\eta^x = \inf\{t : \tilde{X}_t^\varepsilon \in C_{ki}(\varepsilon^{a\beta}/2) \cup C_{ki}(\varepsilon^{(a-d)\beta})\}$, $\tau_0^x = \inf\{t : \tilde{X}_t^\varepsilon \in \gamma_{out}\}$, $\tau_1^x = \inf\{t : \tilde{X}_t^\varepsilon \in \gamma_{in}\}$, $\tau_2^x = \inf\{t > \tau_1^x : \tilde{X}_t^\varepsilon \in \gamma_{out}\}$, and $\hat{\tau}_j^x = \tau_j^x \wedge \eta^x$, $j = 0, 1, 2$. Since σ is bounded, it's not hard to see that, for each $x \in F$ and $\tilde{X}^\varepsilon$ starting at x,

$$\mathbf{P}\left(\sup_{\Delta\in[0,1]} |\sqrt{\varepsilon}\int_t^{t+\Delta} \sigma(\tilde{X}_s^\varepsilon)dW_s| > \varepsilon^{\frac{1}{2}-\frac{1}{2}(1-2(a-d))\beta-2\kappa}\right) < \exp\left(-\varepsilon^{-(1-2(a-d))\beta-4\kappa}\right).$$

We claim that, with $\overline{M}$ defined at the beginning of the Section 3.3 and $x \in \gamma_{in}$,

$$E := \bigcap_{k=0}^{\overline{M}|\log\varepsilon|-1} \left\{\sup_{\Delta\in[0,1]} |\sqrt{\varepsilon}\int_k^{k+\Delta} \sigma(\tilde{X}_s^\varepsilon)dW_s| \le \varepsilon^{\frac{1}{2}-\frac{1}{2}(1-2(a-d))\beta-2\kappa}\right\} \subseteq \left\{\hat{\tau}_0^x \le \overline{M}|\log\varepsilon|\right\}.$$

To verify this, note that, by Gronwall's inequality, the event E is contained in

$$\bigcap_{k=0}^{\overline{M}|\log\varepsilon|-1} \left\{\sup_{\Delta\in[0,1]} |\tilde{X}_{k+\Delta}^\varepsilon - x_{k+\Delta}^k| \le e^L \cdot \varepsilon^{\frac{1}{2}-\frac{1}{2}(1-2(a-d))\beta-2\kappa}\right\},$$

where L is the Lipschitz constant of ∇H and x^k is the deterministic process with random starting point defined by

$$dx_t^k = \nabla^\perp H(x_t^k)dt,\ k \le t \le k+1,\ x_k^k = \tilde{X}_k^\varepsilon.$$

Then, by the recalling that $\|J_{\psi^{-1}}\| < \overline{M}$ and $|\nabla T(\psi^{-1}(x))| \le \frac{C}{|H(x)|}$, it follows that

$$\begin{aligned}
E &\subseteq \bigcap_{k=0}^{\overline{M}|\log\varepsilon|-1} \left\{ \sup_{\Delta\in[0,1]} |\tilde{X}_{k+\Delta}^\varepsilon - x_{k+\Delta}^k| \le e^L \cdot \varepsilon^{\frac{1}{2}-\frac{1}{2}(1-2(a-d))\beta-2\kappa} \right\} \\
&\subseteq \bigcap_{k=0}^{\overline{M}|\log\varepsilon|-1} \left\{ T_1(\tilde{X}_{k+1}^\varepsilon) \le (T_1(\tilde{X}_k^\varepsilon) - 1/2) \vee 0 \right\} \bigcup \left\{ \hat{\tau}_0^x \le \overline{M}|\log\varepsilon| \right\} \qquad (4.16) \\
&\subseteq \left\{ \hat{\tau}_0^x \le \overline{M}|\log\varepsilon| \right\}. \qquad (4.17)
\end{aligned}$$

Here we need to explain why the last two inclusions hold.

(1) Let us start with (4.16). For each $0 \le k \le \overline{M}|\log\varepsilon| - 1$, if $\sup_{\Delta\in[0,1]} |\tilde{X}_{k+\Delta}^\varepsilon - x_{k+\Delta}^k| \le e^L \cdot \varepsilon^{\frac{1}{2}-\frac{1}{2}(1-2(a-d))\beta-2\kappa}$, then one of the following must happen:

  (a) If $\tilde{X}_{k+1}^\varepsilon \notin F$, then $\hat{\tau}_0^x \le \overline{M}|\log\varepsilon|$.

