A fractional kinetic process describing the intermediate time behaviour of cellular flows

July 7, 2016

Martin Hairer[1]**, Gautam Iyer**[2]**, Leonid Koralov**[3]**, Alexei Novikov**[4] **and Zsolt Pajor-Gyulai**[5]

[1] University of Warwick, Email: `M.Hairer@Warwick.ac.uk`
[2] Carnegie Mellon University, Email: `gautam@math.cmu.edu`
[3] University of Maryland, Email: `koralov@math.umd.edu`
[4] Pennsylvania State University, Email: `anovikov@math.psu.edu`
[5] New York University, Email: `zsolt@cims.nyu.edu`

Abstract

This paper studies the *intermediate time* behaviour of a small random perturbation of a periodic cellular flow. Our main result shows that on time scales shorter than the diffusive time scale, the limiting behaviour of trajectories that start close enough to cell boundaries is a fractional kinetic process: A Brownian motion time changed by the local time of an independent Brownian motion. Our proof uses the Freidlin-Wentzell framework, and the key step is to establish an analogous averaging principle on shorter time scales.

As a consequence of our main theorem, we obtain a homogenization result for the associated advection diffusion equation. We show that on intermediate time scales the effective equation is a *fractional time* PDE that arises in modelling anomalous diffusion.

1 Introduction

The purpose of this paper is to study the *intermediate time* behaviour of tracer particles passively advected by a periodic cellular flow. Cellular flows arise in various contexts, most notably as a two-dimensional model for heat transport in Bernard convection cells. Our interest in studying the *intermediate time* behaviour stems from [You88] (see also [YJ91]), which proposes a *fractional kinetic* or non-Fickian model governing the behaviour on intermediate time scales. This is in stark contrast to the well known diffusive behaviour on long time scales, and the deterministic Hamiltonian ODE behaviour on short time scales.

The position of tracer particles diffusing in a cellular flow is governed by the SDE

$$d\tilde{X}_t = v(\tilde{X}_t)\,dt + \sqrt{\varepsilon}\,dW_t, \qquad \text{with } \tilde{X}_0 \sim \mu. \tag{1.1}$$

Here, μ is a probability measure on $\mathbb{R}^2$ representing the initial distribution, ε is twice the molecular diffusivity, W is a standard two-dimensional Brownian motion. For notational convenience we denote the law of the solution by $\mathbf{P}_\varepsilon^\mu$, indicating the μ and ε dependence on the probability measure instead of on the process $\tilde{X}$, which we always take to be the canonical process. Above, v is the velocity field of a periodic cellular flow. Namely, there exists a periodic function $H \colon \mathbb{R}^2 \to \mathbb{R}$ (known as the Hamiltonian, or stream function) such that

$$v = \nabla^\perp H \overset{\text{def}}{=} \begin{pmatrix} -\partial_2 H \\ \partial_1 H \end{pmatrix}.$$

Moreover, all the critical points of H are non-degenerate, and there is a connected level set of H, say $\mathcal{L} = \{x \in \mathbb{R}^2 \ : \ H(x) = 0\}$, called the separatrix, which divides the plane into bounded regions (cells) that are each invariant under the (deterministic) flow of the vector field v (see Figure 1). For simplicity of notation, we assume that H has no saddle points inside the cells. An example commonly used in fluid dynamics is $H(x_1, x_2) = \sin(x_1)\sin(x_2)$, as shown in Figure 2.
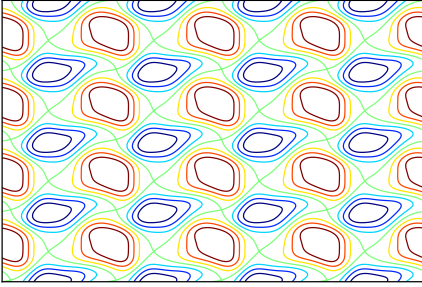

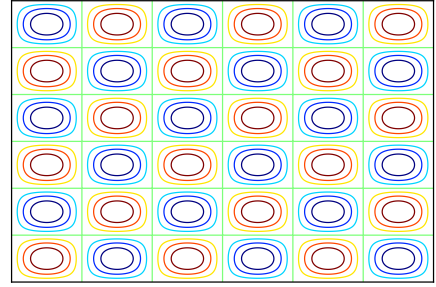Figure 1: A contour plot of the Hamiltonian in a generic cellular flow.

Figure 2: A contour plot of the Hamiltonian $H(x_1, x_2) = \sin(x_1)\sin(x_2)$

The behaviour of $\tilde{X}$ on both short time scales (i.e., time scales of order 1) and long time scales (i.e., time scales larger than $1/\varepsilon$) is well known. On short time scales, a large deviations principle [FW12, Chap. 4, Thm 1.1] guarantees that the trajectories of $\tilde{X}$ deviate from the deterministic trajectories of the flow v with an exponentially small probability. On long time scales, standard homogenization results [Fre64] show that $\tilde{X}$ behaves like a Brownian motion with an enhanced diffusion coefficient.

This paper concerns the effective behaviour of $\tilde{X}$ on *intermediate time scales*, i.e., time scales much larger than 1 and much smaller than $1/\varepsilon$. If the initial condition of $\tilde{X}$ is chosen in such a way that $H(\tilde{X}_0) \neq 0$, then this is again very well understood: at scales of order $1/\varepsilon^\alpha$ with $\alpha \in (0, 1)$, one sees a Brownian motion on the level sets of H. At scale $1/\varepsilon$, one obtains a non-trivial diffusion [FW93], as long as the diffusion in question does not reach the set $H = 0$. This leaves open the question of the behaviour when the initial condition is chosen close to $H = 0$, and this is what we address in this article. For such starting points, the limiting behaviour on both time scales above is a *time changed* Brownian motion. This is a surprising and substantial departure from what is usually expected. The vast majority of results concerning scaling limits of diffusions obtain a limiting behaviour that is again a diffusion, if not a rescaled Brownian motion. A time changed Brownian motion was first obtained in [HKP14] on time scales of order $1/\varepsilon$, and here we extend this result to *much shorter* time scales.

Explicitly, fix $\alpha \in (0,1)$ and consider the time rescaled process

$$Z_t = Z(t) \stackrel{\text{def}}{=} \tilde{X}\Big(\frac{\alpha|\log \varepsilon|t}{\varepsilon^{1-\alpha}}\Big), \tag{1.2}$$

where for notational convenience we sometimes denote time as an argument instead of a subscript. The process Z focuses on the behaviour of $\tilde{X}$ at time scales of order $|\log \varepsilon|/\varepsilon^{1-\alpha}$, and the main result of this paper shows that Z can be spatially rescaled to converge to a *time changed* Brownian motion, provided $\tilde{X}$ starts on (or very close to) cell boundaries. The reason for the extra $|\log \varepsilon|$ factor is the logarithmic slow-down of the underlying dynamical system as it approaches hyperbolic saddles, and is revisited in detail later (see also [Kif81]). Our main result (Theorem 3.4) is a more general version of the following.

Theorem 1.1. *There exists a symmetric strictly positive definite matrix Q such that, if the initial distribution μ^ε is a delta measure at a point that belongs to the separatrix $\mathcal{L}$, then the laws of $\varepsilon^{\frac{1-\alpha}{4}}Z$ converge weakly to the law of W_L^Q. Here W^Q is a Brownian motion on $\mathbb{R}^2$ with covariance matrix Q, and L is the local time at 0 of an independent Brownian motion.*

Note that on the intermediate time scales we consider, if $\tilde{X}$ starts far away from the separatrix, it will simply make many rotations along the flow lines of v without escaping from the cell where it starts. Thus the assumption that $\tilde{X}$ starts on (or very close to) the separatrix is necessary in order to observe a non-trivial limiting behaviour.

As a direct consequence, we also obtain an intermediate time homogenization result for the advection diffusion equation. Let $\tilde{\theta}^\varepsilon$ satisfy the PDE

$$\partial_t \tilde{\theta}^\varepsilon = v \cdot \nabla \tilde{\theta}^\varepsilon + \frac{\varepsilon}{2}\Delta \tilde{\theta}^\varepsilon \qquad \text{on } \mathbb{R}^2 \times (0, \infty), \tag{1.3}$$

with initial data $\tilde{\theta}^\varepsilon(x, 0) = \tilde{\theta}_0^\varepsilon(x)$. Standard homogenization results [PS08, FP94, Fan02] show that on time scales longer than $\mathcal{O}(1/\varepsilon)$, $\tilde{\theta}^\varepsilon$ converges weakly to the solution of the standard heat equation, with an enhanced diffusion coefficient. On intermediate time scales, we show $\tilde{\theta}^\varepsilon$ converges to the solution of a *time fractional* heat equation. Again, this is somewhat unexpected, as the scaling limits of linear parabolic equations usually lead to a parabolic (spatially homogeneous) equation, and not a time fractional equation!

Explicitly, our main PDE result (Theorem 4.1) can be stated as follows.