  (b) If $\tilde{X}_{k+1}^\varepsilon \in F$ and $x_{k+1}^k \in F$, then $T_1(x_{k+1}^k) = T_1(\tilde{X}_k^\varepsilon) - 1$ and $|T_1(\tilde{X}_{k+1}^\varepsilon) - T_1(x_{k+1}^k)| = |T(\psi^{-1}(\tilde{X}_{k+1}^\varepsilon)) - T(\psi^{-1}(x_{k+1}^k))| = o(1)$, so $T_1(\tilde{X}_{k+1}^\varepsilon) \le (T_1(\tilde{X}_k^\varepsilon) - 1/2)$.

  (c) If $\tilde{X}_{k+1}^\varepsilon \in F$, $x_{k+1}^k \notin F$, and $H(x_{k+1}^k) \notin [\frac{1}{2}\varepsilon^{a\beta}, \varepsilon^{(a-d)\beta}]$, then $\tilde{X}_k^\varepsilon \notin F$ and $\hat{\tau}_0^x \le \overline{M}|\log\varepsilon|$.

  (d) If $\tilde{X}_{k+1}^\varepsilon \in F$, $x_{k+1}^k \notin F$, $H(x_{k+1}^k) \in [\frac{1}{2}\varepsilon^{a\beta}, \varepsilon^{(a-d)\beta}]$, and $x_{k+1}^k \in V_0 \setminus V$, then, as in (b), we still have that $T_1(\tilde{X}_{k+1}^\varepsilon) \le (T_1(\tilde{X}_k^\varepsilon) - 1/2)$.

  (e) If $\tilde{X}_{k+1}^\varepsilon \in F$, $x_{k+1}^k \notin F$, $H(x_{k+1}^k) \in [\frac{1}{2}\varepsilon^{a\beta}, \varepsilon^{(a-d)\beta}]$, and $x_{k+1}^k \notin V_0$, then we have a contradiction here, since ψ is a diffeomorphism, $|\tilde{X}^{\varepsilon}_{k+1} - x^k_{k+1}| \leq e^L \cdot \varepsilon^{\frac{1}{2} - \frac{1}{2}(1-2(a-d))\beta - 2\kappa}$, and there is a constant distance between ∂U and ∂U_0.

(2) The last inclusion (4.17) is justified by the fact that $T_1(x) < \overline{M}|\log H(x)| < \frac{1}{2}\overline{M}|\log \varepsilon| - 1$ on F for ε small enough.

Thus, by the strong Markov property,

$$
\begin{aligned}
\mathbf{P}(\hat{\tau}^x_0 \leq \overline{M}|\log \varepsilon|) \geq \mathbf{P}(E) &\geq \left(1 - \exp\left(-\varepsilon^{-(1-2(a-d))\beta - 4\kappa}\right)\right)^{\overline{M}|\log \varepsilon|} \\
&\geq 1 - \overline{M}|\log \varepsilon| \exp\left(-\varepsilon^{-(1-2(a-d))\beta - 4\kappa}\right) \\
&\geq 1 - \exp\left(-\varepsilon^{-(1-2(a-d))\beta - 2\kappa}\right).
\end{aligned}
$$

Now we look at one full rotation. For each $x \in \gamma_{out} \cap D_i(\frac{1}{2}\varepsilon^{a\beta}, \varepsilon^{(a-d)\beta})$ and deterministic and random processes starting at x, consider the events