Theorem 1.2. *For a fixed $\alpha \in (0,1)$, define the rescaled functions θ^ε and θ_0^ε by*

$$\theta^\varepsilon(x, t) = \tilde{\theta}^\varepsilon\Big(\frac{x}{\varepsilon^{(1-\alpha)/4}}, \frac{\alpha|\log \varepsilon|t}{\varepsilon^{1-\alpha}}\Big) \quad and \quad \theta_0^\varepsilon(x) = \tilde{\theta}_0^\varepsilon\Big(\frac{x}{\varepsilon^{(1-\alpha)/4}}\Big). \tag{1.4}$$

If $\theta_0^\varepsilon = \theta_0 \in \mathcal{C}_b(\mathbb{R}^2)$ is independent of ε, then as $\varepsilon \to 0$, θ^ε converges[1] to ϑ, where ϑ satisfies

$$r_0 \mathcal{D}_t^{1/2}\vartheta - \frac{1}{2}[Q : \nabla^2]\vartheta = 0, \qquad \vartheta(x, 0) = \theta_0(x), \tag{1.5}$$

for some constant $r_0 > 0$ that can be computed explicitly in terms of v. Here $\mathcal{D}_t^{1/2}$ denotes the Caputo *derivative of order $1/2$ (see for instance [Die10]) and is defined by*

$$\mathcal{D}_t^{1/2}f = \frac{1}{\sqrt{\pi}}\frac{d}{dt}\int_0^t \frac{f(s) - f(0)}{(t-s)^{1/2}}\,ds, \tag{1.6}$$

---

[1]The notion by which $\theta^\varepsilon \to \vartheta$ is related to the two scale convergence [Ngu89] and is described precisely later. Roughly speaking, one needs to test θ^ε against an ε-dependent measure $\hat{\nu}^\varepsilon$ on $\mathbb{R}^2$, where the family of measures $(\hat{\nu}^\varepsilon)$, when rescaled appropriately, converges to a probability measure supported on the separatrix.

and $Q : \nabla^2 = \sum_{i,j} Q_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$.

Time fractional equations of the form (1.5) often arise when studying anomalous, or non-Fickian diffusions. In this context, it was first suggested by Young [You88] (see also [YPP89, YJ91]) and supported by both numerics and a heuristic explanation. Roughly speaking, on intermediate time scales, the heat near the separatrix diffuses to neighbouring cells, and also gets trapped in cell interiors. This leads to a coupled system governing the effective behaviour, and eliminating the heat in cell interiors from this system leads to (1.5). We elaborate on this and carry out the details in Section 4. We remark, however, that even though this is a purely deterministic result, we prove it using our main probabilistic result (Theorem 3.4) and the Kolmogorov equation. In lieu of a rigorous PDE proof of this result, we provide (in Appendix B) a formal asymptotic expansion motivating it.

Plan of this paper

In order to place our results in the context of the existing literature, Section 2 provides a brief overview of the effective behaviour of tracer particles on both long and short time scales. This section is independent of the rest of the paper and can be skipped by the reader familiar with the literature.

In Section 3, we state the main result of our paper (Theorem 3.4) proving the convergence of $\tilde{X}$ to an effective process on intermediate time scales. An important step in the proof is Theorem 3.2, which is an analogue of the Freidlin-Wentzell averaging principle on these time scales. Before proving the two theorems stated above, we digress and prove an intermediate time homogenization result for the advection diffusion equation governing the density of tracer particles (Theorem 4.1). This is presented in Section 4, and is independent of all subsequent sections (except Appendix B).

The remainder of the paper is devoted to proving our main results. In Section 5 we prove Theorem 3.4, modulo an estimate on how far Z can travel before exiting a small neighbourhood of the separatrix (Proposition 5.1). In Sections 6 and 7, we prove the intermediate time averaging principle (Theorem 3.2). In Appendix A we prove Proposition 5.1. Finally, in Appendix B, we provide a formal asymptotic expansion, which serves as an alternative, purely PDE, approach to derive our intermediate time PDE homogenization result (Theorem 4.1).

2   The effective short time and long time behaviour of tracer particles

This section contains a brief review of results concerning the effective behaviour of tracer particles on long time scales and short time scales. Its main purpose is to place our results in the broader context of existing literature, and the familiar reader can skip directly to Section 3.

2.1   Homogenization: Effective behaviour on long time scales

Well known homogenization results show that on time scales much larger than the diffusive time scale $1/\varepsilon$, the effective behaviour of $\tilde{X}$ is that of a Brownian motion with an enhanced diffusion coefficient. Explicitly, consider the rescaled process $\tilde{Z} = \tilde{Z}^{\varepsilon,\delta}$, defined by

$$\tilde{Z}_t = \tilde{Z}_t^{\varepsilon,\delta} \stackrel{\text{def}}{=} \delta^{1/2} \tilde{X}_{t/\delta}, \tag{2.1}$$

where for clarity we suppress the dependence of $\tilde{X}$ and $\tilde{Z}$ on the parameters ε and δ. Freidlin [Fre64] (see also [Oll94, BLP78, PS08]) proved that for fixed ε we have

$$\tilde{Z} = \tilde{Z}^{\varepsilon,\delta} \xrightarrow[\delta \to 0]{\mathcal{L}} W^{D_{\text{eff}}(\varepsilon)},$$

where $D_{\text{eff}}(\varepsilon)$ is a constant 2×2 positive matrix known as the *effective diffusivity*, and W is a 2D Brownian motion with the covariance matrix $D_{\text{eff}}(\varepsilon)$. Intuitively, the temporal rescaling involves waiting for longer and longer times as $\delta \to 0$. In this time, the process $\tilde{Z}$ spreads out further and further, and rescaling space by a factor of $\sqrt{\delta}$ produces a non-trivial limit. The spatial rescaling is akin to an observer zooming out until the microscopic details of the cellular flow cannot be seen anymore and can effectively be replaced by a homogeneous background.

The effective diffusivity $D_{\text{eff}}(\varepsilon)$ can be computed explicitly by solving a cell problem, and its asymptotic behaviour as $\varepsilon \to 0$ has been extensively studied [Chi79, RBDH87, CS89, FP94, Kor04]. In particular, it is well known that

$$D_{\text{eff}}(\varepsilon) = \mathcal{O}(\sqrt{\varepsilon}) \tag{2.2}$$

as $\varepsilon \to 0$. We observe that $D_{\text{eff}}(\varepsilon)$ is much larger than the molecular diffusivity ε in (1.1) for small ε.

To address the time scales involved, we consider the double limit of $\tilde{Z}$ as both ε and δ approach 0. Using [Fan02] (see also [IKNR14]) it follows that

$$\frac{\tilde{Z}}{\sqrt{D_{\text{eff}}(\varepsilon)}} \xrightarrow[\substack{\varepsilon, \delta \to 0, \\ \delta \ll \varepsilon}]{\mathcal{L}} W. \tag{2.3}$$

Rewriting this in terms of the original process, this means that $\tilde{X}$ behaves like a rescaled Brownian motion on time scales *much larger* than $1/\varepsilon$.

2.2   Averaging and the effective behaviour on the transition time scale

As discussed in the previous section, $\tilde{X}$ homogenizes on time scales larger than $\mathcal{O}(1/\varepsilon)$. Under a compactness assumption (e.g., if the periodic flow is replaced by a flow on a torus) classical results of Freidlin (discussed below) show that $\tilde{X}$ averages along the flow lines of v. In the non-compact setting that we consider, a recent result [HKP14] shows that $\tilde{X}$ transitions between the homogenized and averaged behaviour in a very natural way, and we describe this behaviour here.

To study the behaviour on time scales of order $t \approx 1/\varepsilon$, consider the time rescaled process X defined by

$$X_t = X_t^\varepsilon \stackrel{\text{def}}{=} \tilde{X}_{t/\varepsilon}. \tag{2.4}$$

In this case, X satisfies the SDE

$$dX_t = \frac{1}{\varepsilon} v(X_t)\, dt + dW_t, \qquad \text{with } X_0 \sim \mu.$$

When ε is small, X moves very fast along trajectories of v, and diffuses slowly across them.

To explain further, assume that $H(x)$ is 1-periodic in x_1 and x_2. Let $\mathcal{T} = \mathbb{R}^2/\mathbb{Z}^2$ be the two-dimensional torus, and $\pi : \mathbb{R}^2 \to \mathcal{T}$ be the projection map. The Reeb graph [Ree46] of H (where H is viewed as a function on the torus) is obtained by mapping the connected components of level sets of H to individual points, and using a metric that is locally defined by H. For the Hamiltonians we consider, the Reeb graph is star shaped with each edge corresponding to a cell, and the distance to the vertex corresponding to the absolute value of the Hamiltonian. One example is shown in Figure 3.

Given $x \in \mathcal{T}$, define $\Gamma(x)$ to be the point on the Reeb graph corresponding to the connected component of the level set of H that contains x. Freidlin and Wentzell [FW93] proved that

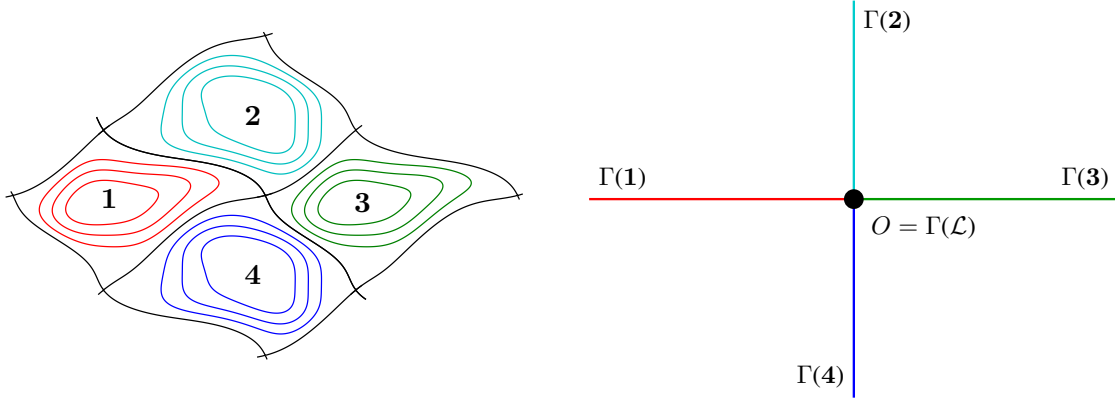$$\Gamma(\pi(X)) \xrightarrow[\varepsilon \to 0]{\mathcal{L}} Y,$$

Figure 3: The graph corresponding to the structure of the level sets of H on $\mathcal{T}$

where Y is a diffusion on the Reeb graph with $Y_0 = \Gamma(X_0)$ and with a specific gluing condition at the interior vertex that can be determined explicitly in terms of the Hamiltonian H. (The exterior vertices are inaccessible and require no boundary condition.) This is the averaging principle.[2]

We emphasise that this only determines the effective behaviour of X *projected* onto the compact Reeb graph of H, when H is viewed as a function on the torus. A recent paper [HKP14] showed how this can be used to obtain the effective behaviour of X on the whole plane $\mathbb{R}^2$. The main theorem in [HKP14] shows that

$$\varepsilon^{1/4} X \xrightarrow[\varepsilon \to 0]{\mathcal{L}} W_L^Q. \tag{2.5}$$

Here Q is a strictly positive definite matrix, and W^Q is a Brownian motion with the covariance matrix Q. The process L is the local time of the limiting diffusion Y at the vertex of the Reeb graph, and is independent of W^Q. The notation W_L^Q in (2.5) above refers to the process W^Q, time changed by the process L.

To relate this to the classical homogenization results, note that equation (2.5) provides information on the effective behaviour of $\tilde{X}$ on time scales of order $1/\varepsilon$; the borderline time scale, beyond which homogenization results are valid. In fact, for the process $\tilde{Z}^{\varepsilon,\delta}$ defined by (2.1), the result of [HKP14] states that for $\delta = \varepsilon$, the limiting process in (2.3) is now a *subordinated* Brownian motion. In contrast, for $\delta \ll \varepsilon$ (as we had in (2.1)), the limiting process is simply an effective Brownian motion without any subordination.

In this spirit, even though the construction of the covariance matrix Q in [HKP14] is not explicit, we can find Q by a matching argument with the existing literature on the effective diffusivity. Indeed, since the process Y is ergodic on the Reeb graph, we must have

$$\lim_{t \to \infty} \frac{L(t)}{t} = \rho,$$

for some $\rho \in (0, \infty)$. This implies that for large t, $W_{L(t)}^Q$ has approximately the same law as a Brownian motion with covariance matrix ρQ. Comparing this with (2.3) and (2.5), and taking the time change (2.4) into account, we get

$$Q = \frac{1}{\rho} \cdot \lim_{\varepsilon \to 0} \frac{D_{\text{eff}}(\varepsilon)}{\sqrt{\varepsilon}}. \tag{2.6}$$

---

[2]Strictly speaking, the classical averaging principle [FW93] requires $H(x) \to \infty$ as $|x| \to \infty$ instead of compactness. These results can, however, be readily adapted to the scenario where the domain is compact.

2.3 Large Deviations: Effective behaviour on short time scales

The next natural asymptotic regime is "intermediate" time scales for which $1 \ll t \ll 1/\varepsilon$. This, however, is the main focus of our paper and is described along with our main results in Section 3. Instead, we conclude this section by briefly describing short time scales.

On time scales of order 1, the trajectories of $\tilde{X}$ deviate from the flow lines of v with an exponentially small probability. To elaborate, let $\phi \colon \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}^2$ be[3] the flow of the vector field v, defined by the ordinary differential equation

$$\partial_t \phi_t(x) = v \circ \phi_t(x), \qquad \text{with } \phi_0(x) = x. \tag{2.7}$$

Then, for every $T, \eta > 0$, we have

$$-\log\Big(\mathbf{P}^x_\varepsilon\Big(\sup_{t\in[0,T]} |\tilde{X}_t - \phi_t(x)| > \eta\Big)\Big) = \mathcal{O}(\varepsilon),$$

where we write $\mathbf{P}^x_\varepsilon = \mathbf{P}^{\delta_x}_\varepsilon$ for brevity. For details, we refer the reader to [FW12, Chapter 4, Theorem 1.1].

This is not surprising as the qualitative effect of the noise is a motion across the flow lines on a time scale of order $1/\varepsilon$, which is much longer than the order one natural time scale of the deterministic motion. We remark, however, that at the slightly longer time scale $t \approx |\log(\varepsilon)|$, an interesting behaviour is observed near the separatrices. The effective process in this regime is a piecewise constant *non-Markovian* process that jumps between the saddle points of H, and we refer the reader to [Bak11, AMB11] for details.

3 Main results: Effective behaviour on intermediate time scales

The main contribution of this paper is the precise description of the effective behaviour of $\tilde{X}$ on *intermediate* time scales where $1 \ll t \ll 1/\varepsilon$. As we have outlined earlier, the effective behaviour of $\tilde{X}$ on these time scales might seem trivial at first glance. Indeed, convection only transports $\tilde{X}$ along flow lines of v, which are all closed orbits inside each cell. On the other hand, for diffusion to transport $\tilde{X}$ to a different cell, it will take time of order $1/\varepsilon$, which is much longer than the time scales under consideration. Thus, if $\tilde{X}$ starts at a generic point inside one of the cells, it will simply make many rotations along the flow lines of v without escaping the cell.

The interesting behaviour is observed when $\tilde{X}$ starts close enough to (or on) the separatrix. The diffusion is then strong enough to transport $\tilde{X}$ from one cell to another and, combined with the effect of the drift, the process $\tilde{X}$ can conceivably travel large distances in a short time. Indeed, a recent result [IN16] proves that on time scales for which $1 \ll t \ll 1/\varepsilon$, the variance of $\tilde{X}_t$ is of order $\sqrt{t}$, up to a logarithmic correction. The main result of the present article goes much further than a variance estimate, and provides an effective process on these intermediate time scales.

For a given $\alpha \in (0, 1)$, we study the behaviour of $\tilde{X}$ on time scales of order[4] $|\log \varepsilon|/\varepsilon^{1-\alpha}$ using the time rescaled process $Z = Z^\varepsilon$ defined by (1.2). As before, we suppress the ε-dependence of the process Z and use time as an argument instead of a subscript when notationally convenient. Clearly, time scales of order $|\log \varepsilon|/\varepsilon^{1-\alpha}$ are shorter than time scales of order $1/\varepsilon$, and longer than time scales of order 1.

---

[3]Throughout this paper we use the convention that $\mathbb{R}_+ = [0, \infty)$.

[4]Choosing $\alpha \in (0, 1)$ and restricting to time scales of order $|\log \varepsilon|/\varepsilon^{1-\alpha}$ is performed mainly for convenience, and does not have any bearing on the final result. In fact, the main results of this paper can be formulated more generally by choosing a parameter $\delta = \delta(\varepsilon)$ such that both $\delta \to 0$ and $\varepsilon/\delta \to 0$ as $\varepsilon \to 0$. Now a description of $\tilde{X}$ on time scales of order δ can be obtained from our main results by replacing all occurrences of $\varepsilon^{1-\alpha}$, $\varepsilon^{\alpha/2}$ and $\alpha|\ln \varepsilon|$ with δ, $(\delta/\varepsilon)^{1/2}$ and $\ln(\delta/\varepsilon)$ respectively.

We describe the effective behaviour of Z in two steps: First, we compactify the state space by projecting Z onto the periodic torus. In this case, we prove a direct analogue of the classical Freidlin-Wentzell averaging principle [FW12] on shorter time scales, and show that the limiting process is a diffusion Y on a (rescaled) Reeb graph. Next, we show that the limiting behaviour of Z on $\mathbb{R}^2$ is exactly an independent two-dimensional Brownian motion time changed by the local time of Y at the vertex of the rescaled Reeb graph. That is, the effective process only moves when the graph diffusion Y is at the vertex. These steps are described below in Sections 3.1 and 3.2 respectively.

3.1 Intermediate time averaging on the torus

The purpose of this section is to state an analogue of the classical Freidlin-Wentzell averaging principle [FW12] when Z is projected onto the torus. While we state our result in the context of cellular flows, it is applicable more generally to behaviour of Hamiltonian systems around heteroclinic connections.