$$
\begin{aligned}
E_1 &= \{\eta^x \leq \tau^x_1\} \cap \{\sup_{t \in [0, 2T_2(x)]} |\tilde{X}^{\varepsilon}_t - x_t| < \varepsilon^{\frac{1}{2} - \frac{1}{2}(1-2(a-d))\beta - 2\kappa}\}, \\
E_2 &= \{\eta^x > \tau^x_1\} \cap \{\sup_{t \in [0, 2T_2(x)]} |\tilde{X}^{\varepsilon}_t - x_t| < \varepsilon^{\frac{1}{2} - \frac{1}{2}(1-2(a-d))\beta - 2\kappa}\} \cap \{\hat{\tau}^x_2 - \hat{\tau}^x_1 \leq \overline{M}|\log \varepsilon|\}, \\
E_3 &= \{\eta^x > \tau^x_1\} \cap \{\sup_{t \in [0, 2T_2(x)]} |\tilde{X}^{\varepsilon}_t - x_t| < \varepsilon^{\frac{1}{2} - \frac{1}{2}(1-2(a-d))\beta - 2\kappa}\} \cap \{\hat{\tau}^x_2 - \hat{\tau}^x_1 > \overline{M}|\log \varepsilon|\}.
\end{aligned}
$$

By the strong Markov property of $\tilde{X}^{\varepsilon}_t$, $\mathbf{P}(E_3) \leq \exp\left(-\varepsilon^{-(1-2(a-d))\beta - 2\kappa}\right)$. Furthermore, note that

$$
\begin{aligned}
\mathbf{P}(E_1 \sqcup E_2 \sqcup E_3) &= \mathbf{P}\left(\sup_{t \in [0, 2T_2(x)]} |\tilde{X}^{\varepsilon}_t - x_t| < \varepsilon^{\frac{1}{2} - \frac{1}{2}(1-2(a-d))\beta - 2\kappa}\right) \\
&> 1 - \exp\left(-\varepsilon^{-(1-2(a-d))\beta - 4\kappa}\right),
\end{aligned}
$$

where last inequality follows from the fact that $T_2(x)$ is bounded by a constant on γ_{out}. We deduce that

$$\mathbf{P}(E_1 \cup E_2) \geq 1 - 2\exp\left(-\varepsilon^{-(1-2(a-d))\beta-2\kappa}\right). \tag{4.18}$$

Observe that, on the event $E_1 \cup E_2$, $\int_0^t |\log H(\tilde{X}_s^\varepsilon)\nabla H(\tilde{X}_s^\varepsilon)^*\sigma(\tilde{X}_s^\varepsilon)|^2 ds > t$ holds for all $t \leq \hat{\tau}_2^x$. Using the strong Markov property of $X_t^\varepsilon = \tilde{X}_{t\varepsilon^{\beta-1}}^\varepsilon$ and (4.18) for every rotation prior to $\eta \wedge T$ (here the term "rotation" means that the process travels from γ_{out} to γ_{in} and then back to γ_{out}), we obtain the desired result by noting that the number of rotations N is then bounded by $C\varepsilon^{\beta-1}$ provided that the analogue of $E_1 \cup E_2$ holds for every rotation,

$$\begin{aligned}
&\mathbf{P}\left(\int_0^t |\log H(X_s^\varepsilon)\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2 ds > t,\ \forall\, t \leq \eta \wedge T\right) \\
&\geq \left(1 - 2\exp\left(-\varepsilon^{-(1-2(a-d))\beta-2\kappa}\right)\right)^N \\
&\geq 1 - 2N\exp\left(-\varepsilon^{-(1-2(a-d))\beta-2\kappa}\right) \\
&\geq 1 - \exp\left(-\varepsilon^{-(1-2(a-d))\beta-\kappa}\right). \qquad \square
\end{aligned}$$

Part 5. The last lemma is needed for the proof in Section 4.1.

Lemma 4.7.9. *With the same assumptions as in Lemma 4.4.1, we have, for every $0 \leq k < n$,*

$$\mathbf{E}\left(\chi_{B_k \cap A_k} \cdot \exp\left(\frac{1-\mu}{2\varepsilon^\beta} \cdot \frac{n}{B^2\left(H(X_{\frac{k}{n}}^\varepsilon)\right) - \frac{L\delta'}{2}} \left|H(X_{\frac{k+1}{n}}^\varepsilon) - H(X_{\frac{k}{n}}^\varepsilon)\right|^2\right) \mid \mathcal{F}_{\frac{k}{n}}\right) \leq C_\mu < \infty,$$

where C_μ is a constant that depends solely on μ.

Proof. Without loss of generality, we can assume that $k = 0$, the initial point is $x \in C$, and drop the conditioning on $\mathcal{F}_{\frac{k}{n}}$. Note that

$$|H(X^{\varepsilon}_{1/n}) - H(x)| \leq \varepsilon^{\beta} \left| \int_0^{1/n} AH(X^{\varepsilon}_s) ds \right| + \varepsilon^{\beta/2} \left| \int_0^{1/n} \nabla H(X^{\varepsilon}_s)^* \sigma(X^{\varepsilon}_s) dW_s \right|. \tag{4.19}$$

The first term on the right-hand side is negligible compared to the second term, which needs to be studied further. Note that, on the event B_0, $H(X^{\varepsilon}_t) \in H(C)$ for $0 \leq t \leq 1/n$. For each $0 \leq p \leq \lfloor \frac{1}{nT^{\varepsilon}_0} \rfloor$ and $t \geq pT^{\varepsilon}_0$, define the deterministic process with the random starting point:

$$d\xi^{\varepsilon}_t = \varepsilon^{\beta-1} \nabla^{\perp} H(\xi^{\varepsilon}_t) dt, \quad \xi^{\varepsilon}_{pT^{\varepsilon}_0} = X^{\varepsilon}_{pT^{\varepsilon}_0}.$$

Then, on the event $A_0 \cap B_0$,

$$\begin{aligned} |X^{\varepsilon}_t - \xi^{\varepsilon}_t| &\leq \varepsilon^{\beta-1} \int_{pT^{\varepsilon}_0}^{t} \left| \nabla H^{\perp}(X^{\varepsilon}_s) - \nabla H^{\perp}(\xi^{\varepsilon}_s) \right| ds + \varepsilon^{\beta/2} \left| \int_{pT^{\varepsilon}_0}^{t} \sigma(X^{\varepsilon}_s) dW_s \right| \\ &\leq \varepsilon^{\beta-1} \int_{pT^{\varepsilon}_0}^{t} L \left| X^{\varepsilon}_s - \xi^{\varepsilon}_s \right| ds + M \cdot \varepsilon^{\frac{1-\beta}{4}}. \end{aligned}$$

So, for ε small enough, by Gronwall's inequality,

$$|X^{\varepsilon}_t - \xi^{\varepsilon}_t| \leq M \exp(L \cdot T(H(x))) \cdot \varepsilon^{\frac{1-\beta}{4}} \leq M \exp(LM) \cdot \varepsilon^{\frac{1-\beta}{4}} \leq \frac{\delta'}{M}. \tag{4.20}$$

Using (4.20) and Lipschitz continuity and boundedness of $|\nabla H(x)^* \sigma(x)|^2$, $B^2(H)$, and $T(H)$, we estimate $|\nabla H(X^{\varepsilon}_s)^* \sigma(X^{\varepsilon}_s)|^2$ in terms of $B^2(X^{\varepsilon}_s)$ on the interval $[pT^{\varepsilon}_0, (p+1)T^{\varepsilon}_0]$, on the event

$A_0 \cap B_0$:

$$\begin{aligned}
&\int_{pT_0^\varepsilon}^{(p+1)T_0^\varepsilon} |\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds \\
&\le \int_{pT_0^\varepsilon}^{(p+1)T_0^\varepsilon} |\nabla H(\xi_s^\varepsilon)^* \sigma(\xi_s^\varepsilon)|^2 ds + \varepsilon^{1-\beta} T(H(x)) L \cdot \frac{\delta'}{M} \\
&\le \int_{pT_0^\varepsilon}^{pT_0^\varepsilon + \varepsilon^{1-\beta} T(H(X_{pT_0^\varepsilon}^\varepsilon))} |\nabla H(\xi_s^\varepsilon)^* \sigma(\xi_s^\varepsilon)|^2 ds + \varepsilon^{1-\beta} \cdot L\delta' M + \varepsilon^{1-\beta} T(H(x)) L \cdot \frac{\delta'}{M} \\
&\le \varepsilon^{1-\beta} T(H(X_{pT_0^\varepsilon}^\varepsilon)) B^2(H(X_{pT_0^\varepsilon}^\varepsilon)) + \varepsilon^{1-\beta} \cdot L\delta'(M+1) \\
&\le T_0^\varepsilon B^2(H(X_{pT_0^\varepsilon}^\varepsilon)) + \varepsilon^{1-\beta} \cdot L\delta'(2M+1).
\end{aligned}$$