We begin with some notation describing the geometry of the Hamiltonian and the projection on the Reeb graph. We recall that we normalised H so that it has period 1, and the separatrix, denoted by $\mathcal{L}$, is exactly

$$\mathcal{L} = \{x \in \mathbb{R}^2 \ : \ H(x) = 0\},$$

and is assumed to be connected. Let $\mathcal{T} \stackrel{\text{def}}{=} \mathbb{R}^2/\mathbb{Z}^2$ be the torus, $\pi \colon \mathbb{R}^2 \to \mathcal{T}$ be the projection map, and define $\mathcal{L}_\mathcal{T} = \pi(\mathcal{L})$. Let $A_1, \ldots, A_M$ denote the saddle points of H on the separatrix $\mathcal{L}_\mathcal{T}$. Then $\mathcal{L}_\mathcal{T}$ (or $\mathcal{L}$) is the union of the saddles $\{A_i\}$ (or $\pi^{-1}(\{A_i\})$, respectively), and the heteroclinic orbits connecting these saddles. For notational simplicity in the proof, we assume that there are no homoclinic orbits (i.e., orbits that connect a saddle to itself).

By Euler's polyhedron formula (recall that the torus has Euler characteristic zero), there are exactly M connected components of the complement of the separatrix $\mathcal{T} \setminus \mathcal{L}_\mathcal{T}$, and we denote these domains by U_1, $\ldots, U_M$. (There is, however, no particular relation between the numbering of the U_i's and that of the A_i's.) For convenience, we further assume that there are no saddle points of H in the interior of the sets $U_1, \ldots, U_M$.

We now define the space $\mathcal{G}$ that serves as the rescaled Reeb graph of H. Let $\mathcal{G}$ be the topological quotient space obtained from $\{1, \ldots, M\} \times \mathbb{R}^+$ by identifying all the points $(1, 0), \ldots, (M, 0)$ with each other. We observe that $\mathcal{G}$ is a star shaped graph with semi-infinite edges $I_i \stackrel{\text{def}}{=} \{i\} \times \mathbb{R}^+$, corresponding to the rescaled distance into the interior of U_i, and one interior vertex $O = (1, 0) = \cdots = (M, 0)$ corresponding to the separatrix $\mathcal{L}_\mathcal{T}$. A natural metric on $\mathcal{G}$ is given by

$$d_\mathcal{G}((i, y), (j, \bar{y})) = \left\{ \begin{array}{ll} |y - \bar{y}| & \text{if } i = j, \\ |y| + |\bar{y}| & \text{otherwise.} \end{array} \right. \tag{3.1}$$

Define the projection $\Gamma_\varepsilon \colon \mathcal{T} \to \mathcal{G}$ by

$$\Gamma_\varepsilon(x) = (i, \varepsilon^{-\alpha/2}|H(x)|) \quad \text{if } x \in \overline{U}_i, \tag{3.2}$$

and extend it periodically to $\mathbb{R}^2$. Note that Γ_ε projects each invariant region U_i into an edge I_i of the graph and $\Gamma_\varepsilon(\mathcal{L}_\mathcal{T}) = O$. Even though the sets $\overline{U}_i$ overlap, the map Γ_ε is well-defined since the points $(i, 0)$ have all been identified.

We claim that the law of the projected process $(\Gamma_\varepsilon \circ \pi)(Z)$ converges weakly to a process Y on the graph $\mathcal{G}$. The limiting process Y is a diffusion, and it can be characterised by its generator $\mathcal{A}$. We describe this before stating the main convergence result of this section.

On the i-th edge of the graph, define the operator $\mathcal{A}_i$ by

$$\mathcal{A}_i = \frac{a_i}{2} D_i^2.$$

Here D_i denotes the derivative along the i-th edge of $\mathcal{G}$, and the coefficients a_i are defined by

$$a_i \stackrel{\text{def}}{=} q_i \cdot \lim_{\varepsilon \to 0} \frac{|\log \varepsilon|}{T_i(\varepsilon^{1/2})}, \tag{3.3}$$

where

$$q_i = \oint_{\partial U_i} |\nabla H| \, dl \qquad \text{and} \qquad T_i(h) = \oint_{\{|H(x)|=h\} \cap U_i} |\nabla H|^{-1} \, dl. \tag{3.4}$$

We recall that $T_i(h)$ is the time the flow ϕ (defined in equation (2.7)) takes to complete one rotation along the periodic orbit starting from any point $x \in U_i$ for which $|H(x)| = h$. Since ∂U_i contains hyperbolic saddles, we know that as $h \to 0$, the period $T_i(h)$ diverges at a logarithmic rate, and hence a_i is finite and strictly positive.

Now we define the domain $D(\mathcal{A})$ to be the set of all functions F that satisfy the following conditions:

(a) $F \in \mathcal{C}_0(\mathcal{G}) \cap \mathcal{C}^2(\mathcal{G} \setminus \{O\})$. That is, F is continuous on $\mathcal{G}$, tends to zero at infinity, and is twice continuously differentiable away from the interior vertex O.

(b) Writing $\mathbf{1}_A$ for the indicator function of a set A, the function

$$y \mapsto \sum_{i=1}^{M} \mathbf{1}_{\{y \in I_i\}} \mathcal{A}_i F(y), \tag{3.5}$$

defined on $\mathcal{G} \setminus \{O\}$, extends to a $\mathcal{C}_0$ function on all of $\mathcal{G}$.

(c) The function F satisfies the flux condition

$$\sum_{i=1}^{M} q_i D_i F(O) = 0,$$

Finally, for $F \in D(\mathcal{A})$ we define $\mathcal{A}F$ to be the unique $\mathcal{C}_0(\mathcal{G})$ extension of the function defined in (3.5).

Replicating results from [Man68] for our one-dimensional operator, we can show that for every $u \in D(\mathcal{A})$ and $\lambda > 0$ the resolvent equation $\lambda f - \mathcal{A}f = u$ has a unique solution $f \in \mathcal{D}(\mathcal{A})$. Since $\mathcal{A}$ is a closed operator, the Hille-Yosida theorem [EK86, Theorem 2.2 in Ch 4], shows that $\mathcal{A}$ generates a strongly continuous positive contraction semigroup on $\mathcal{C}_0(\mathcal{G}) = \overline{D(\mathcal{A})}$, the L^∞ closure of $D(\mathcal{A})$. Therefore there is a Fellerian Markov family Y with generator $\mathcal{A}$ (see [EK86, §4.1-4.2]), and we use the Kolmogorov continuity theorem to replace Y with a modification with continuous trajectories on $\mathcal{G}$. The process Y will arise in the main result of this subsection (Theorem 3.2, below) as the weak limit of $\Gamma_\varepsilon(Z)$.

As with the process Z, we transfer the dependence of Y on its initial position to the associated probability measure. When the law of Y_0 is μ, we denote the corresponding probability measure on $\mathcal{C}(\mathbb{R}_+, \mathcal{G})$ by $\bar{\mathbf{P}}^\mu$, and the associated expectation operator by $\bar{\mathbf{E}}^\mu$. When μ is concentrated at a point $y \in \mathcal{G}$, we will simply write $\bar{\mathbf{P}}^y$ and $\bar{\mathbf{E}}^y$ as appropriate.

Remark 3.1. *The process Y can alternately be constructed directly as follows. Take a standard Brownian motion B and decompose it into excursions away from the origin. Since there are countably many such excursions, we can enumerate them. Say that the kth excursion happens during the interval (s_k, t_k), then these intervals are all disjoint and the complement of their union consists precisely of the null set of times for which $B(t) = 0$. Consider now a sequence i_k of i.i.d. $\{1, \ldots, M\}$-valued random variables independent of B with $\mathbf{P}(i_k = j)$ proportional to $q_j / \sqrt{a_j}$. We then define a $\mathcal{G}$-valued process Y in the following way. If $t \in (s_k, t_k)$ for some k and $i_k = j$, we set $Y_t = (j, \sqrt{a_j}|B(t)|) \in \mathcal{G}$, otherwise we set $Y_t = O$. In order to start this process with an initial condition $y = (j, c) \neq O$, one can perform the same construction, with the difference that one sets $B(0) = c/\sqrt{a_j}$, and we set the value i_k corresponding to the excursion containing time 0 to j. We refer the reader to [Lej06] for more details and various other constructions.*

We can now state the main result of this section.

Theorem 3.2. *Let μ^ε be a family of probability measures on $\mathcal{T}$ such that the push forward measures $\Gamma_\varepsilon^* \mu^\varepsilon$ converge weakly, as $\varepsilon \to 0$, to a probability measure μ on $\mathcal{G}$. Then the laws of $\Gamma_\varepsilon(Z)$ under $\mathbf{P}_\varepsilon^{\mu^\varepsilon}$ converge weakly to that of the process Y under $\bar{\mathbf{P}}^\mu$. That is, for every bounded continuous $f : \mathcal{C}(\mathbb{R}_+; \mathcal{G}) \to \mathbb{R}$, we have*

$$\lim_{\varepsilon \to 0} \mathbf{E}_\varepsilon^{\mu^\varepsilon} f(\Gamma_\varepsilon(Z.)) = \bar{\mathbf{E}}^\mu f(Y.) .$$

Moreover, for fixed μ and f, the convergence is uniform with respect to all choices of μ^ε such that $\Gamma_\varepsilon^* \mu^\varepsilon = \mu$.

The proof of Theorem 3.2 is given in Section 6. As mentioned earlier, Theorem 3.2 is not restricted to the cellular flow setting of this paper. It describes the behaviour of a generic Hamiltonian system around heteroclinic connections, and serves as a direct analogue of the classical Freidlin-Wentzell averaging principle [FW12] at shorter time scales.