This is valid for every $0 \le p \le \lfloor \frac{1}{nT_0^\varepsilon} \rfloor$. Combining the contributions from all the time intervals, since $B^2(H)$ is Lipschitz and δ' is chosen to be small enough, it follows that on the event $A_0 \cap B_0$:

$$\begin{aligned}
&\int_0^{1/n} |\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds \\
&\le \sum_{p=0}^{\lfloor \frac{1}{nT_0^\varepsilon} \rfloor - 1} \int_{pT_0^\varepsilon}^{(p+1)T_0^\varepsilon} |\nabla H(X_s^\varepsilon)^* \sigma(X_s^\varepsilon)|^2 ds + T_0^\varepsilon M \\
&\le \sum_{p=0}^{\lfloor \frac{1}{nT_0^\varepsilon} \rfloor - 1} \Big(T_0^\varepsilon B^2(H(X_{pT_0^\varepsilon}^\varepsilon)) + \varepsilon^{1-\beta} \cdot L\delta'(2M+1) \Big) + T_0^\varepsilon M \\
&\le \frac{1+\mu}{n} \left[B^2(H(x)) + \frac{L\delta'}{2} \right],
\end{aligned}$$

for ε sufficiently small. Let $Q = \frac{1+\mu}{n} \left[B^2(H(x)) + \frac{L\delta'}{2} \right]$. Since $|AH(x)| \le M$ in C and (4.19) holds,

$$\mathbf{E} \left(\chi_{B_0 \cap A_0} \cdot \exp \left(\frac{1-\mu}{2\varepsilon^\beta} \cdot \frac{n}{B^2\left(H(x)\right) - \frac{L\delta'}{2}} \left| H(X_{1/n}^\varepsilon) - H(x) \right|^2 \right) \right)$$

$$\leq \mathbf{E}\left(\chi_{B_0\cap A_0}\cdot\exp\left(\frac{n(1-\mu)}{2B^2\left(H(x)\right)-L\delta'}\left(\left|\int_0^{\frac{1}{n}}\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)dW_s\right|+\frac{M}{n}\varepsilon^{\beta/2}\right)^2\right)\right)$$
$$=\mathbf{E}\left(\chi_{B_0\cap A_0}\cdot\exp\left(\frac{n(1-\mu)}{2B^2\left(H(x)\right)-L\delta'}\left(\left|\tilde{W}\left(\int_0^{\frac{1}{n}}|\nabla H(X_s^\varepsilon)^*\sigma(X_s^\varepsilon)|^2ds\right)\right|+\frac{M}{n}\varepsilon^{\beta/2}\right)^2\right)\right)$$
$$\leq \mathbf{E}\left(\chi_{B_0\cap A_0}\cdot\exp\left(\frac{1-\mu}{2}\cdot\frac{n}{B^2\left(H(x)\right)-\frac{L\delta'}{2}}\cdot\left(\sup_{0\leq t\leq Q}|\tilde{W}(t)|+\frac{M}{n}\varepsilon^{\beta/2}\right)^2\right)\right)$$
$$\leq \mathbf{E}\left(\exp\left(\frac{1-\mu^2}{2}\cdot\frac{B^2\left(H(x)\right)+\frac{L\delta'}{2}}{B^2\left(H(x)\right)-\frac{L\delta'}{2}}\cdot\frac{1}{Q}\left(\sup_{0\leq t\leq Q}|\tilde{W}(t)|+\frac{M}{n}\varepsilon^{\beta/2}\right)^2\right)\right)$$
$$\leq \mathbf{E}\left(\exp\left(\frac{1-\mu^2/2}{2}\cdot\frac{1}{Q}\left(\sup_{0\leq t\leq Q}|\tilde{W}(t)|+\frac{M}{n}\varepsilon^{\beta/2}\right)^2\right)\right)$$
$$\leq \mathbf{E}\left(\exp\left(\frac{1-\mu^2/3}{2}\cdot\frac{1}{Q}\sup_{0\leq t\leq Q}|\tilde{W}(t)|^2\right)\right)+\exp\left(\frac{1-\mu^2/2}{2}\right),$$

where the last inequality follows, for ε small, by integrating over the sets $\{\sup_{0\leq t\leq Q}|\tilde{W}(t)|>\sqrt{Q}/2\}$ and $\{\sup_{0\leq t\leq Q}|\tilde{W}(t)|\leq\sqrt{Q}/2\}$ separately. For every $a>0$, by the reflection principle,

$$\mathbf{P}\left(\sup_{0\leq t\leq Q}|\tilde{W}(t)|^2>a^2\right)\leq 4\mathbf{P}\left(\tilde{W}_Q>a\right)=2\mathbf{P}\left(|\tilde{W}_Q|^2>a^2\right).$$