3.2 The intermediate time behaviour on the plane

Theorem 3.2, stated in the previous section, shows that a limiting behaviour of the projection $\Gamma_\varepsilon(Z)$ is a diffusion on the rescaled Reeb graph $\mathcal{G}$. In this section, we show that the limiting behaviour of Z itself is an independent Brownian motion on the plane time changed by the local time of Y at O. We begin by recalling the abstract definition of the local time of Y.

Definition 3.3. *The local time of Y is the unique nonnegative random field*

$$L \overset{\text{def}}{=} \{L_t(y) : (t, y) \in \mathbb{R}_+ \times \mathcal{G}\}$$

such that the following hold:

(a) *The mapping $(t, y) \to L_t(y)$ is jointly measurable, and $L_t(y)$ is adapted.*

(b) *For every $y \in \mathcal{G}$, the mapping $t \to L_t(y)$ is non-decreasing and constant on all open intervals for which $Y_t \neq y$.*

(c) *For every bounded Borel measurable $f : \mathcal{G} \to \mathbb{R}_+$ and every $y_0 \in \mathcal{G}$, we have*

$$\int_0^t f(Y_s) a(Y_s) \, ds = 2 \int_{\mathcal{G}} f(y) L_t(y) \, dy \qquad \bar{\mathbf{P}}^{y_0}\text{-a.s.,}$$

> where $a : \mathcal{G} \to \mathbb{R}^+$ is defined[5] by $a(y) = a_i$ if $y \in I_i$.

(d) For every $y_0 \in \mathcal{G}$, $L_t(y)$ is $\bar{\mathbf{P}}^{y_0}$-a.s. jointly continuous in t and y for $y \neq O$. Moreover, at O, we have

$$L_t(O) = \sum_{i=1}^{M} \lim_{\substack{y \to O, \\ y \in I_i}} L_t(y) .$$

The existence and uniqueness of local time for diffusions on the real line is relatively well studied (see for instance [RW00a, RW00b]). These standard results, together with [FS00, Lemma 2.2], give the existence and uniqueness of L. Moreover, in view of Remark 3.1, the process $L_t(O)$ is a constant multiple of the local time of a one-dimensional Brownian motion B at the origin.

With this definition we state our result concerning the limiting behaviour of Z. To state our full convergence result, we introduce the state space

$$\bar{\mathcal{G}} \stackrel{\text{def}}{=} \mathbb{R}^2 \times \mathcal{G} , \quad \text{with metric } d_{\bar{\mathcal{G}}}((x,g),(\bar{x},\bar{g})) \stackrel{\text{def}}{=} |x - \bar{x}| + d_{\mathcal{G}}(g,\bar{g}) .$$

Our main probabilistic convergence result then reads as follows.

Theorem 3.4. *Define $\widehat{\Gamma}_\varepsilon : \mathbb{R}^2 \to \bar{\mathcal{G}}$ by*

$$\widehat{\Gamma}_\varepsilon(x) = (\varepsilon^{(1-\alpha)/4} x, (\Gamma_\varepsilon \circ \pi)(x)) ,$$

and let ν_ε be a family of probability measures on $\mathbb{R}^2$ such that the push forward measures $\widehat{\Gamma}_\varepsilon^* \nu_\varepsilon$ converge weakly to a probability measure $\hat{\nu}$ on $\bar{\mathcal{G}}$. Then, there exists a pair of processes $(\tilde{W}^Q, Y)$ defined on some probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}}^{\hat{\nu}})$ such that the following hold.

(a) *The initial distribution of $(\tilde{W}^Q, Y)$ under $\bar{\mathbf{P}}^{\hat{\nu}}$ is $\hat{\nu}$.*

(b) *The process $\tilde{W}_\cdot^Q - \tilde{W}_0^Q$ is a Brownian motion on $\mathbb{R}^2$ with strictly positive definite covariance matrix Q.*

(c) *The process Y is a diffusion on the graph $\mathcal{G}$ with generator $\mathcal{A}$. Moreover, conditioned on $(\tilde{W}_0^Q, Y_0)$, the processes $\tilde{W}^Q$ and Y are independent.*

(d) *As $\varepsilon \to 0$, the law of the process $\widehat{\Gamma}_\varepsilon(Z_\cdot)$ under $\mathbf{P}^{\nu_\varepsilon}$ converges weakly to that of $\Xi \stackrel{\text{def}}{=} (\tilde{W}_{L_\cdot}^Q, Y)$ under $\bar{\mathbf{P}}^{\hat{\nu}}$, where $L_t = L_t(O)$ is the local time of Y at the interior vertex O.*

Since $L_t(O)$ is simply a constant multiple of Brownian local time, $\tilde{W}_L^Q$ is a fractional kinetic process of index $1/2$. This process arises naturally as the scaling limit of many trap models, such as continuous time random walks with heavy tailed jump times [MS04, MS08] or the Bouchaud trap model [Bou92, BAČ07]. Intuitively, the scaling limit of the time of an excursion of X_t away from the separatrix (when the process is trapped inside a cell) is approximately an excursion of a Brownian motion, and its length is accordingly heavy tailed with index $1/2$.

If the support of $\hat{\nu}$ concentrates on $\mathbb{R}^2 \times \{O\}$, then by Brownian scaling, $\bar{\mathbf{E}}^O L_t = c\sqrt{t}$ for some constant $c(a_i, q_i) > 0$. In this case the variance of the limit process $\tilde{W}_L^Q$ is proportional to $\sqrt{t}$ for all time. This was proved earlier in [IN16] in the case $H(x_1, x_2) = \sin(x_1)\sin(x_2)$. The proof of Theorem 3.4 is presented in Section 5. Even though many ingredients in the proofs rely on the corresponding techniques from the companion paper [HKP14], we keep the current paper self contained by sketching the main steps and highlighting the differences involved.

---

[5]Strictly speaking, $a(O)$ is not well defined. This, however, does not affect the left hand side since, with probability one, the process Y spends time of measure zero at the interior vertex O.

4 A PDE Application: Intermediate time homogenization of the advection diffusion equation

This section is devoted to an intermediate time homogenization result for the advection diffusion equation. We emphasize that the proof of the probabilistic result (Theorem 3.4) does not rely on the arguments in this section. The PDE (1.3) is closely related to the process $\tilde{X}$, and understanding the behaviour of $\tilde{X}$ on intermediate time scales yields an intermediate time homogenization result for (1.3) in a natural way. Since $\tilde{X}$ behaves like a fractional kinetic process on these time scales, it is natural to expect that $\tilde{\theta}$ satisfies the (time) fractional heat equation (1.5), and this was heuristically derived by Young [You88]. We prove it rigorously below (Theorem 4.1) using Theorem 3.4.

Theorem 4.1. *Let $\tilde{\theta}^\varepsilon$ satisfy the PDE (1.3) for $x \in \mathbb{R}^2$, $t > 0$ with initial data $\tilde{\theta}_0^\varepsilon$. For $\alpha \in (0,1)$ define the rescaled functions θ^ε and θ_0^ε by (1.4), and suppose $\theta_0^\varepsilon = \theta_0 \in \mathcal{C}_b(\mathbb{R}^2)$ and is independent of ε. Define the rescaled projection $\overline{\Gamma}_\varepsilon$ by $\overline{\Gamma}_\varepsilon(x) = \widehat{\Gamma}_\varepsilon(x/\varepsilon^{(1-\alpha)/4})$. Then for any family of probability measures $\bar{\nu}^\varepsilon$ on $\mathbb{R}^2$ such that $\overline{\Gamma}_\varepsilon^* \bar{\nu}^\varepsilon$ converges weakly to a probability measure ν on $\bar{\mathcal{G}}$, we have*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^2} \theta^\varepsilon(x,t) \, d\bar{\nu}^\varepsilon(x) = \int_{\mathbb{R}^2 \times \mathcal{G}} \theta(x,y,t) \, d\nu(x,y). \tag{4.1}$$

Here θ is the unique classical solution to the system

$$\partial_t \theta - \mathcal{A}_y \theta = 0 \qquad \text{for } y \neq O,\, t > 0, \tag{4.2a}$$

$$\frac{1}{2}[Q : \nabla_x^2]\theta + \sum_{i=1}^M \bar{q}_i D_i^y \theta = 0 \qquad \text{for } y = O,\, t > 0, \tag{4.2b}$$

$$\theta(x,y,0) = \theta_0(x). \tag{4.2c}$$

Here $\bar{q}_i = q_i / \sum_{j=1}^M q_j$, $D_i^y = D_i$ denotes the derivative along the ith edge of $\mathcal{G}$, and $\mathcal{A}_y$ is the generator of the process Y acting only on the variable y.

    Moreover, if $\nu = \nu' \times \delta_O$, then

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^2} \theta^\varepsilon(x,t) \, d\bar{\nu}^\varepsilon(x) = \int_{\mathbb{R}^2} \vartheta(x,t) \, d\nu'(x),$$

where $\vartheta(x,t) \stackrel{def}{=} \theta(x,O,t)$ satisfies the Caputo time fractional equation

$$\Big(\sum_{i=1}^M \frac{\bar{q}_i}{\sqrt{a_i/2}}\Big) \mathcal{D}_t^{1/2} \vartheta - \frac{1}{2}[Q : \nabla^2]\vartheta = 0, \tag{4.3}$$

with initial data θ_0. Here $\mathcal{D}_t^{1/2}$ denotes the Caputo derivative *of order $1/2$ defined by (1.6).*

Remark. *Even though this is a purely deterministic result, our proof is probabilistic and relies on Theorem 3.4. In lieu of additionally presenting a direct PDE proof, we provide in Appendix B a formal asymptotic expansion motivating (4.2).*

    Well-posedness and regularity of solutions to (4.2) is standard. Parabolic problems similar to (4.3) and the regularity of their solutions are discussed in, e.g. [ACV16]. The notion of convergence used in equation (4.1) is known as two-scale convergence, and was introduced by Nguetseng [Ngu89]. It has

proved to be an invaluable tool in the theory of homogenization and has been applied various contexts. In most situations, however, the underlying small-scale manifold is the torus. The key difference in (4.1) is that the underlying small scale naturally arises as the rescaled Reeb graph of the Hamiltonian.