Therefore, for every $a>0$, $\mathbf{P}(\xi>a)\leq 2\mathbf{P}(\eta>a)$, where

$$\xi=\exp\left(\frac{1-\mu^2/3}{2}\cdot\frac{1}{Q}\sup_{0\leq t\leq Q}|\tilde{W}(t)|^2\right)>0,$$

$$\eta=\exp\left(\frac{1-\mu^2/3}{2}\cdot\frac{1}{Q}|\tilde{W}(Q)|^2\right)>0.$$

Since

$$\mathbf{E}\left(\eta\right)=\mathbf{E}\left(\exp\left(\frac{1-\mu^2/3}{2}\cdot\frac{1}{Q}|\tilde{W}(Q)|^2\right)\right):=\tilde{c}_\mu<\infty,$$

we have

$$\mathbf{E}\left(\xi\right) = \int_0^\infty \mathbf{P}\left(\xi > x\right) dx \leq 2\int_0^\infty \mathbf{P}\left(\eta > x\right) dx = 2\tilde{c}_\mu < \infty,$$

which completes the proof of the Lemma. $\square$

Bibliography

[1] Leonid Koralov and Shuo Yan. Local limit theorem for time-inhomogeneous functions of Markov processes, 2023.

[2] Shuo Yan. Fast-oscillating random perturbations of Hamiltonian systems, 2023.

[3] Shuo Yan. Large deviations for Hamiltonian systems on intermediate time scales. *Stoch. Dyn.*, 22(6):Paper No. 2250025, 46, 2022.

[4] Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.

[5] R. Z. Has'minskii. A limit theorem for solutions of differential equations with a random right hand part. *Teor. Verojatnost. i Primenen.*, 11:444–462, 1966.

[6] A. N. Borodin. A limit theorem for the solutions of differential equations with a random right-hand side. *Teor. Verojatnost. i Primenen.*, 22(3):498–512, 1977.

[7] Mark I. Freidlin and Alexander D. Wentzell. Random perturbations of Hamiltonian systems. *Mem. Amer. Math. Soc.*, 109(523):viii+82, 1994.

[8] Georges Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *C. R. Acad. Sci. Paris*, 222:847–849, 1946.

[9] Mark Freidlin and Matthias Weber. Random perturbations of nonlinear oscillators. *Ann. Probab.*, 26(3):925–967, 1998.

[10] Mark Freidlin and Matthias Weber. A remark on random perturbations of the nonlinear pendulum. *Ann. Appl. Probab.*, 9(3):611–628, 1999.

[11] Michael Brin and Mark Freidlin. On stochastic behavior of perturbed Hamiltonian systems. *Ergodic Theory Dynam. Systems*, 20(1):55–76, 2000.

[12] Mark Freidlin and Matthias Weber. On random perturbations of Hamiltonian systems with many degrees of freedom. *Stochastic Process. Appl.*, 94(2):199–239, 2001.

[13] Mark Freidlin and Matthias Weber. Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab. Theory Related Fields*, 128(3):441–466, 2004.

[14] Mark I. Freidlin. Perturbation of systems with a first integral: motion on the Reeb graph. *J. Stat. Phys.*, 189(3):Paper No. 36, 22, 2022.

[15] Martin Hairer, Leonid Koralov, and Zsolt Pajor-Gyulai. From averaging to homogenization in cellular flows—an exact description of the transition. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1592–1613, 2016.

[16] Martin Hairer, Gautam Iyer, Leonid Koralov, Alexei Novikov, and Zsolt Pajor-Gyulai. A fractional kinetic process describing the intermediate time behaviour of cellular flows. *Ann. Probab.*, 46(2):897–955, 2018.

[17] M. I. Freidlin and A. D. Wentzell. Averaging principle for stochastic perturbations of multifrequency systems. *Stoch. Dyn.*, 3(3):393–408, 2003.

[18] M. I. Freidlin and A. D. Wentzell. Diffusion processes on an open book and the averaging principle. *Stochastic Process. Appl.*, 113(1):101–126, 2004.

[19] M. I. Freidlin and A. D. Wentzell. Long-time behavior of weakly coupled oscillators. *J. Stat. Phys.*, 123(6):1311–1337, 2006.

[20] Richard B. Sowers. Random perturbations of two-dimensional pseudoperiodic flows. *Illinois J. Math.*, 50(1-4):853–959, 2006.

[21] Xue-Mei Li. An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity*, 21(4):803–822, 2008.

[22] Dmitry Dolgopyat and Leonid Koralov. Averaging of Hamiltonian flows with an ergodic component. *Ann. Probab.*, 36(6):1999–2049, 2008.

[23] Dmitry Dolgopyat and Leonid Koralov. Averaging of incompressible flows on two-dimensional surfaces. *J. Amer. Math. Soc.*, 26(2):427–449, 2013.

[24] Dmitry Dolgopyat, Mark Freidlin, and Leonid Koralov. Deterministic and stochastic perturbations of area preserving flows on a two-dimensional torus. *Ergodic Theory Dynam. Systems*, 32(3):899–918, 2012.

[25] A. N. Borodin and M. I. Freidlin. Fast oscillating random perturbations of dynamical systems with conservation laws. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(3):485–525, 1995.

[26] M. I. Freidlin and A. D. Wentzell. Diffusion approximation for noise-induced evolution of first integrals in multifrequency systems. *J. Stat. Phys.*, 182(3):Paper No. 45, 24, 2021.

[27] Mark Freidlin and Matthias Weber. Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab. Theory Related Fields*, 128(3):441–466, 2004.

[28] Martin Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[29] David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[30] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[31] M. Freidlin and L. Koralov. Averaging in the case of multiple invariant measures for the fast system. *Electron. J. Probab.*, 26:Paper No. 138, 17, 2021.

[32] R. Z. Khas'minskii. Ergodic properties of recurrent diffusion processes and stabilization of the solution to the cauchy problem for parabolic equations. *Theory of Probability & Its Applications*, 5(2):179–196, 1960.

[33] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.

[34] Masuyuki Hitsuda and Akinobu Shimizu. The central limit theorem for additive functionals of Markov processes and the weak convergence to Wiener measure. *J. Math. Soc. Japan*, 22:551–566, 1970.

[35] Déborah Ferré, Loïc Hervé, and James Ledoux. Limit theorems for stationary Markov processes with L^2-spectral gap. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(2):396–423, 2012.

[36] Loïc Hervé and Françoise Pène. The Nagaev-Guivarc'h method via the Keller-Liverani theorem. *Bulletin de la Société Mathématique de France*, 138(3):415–489, 2010.

[37] Pascal Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Statist.*, 5:183–201, 2001.

[38] Yeor Hafouta and Yuri Kifer. *Nonconventional limit theorems and random dynamics*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2018.

[39] Leo Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.

[40] Hubert Hennion. Quasi-compactness and absolutely continuous kernels. *Probab. Theory Related Fields*, 139(3-4):451–471, 2007.

[41] Mark I. Freidlin and Alexander D. Wentzell. On small random perturbation of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, 1970.

[42] Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012.

[43] Mark Freidlin. Reaction-diffusion in incompressible fluid: asymptotic problems. *J. Differential Equations*, 179(1):44–96, 2002.

[44] Robert Sh. Liptser and Anatolii A. Pukhalskii. Limit theorems on large deviations for semimartingales. *Stochastics Stochastics Rep.*, 38(4):201–249, 1992.

[45] A. Puhalskii. On functional principle of large deviations. In *New trends in probability and statistics, Vol. 1 (Bakuriani, 1990)*, pages 198–218. VSP, Utrecht, 1991.

[46] Robert Liptser. Large deviations for two scaled diffusions. *Probab. Theory Related Fields*, 106(1):71–104, 1996.

[47] R. Sh. Liptser and A. N. Shiryayev. *Theory of martingales*, volume 49 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1989.

[48] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[49] Frigyes Riesz and Béla Sz.-Nagy. *Functional analysis*. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1990.