Before proving Theorem 4.1, we momentarily pause to consider an illustrative special case. For $x \in \mathbb{R}^2$, take the sequence of measures $\bar{\nu}^\varepsilon$ defined by

$$\bar{\nu}^\varepsilon = \delta\left(\varepsilon^{(1-\alpha)/4}\left\lfloor\frac{x}{\varepsilon^{(1-\alpha)/4}}\right\rfloor\right),$$

where $\delta(z)$ denotes the delta measure supported at the point $z \in \mathbb{R}^2$, and we assume that the origin belongs to the separatrix (and therefore so does every point with integer coordinates). Applying Theorem 4.1 now shows that

$$\theta^\varepsilon\left(\varepsilon^{(1-\alpha)/4}\left\lfloor\frac{x}{\varepsilon^{(1-\alpha)/4}}\right\rfloor, t\right) \xrightarrow{\varepsilon \to 0} \vartheta(x, t),$$

for all $x \in \mathbb{R}^2$ and $t > 0$. That is, at time t the value of the temperature θ^ε at the corner of the domain of periodicity containing x converges to $\vartheta(x, t)$.

At first sight, this is extremely surprising. Long time scaling limits of (1.3) have been studied extensively, and the limiting behaviour is simply the heat equation with an enhanced diffusion coefficient. In our situation, the "intermediate time" scaling limit of (1.3) is a *time fractional* heat equation (4.3)!

The heuristic explanation of this is as follows. On time scales shorter than $1/\varepsilon$, any heat trapped in the interior of one cell will not escape the cell. Thus to observe a non-trivial limiting behaviour, at these time scales one needs to zoom in close to the separatrix. Indeed, if the family of measures $\widehat{\Gamma}_\varepsilon^* \hat{\nu}^\varepsilon$ converge weakly to a probability measure (as required in Theorem 3.4), the supports of $\hat{\nu}^\varepsilon$ must asymptotically concentrate on the separatrix.

Now, in a small neighbourhood of the separatrix, there are two effects at play: heat diffuses to neighbouring cells, and heat is "trapped" in the cell interior. Thus the limiting behaviour should be a coupled system balancing these two effects. This is precisely what (4.2a) and (4.2b) accomplish. Many similar models for anomalous diffusion have been studied by various authors. For example, Young [You88] (see also [YJ91]) heuristically derived a similar system in the context of cellular flows. Theorem 4.1 establishes this rigorously, and we prove it below.

Proof of Theorem 4.1. By the Feynman-Kac formula

$$\theta^\varepsilon(x, t) = \tilde{\theta}^\varepsilon\left(\frac{x}{\varepsilon^{(1-\alpha)/4}}, \frac{\alpha|\log \varepsilon|t}{\varepsilon^{1-\alpha}}\right) = \mathbf{E}_\varepsilon^{x/\varepsilon^{(1-\alpha)/4}}\theta_0(\varepsilon^{(1-\alpha)/4}Z_t).$$

Hence, by Theorem 3.4,

$$\int_{\mathbb{R}^2} \theta^\varepsilon(x, t)\, d\bar{\nu}^\varepsilon(x) = \mathbf{E}_\varepsilon^{\hat{\nu}^\varepsilon}\theta_0\left(\varepsilon^{(1-\alpha)/4}Z_t\right) \xrightarrow{\varepsilon \to 0} \bar{\mathbf{E}}^\nu\theta_0(\tilde{W}_{L(t)}^Q), \qquad (4.4)$$

where $\hat{\nu}^\varepsilon$ is the rescaled measure defined by

$$d\hat{\nu}^\varepsilon(x) = d\bar{\nu}^\varepsilon(\varepsilon^{(1-\alpha)/4}x),$$

and $\tilde{W}^Q$ and L are as in Theorem 3.4. Define the function $\theta \colon \mathbb{R}^2 \times \mathcal{G} \times \mathbb{R}_+ \to \mathbb{R}$ by

$$\theta(x, y, t) \stackrel{\text{def}}{=} \bar{\mathbf{E}}^{(x,y)}\theta_0(\tilde{W}_{L(t)}^Q), \qquad (4.5)$$

where we recall $\bar{\mathbf{E}}^{(x,y)}$ is the expectation operator with respect to the probability measure $\bar{\mathbf{P}}^{(x,y)}$ under which $\bar{\mathbf{P}}^{(x,y)}(\tilde{W}_0^Q = x \ \& \ Y_0 = y) = 1$. Now,

$$\bar{\mathbf{E}}^\nu \theta_0(\tilde{W}_{L(t)}^Q) = \int_{\mathbb{R}^2 \times \mathcal{G}} \theta(x, y, t) \, d\nu(x, y)$$

and hence (4.1) follows from (4.4).

The fact that θ satisfies the system (4.2) follows from (4.5) and an Itô formula for Y that was proved in [FS00]. Since this is interesting in its own right, we single it out as a proposition (Proposition 4.2, below) and defer it to the end of this section.

Finally, to prove (4.3) when $\nu = \nu' \times \delta_O$, we only need to show that given a solution to (4.2), the function $\vartheta(x, t) \stackrel{\text{def}}{=} \theta(x, O, t)$ satisfies (4.3). This follows from the explicit solution formula for the heat equation on the half line, and similar results are readily available in the literature (see for instance [MS12, §4.5]). For convenience, we derive it below.

Along the ith edge, equation (4.2a) is simply the one dimensional heat equation. Rearranging (4.2b), we obtain the boundary condition

$$D_i^y \theta(x, O, t) = -\frac{1}{\bar{q}_i}\Big(\frac{1}{2}[Q : \nabla_x^2]\theta(x, O, t) + \sum_{j \neq i}^M \bar{q}_j D_j^y \theta(x, O, t)\Big). \tag{4.6}$$

Treating the right hand side of (4.6) as a given function, we can explicitly solve (4.2a) on the ith edge, with boundary condition (4.6) and constant (in y) initial data $\theta_0(x)$. This gives

$$\theta(x, O, t) = \theta_0(x) + \frac{1}{\bar{q}_i}\Big(\frac{a_i}{2\pi}\Big)^{1/2} \int_0^t \Big(\frac{1}{2}[Q : \nabla_x^2]\theta(x, O, s) + \sum_{j \neq i}^M \bar{q}_j D_j^y \theta(x, O, s)\Big) \frac{ds}{\sqrt{t-s}},$$

for every $i \in \{1, \ldots, M\}$. Multiplying both sides by $\bar{q}_i/\sqrt{a_i/2}$, summing over i and using (4.2b) yields

$$\sum_{i=1}^M \frac{\bar{q}_i}{\sqrt{a_i/2}}\Big(\theta(x, O, t) - \theta_0(x)\Big) = \frac{1}{2\sqrt{\pi}} \int_0^t [Q : \nabla_x^2]\theta(x, O, s) \frac{ds}{\sqrt{t-s}}.$$

Applying $\mathcal{D}_t^{1/2}$ to both sides and using $\vartheta(x, t) = \theta(x, O, t)$ yields (4.3) as desired. $\qquad \square$

In the above proof we used the fact that θ defined by (4.5) satisfies the system (4.2). We state and prove this next (see also [PGS15] for a related result).

Proposition 4.2. *Let $\theta_0 \in \mathcal{C}_b(\bar{\mathcal{G}})$, and define θ by (4.5). Then θ satisfies the system (4.2) for $t > 0$, and is continuous at $t = 0$.*

Proof. The first step is to obtain an Itô formula for the process $\Xi \stackrel{\text{def}}{=} (\tilde{W}_L^Q, Y)$. For the process Y alone, an Itô formula is known and can be found in Freidlin and Sheu [FS00]. Explicitly, there exists[6] a Brownian motion B such that

$$f(Y_t) - f(Y_0) = \sum_{i=1}^M \int_0^t D_i f(Y_s) \sigma_i(Y_s) \, dB_s + \int_0^t \mathcal{A}_y f(Y_s) \, ds + \sum_{i=1}^M \bar{q}_i D_i f(O) \, L_t, \tag{4.7}$$

---

[6] The Brownian motion B can be directly obtained from the construction outlined in Remark 3.1. Indeed, if $\tilde{B}$ denotes the Brownian motion in Remark 3.1, then we have $dB = \text{sign}(\tilde{B}) \, d\tilde{B}$.

holds any $f \in \mathcal{C}_b^2(\mathcal{G})$. Here $\sigma_i(y) = \sqrt{a_i}$ if $y \in I_i$ and $\sigma_i(y) = 0$ otherwise.

Now, since $\tilde{W}^Q$ and Y are independent, the time changed process $\tilde{W}_L^Q$ is a martingale with joint quadratic variations given by

$$d\langle \tilde{W}_L^{Q,i}, \tilde{W}_L^{Q,j} \rangle_t = Q_{i,j} dL_t \qquad \text{and} \qquad d\langle \tilde{W}_L^{Q,i}, B \rangle_t = 0$$

for all $i, j \in \{1, 2\}$. Here $\tilde{W}^{Q,i}$ denotes the ith component of $\tilde{W}^Q$, and $Q_{i,j}$ is the i-j^{th} entry of the matrix Q. For $f \in \mathcal{C}_b^2(\mathbb{R}^2 \times \mathcal{G})$, we thus obtain by Itô's formula

$$f(\Xi_t) - f(\Xi_0) = \sum_{i=1}^M \int_0^t D_i^y f(\Xi_s) \sigma_i(Y_s) dB_s + \int_0^t \mathcal{A}_y f(\Xi_s)\, ds$$

$$+ \sum_{i=1}^2 \int_0^t \partial_{x_i} f(\Xi_s)\, d\tilde{W}_{L_s}^{Q,i} + \int_0^t \Big(\frac{1}{2}[Q : \nabla_x^2] + \sum_{i=1}^M \bar{q}_i D_i^y \Big) f(\Xi_s)\, dL_s. \quad (4.8)$$

Now we use the Itô formula to compute $\mathcal{A}_\Xi$, the generator of Ξ. Indeed, for $f \in \mathcal{C}_b^2(\mathbb{R}^2 \times \mathcal{G})$ we have

$$\mathcal{A}_\Xi f(x, y) = \lim_{t \to 0} \frac{1}{t} \bar{\mathbf{E}}^{(x,y)} (f(\Xi_t) - f(\Xi_0))$$

$$= \lim_{t \to 0} \frac{1}{t} \bar{\mathbf{E}}^{(x,y)} \Big(\int_0^t \mathcal{A}_y f(\Xi_s)\, ds + \int_0^t \Big(\frac{1}{2}[Q : \nabla_x^2] + \sum_{i=1}^M \bar{q}_i D_i^y \Big) f(\Xi_s)\, dL_s \Big), \quad (4.9)$$

since the other two terms on the right of (4.8) are martingales and have expectation 0. Now, as $t \to 0$, the first term on the right of (4.9) converges to $\mathcal{A}_y f(x, y)$. For the second term on the right of (4.9), the fact that L is a constant multiple of Brownian local time gives

$$\bar{\mathbf{E}}^{(x,y)} \int_0^t \Big(\frac{1}{2}[Q : \nabla_x^2] + \sum_{i=1}^M \bar{q}_i D_i^y \Big) f(\Xi_s)\, dL_s = \begin{cases} o(t) & y \neq O, \\ \mathcal{O}(\sqrt{t}) \Big(\frac{1}{2}[Q : \nabla_x^2] + \sum_{i=1}^M \bar{q}_i D_i^y \Big) f(x, O) & y = O. \end{cases}$$

After dividing by t and taking the limit as $t \to 0$, this vanishes without any further restriction on f if $y \neq O$. For $y = O$, this limit only exists provided that the compatibility condition

$$\frac{1}{2}[Q : \nabla_x^2] f(x, O) + \sum_{i=1}^M \bar{q}_i D_i^y f(x, O) = 0 \quad (4.10)$$

holds. This shows that if $f \in \mathcal{C}_b^2(\bar{\mathcal{G}}) \cap D(\mathcal{A}_\Xi)$, then for every $x \in \mathbb{R}^2$ we must have $f(x, \cdot) \in D(\mathcal{A}_y)$, $\mathcal{A}_\Xi f(x, y) = \mathcal{A}_y f(x, y)$ for every $y \in \mathcal{G} \setminus O$, and the compatibility condition (4.10) must be satisfied as well.

From this, equation (4.2) follows from standard techniques. Indeed, for θ defined by (4.5), standard results imply that θ is continuous at $t = 0$ and satisfies the Kolmogorov equation $\partial_t \theta - \mathcal{A}_\Xi \theta = 0$ giving (4.2a) and (4.2c). Moreover, for positive time we must have $\theta(\cdot, t) \in \mathcal{C}_b^2(\bar{\mathcal{G}}) \cap D(\mathcal{A}_\Xi)$, and (4.10) gives the flux balance condition (4.2b) as desired. $\qquad \square$

5 Proof of Theorem 3.4

We devote this section to proving Theorem 3.4. Our proof resembles the proof in [HKP14], where a similar result appeared. The main difference in our situation is that we rely on Theorem 3.2 instead of the classical averaging principle. Our first task is to describe how far $Z(t)$ can travel inside a small neighbourhood of the separatrix. Given $\delta > 0$, define $\mathcal{V}^\delta \subset \mathcal{G}$ by

$$\mathcal{V}^\delta \stackrel{\text{def}}{=} \{(i, y) \in \mathcal{G} : |y| \leqslant \delta\} \, ,$$

and introduce two sequences of stopping times $\mu_n^{\varepsilon, \delta}$ and $\kappa_n^{\varepsilon, \delta}$ corresponding to successive visits to O and $\partial \mathcal{V}^\delta$. Namely, let $\mu_0^{\varepsilon, \delta} = \kappa_{-1}^{\varepsilon, \delta} = 0$ and then define recursively

$$\mu_n^{\varepsilon, \delta} = \inf\{t \geqslant \kappa_{n-1}^{\varepsilon, \delta} : \Gamma_\varepsilon(Z_t) \in \partial \mathcal{V}^\delta\} \qquad \text{and} \qquad \kappa_n^{\varepsilon, \delta} = \inf\{t \geqslant \mu_n^{\varepsilon, \delta} : \Gamma_\varepsilon(Z_t) \in O\} \, , \qquad (5.1)$$

for $n \geqslant 1$ and $n \geqslant 0$ respectively. Let $\Delta_n^\varepsilon = Z(\kappa_n^{\varepsilon, \delta}) - Z(\kappa_{n-1}^{\varepsilon, \delta})$, be the displacement between successive visits to $\mathcal{L}$. With this notation, the distance covered by $Z(t)$ before hitting $\Gamma_\varepsilon^{-1}(\partial \mathcal{V}^\delta)$, as well as the cell it then hits can be described as follows.

Proposition 5.1. *There exists a 2×2 non-degenerate matrix Q and a vector $(p_1, \ldots, p_M)$ such that the distributions of $(\varepsilon^{\frac{1-\alpha}{4}} \Delta_1^\varepsilon, \Gamma_\varepsilon(Z(\mu_1^{\varepsilon, \delta})))$ under $\mathbf{P}_\varepsilon^x$ converge, as $\varepsilon \to 0$, to the distribution of $(\sqrt{\delta \xi} \mathcal{N}(0, Q), \zeta)$, uniformly for $x \in \mathcal{L}$. Here, ξ, ζ, and $\mathcal{N}(0, Q)$ are three independent random variables such that ξ is exponentially distributed with parameter one, $\mathcal{N}(0, Q)$ is a two-dimensional normally distributed random variable with mean 0 and covariance matrix Q, and ζ is a $\mathcal{G}$-valued random variable that is almost surely at distance δ from O and $\mathbf{P}(\zeta \in I_i) = p_i$.*

Moreover, for each $\eta > 0$ there is $\delta_0 > 0$ such that

$$\lim_{\varepsilon \to 0} \sup_{x \in \mathcal{T}_0} \mathbf{P}_\varepsilon^x \Big(\varepsilon^{\frac{1-\alpha}{4}} \sup_{0 \leqslant t \leqslant \kappa_1^{\varepsilon, \delta}} |Z_t| > \eta\Big) < \eta, \qquad (5.2)$$

whenever $0 < \delta \leqslant \delta_0$.

A similar result was proved in Section 2 of [HKP14]. However, in order to make this paper self-contained, we sketch the main steps involved in the proof and explain the necessary modifications in Appendix A. Although we will not use it explicitly, we remark that the p_i's above are proportional to the q_i. This follows from the proof of Proposition 5.1, and Corollary 2.4 in [FS00].

Let now $\mathcal{X}_\mathcal{G}$ denote the space of $\mathcal{G}$-valued excursions. In other words, elements $h \in \mathcal{X}_\mathcal{G}$ are continuous functions $h \in \mathcal{C}(\mathbb{R}_+, \mathcal{G})$ with the property that, if $h(t) = O$ for some $t \geqslant 0$, then $h(s) = O$ for all $s \geqslant t$. Furthermore, we impose that $T(h) = \inf\{t \geqslant 0 : h(t) = O\}$ is finite for every $h \in \mathcal{X}_\mathcal{G}$. We turn $\mathcal{X}_\mathcal{G}$ into a metric space by setting

$$d(h, \bar{h}) = |T(h) - T(\bar{h})| + \sup_{t \geqslant 0} d_\mathcal{G}(h(t), \bar{h}(t)) \, , \qquad (5.3)$$

with $d_\mathcal{G}$ as in (3.1).

We also write $\mathcal{X}^\infty = (\mathbb{R}^2 \times \mathcal{X}_\mathcal{G})^{\mathbb{N}}$, endowed with the topology of pointwise convergence, and we define a "projection" $\mathcal{P}_\delta : \mathcal{C}(\mathbb{R}_+; \bar{\mathcal{G}}) \to \mathcal{X}^\infty$ as follows. Given an element $\omega \in \mathcal{C}(\mathbb{R}_+; \bar{\mathcal{G}})$, we write $\omega = (V, G)$ where V and G are continuous $\mathbb{R}^2$-valued and $\mathcal{G}$-valued functions respectively. We first define the "stopping times" $\mu_n^\delta(\omega)$ and $\kappa_n^\delta(\omega)$ as in (5.1), with $\Gamma_\varepsilon(Z)$ replaced by G. We then write $J_n(\omega) \in \mathcal{X}_\mathcal{G}$

for the nth downcrossing of the process G. In other words, suppressing the argument ω for conciseness, we have

$$J_n(t) = (G((t + \mu_n^\delta) \wedge \kappa_n^\delta)) \,,$$

so that in particular $|J_n(\omega)(0)| = \delta$ for $n > 0$ and $T(J_n(\omega)) = \kappa_n^\delta - \mu_n^\delta$. We also define $U_n(\omega) \in \mathbb{R}^2$ by $U_n(\omega) = V(\mu_{n+1}^\delta) - V(\kappa_n^\delta)$. With these notations at hand, we set

$$\mathcal{P}_\delta(\omega) = (U_n(\omega), J_n(\omega))_{n \geqslant 0} \,.$$

We then have the following lemma.

Lemma 5.2. *Let $\hat{\nu}_\varepsilon$ be a family of probability measures on $\mathcal{T}_0$ such that the push forward measures $\Gamma_\varepsilon^* \hat{\nu}_\varepsilon$ converge weakly, as $\varepsilon \to 0$, to a probability measure $\hat{\nu}$ on $\mathcal{G}$. Then, the law of $\mathcal{P}_\delta(\widehat{\Gamma}_\varepsilon(Z))$ converges weakly under $\mathbf{P}_\varepsilon^{\hat{\nu}_\varepsilon}$, as $\varepsilon \to 0$, to the law of $\mathcal{P}_\delta(\Xi)$ under $\bar{\mathbf{P}}^{\hat{\nu}}$.*

Proof. We first note that, under $\bar{\mathbf{P}}^{\hat{\nu}}$, $\mathcal{P}_\delta(\Xi)$ is a random vector $(U_n, J_n)_{n \geqslant 0}$ with independent components. The distribution of U_n is as in Proposition 5.1, i.e., it is equal to the distribution of $\sqrt{\delta}\xi \mathcal{N}(0, Q)$. This follows from the fact that the distribution of the local time accumulated up to μ_1^δ under $\bar{\mathbf{P}}^O$ is the same as that of $\delta\xi$. Indeed, exponentiality follows from the fact that L_t can only grow when $Y_t = 0$, while the expectation is given by applying (4.7) to the function $f(y) = d_\mathcal{G}(y, O)$, plugging in $t = \mu_1^\delta$, and taking expectations (see e.g. Exercise 4.12 Chapter VI in [RY99]). The distribution of J_0 is the distribution of $J_0(Y)$ under $\bar{\mathbf{P}}^\mu$, while the distribution for each of the J_n for $n > 0$ is equal to the distribution of $J_0(Y)$ under $\bar{\mathbf{P}}^\zeta$, where ζ is as in Proposition 5.1. The fact that these are independent follows from the strong Markov property, combined with the fact that the location at which the Y-component of the process first hits $\mathcal{V}^\delta$ is independent of the local time accumulated until then.

We then see that, by Theorem 3.2, the law of $J_0(\widehat{\Gamma}_\varepsilon(Z))$ under $\mathbf{P}_\varepsilon^{\hat{\nu}_\varepsilon}$ does indeed converge weakly as $\varepsilon \to 0$ to the law of $J_0(Y)$ under $\bar{\mathbf{P}}^{\hat{\mu}}$. This is because, although the map $Y \mapsto J_0(Y)$ is not continuous, its points of discontinuity, which consist precisely of those paths which either never hit O or such that their first hit of O is not transverse, are of measure 0 under $\bar{\mathbf{P}}^{\hat{\nu}}$.

The convergence of the other components of the random vector $(U_n, J_n)_{n \geqslant 0}$ to their respective limits follows in the same way from Proposition 5.1, combined with Theorem 3.2. The independence of the components of the limiting vector immediately follows from the strong Markov property of the process Z, the fact that the convergences in Proposition 5.1 and Theorem 3.2 are uniform with respect to the initial condition, and the fact that ζ is independent of the other limiting random variables in Proposition 5.1. $\quad\square$

Proof of Theorem 3.4. We first note that as a consequence of the periodicity of the problem, we can (and will henceforth) restrict ourselves to the case when the probability measure ν_ε is concentrated on $\mathcal{T}_0$, so that the limiting measure $\hat{\nu}$ is of the form $\hat{\nu} = \delta_0 \otimes \nu$ for some probability measure ν on $\mathcal{G}$. We thus only need to prove that, under the conditions of the theorem, $\widehat{\Gamma}_\varepsilon(Z)$ converges in law to Ξ with initial measure $\delta_0 \otimes \nu$. We begin by defining a "concatenation" map $\mathcal{R}_\delta \colon \mathcal{X}^\infty \to \mathcal{C}(\mathbb{R}_+; \bar{\mathcal{G}})$ as follows. Given $\delta > 0$, $U, V \in \mathbb{R}^2$ and $G = (i, y) \in \mathcal{G}$, we define the interpolation $L_\delta(U, V, G) \colon [0, \delta^2] \to \bar{\mathcal{G}}$ by

$$L_\delta(U, V, G)(t) \stackrel{\text{def}}{=} (U + \delta^{-2}t(V - U), (i, \delta^{-2}ty)) \,,$$

so that $L_\delta(U, V, G)(0) = (U, O)$ and $L_\delta(U, V, G)(\delta^2) = (V, G)$. Given $\mathcal{X} = (\mathcal{X}_n)_{n \geqslant 0}$ with $\mathcal{X}_n = (U_n, J_n) \in \mathbb{R}^2 \times \mathcal{X}_\mathcal{G}$, we define recursively two sequences of "excursion times" $E_n, E_n' \in \mathbb{R}_+$ and
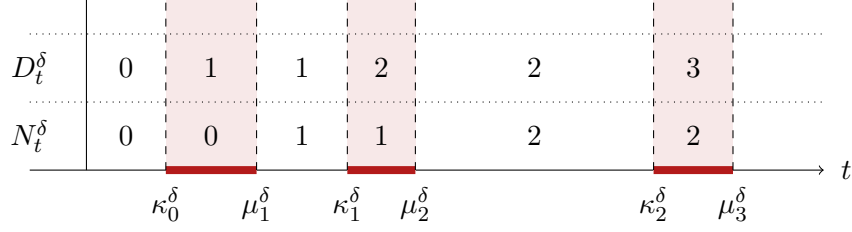
Figure 4: Values of D_t^δ and N_t^δ in relation to the stopping times κ_i and μ_i. The highlighted regions are when the upcrossings from O to $\mathcal{V}^\delta$ occur. The range of e_δ falls into the non-highlighted regions, while the range of $\bar{e}_\delta$ falls into the region obtained by shrinking / expanding the highlighted regions in such a way that each of them has length δ^2.

locations $W_n \in \mathbb{R}^2$ by

$$E_n(\mathscr{X}) \overset{\text{def}}{=} n\delta^2 + \sum_{j=0}^{n-1} T(J_j)\,, \qquad E_n'(\mathscr{X}) = E_n(\mathscr{X}) + T(J_n)\,, \qquad W_n = \sum_{j=0}^{n-1} U_j\,.$$

with the natural conventions that $E_0 = 0$ and $W_0 = 0$. With these notations at hand, we then set

$$\mathcal{R}_\delta(\mathscr{X})(t) = \begin{cases} (W_n, J_n(t - E_n(\mathscr{X}))) & \text{for } t \in [E_n(\mathscr{X}), E_n'(\mathscr{X})], \\ L_\delta(W_n, W_{n+1}, J_{n+1}(0))(t - E_n'(\mathscr{X})) & \text{for } t \in [E_n'(\mathscr{X}), E_n'(\mathscr{X}) + \delta^2]. \end{cases}$$

This definition is unambiguous (and the function $\mathcal{R}_\delta(\mathscr{X})$ is continuous) since, at $t = E_n'(\mathscr{X})$, both expressions equal (W_n, O), while at $t = E_{n+1}(\mathscr{X}) = E_n'(\mathscr{X}) + \delta^2$ both expressions equal $(W_{n+1}, J_{n+1}(0))$.

It is straightforward to see that $\mathcal{R}_\delta$ is a right inverse for $\mathcal{P}_\delta$, i.e. $\mathcal{P}_\delta \mathcal{R}_\delta = \text{id}$. On the other hand, clearly $\mathcal{R}_\delta \mathcal{P}_\delta \neq \text{id}$, however, we will construct a set of trajectories ω on which $(\mathcal{R}_\delta \mathcal{P}_\delta)(\omega)$ is close to ω. For this, we need a bit of additional notation. Given a trajectory $\omega \in \mathcal{C}(\mathbb{R}_+; \bar{\mathcal{G}})$ and times $\mu_n^\delta(\omega)$ and $\kappa_n^\delta(\omega)$ as above, we define the corresponding downcrossing and upcrossing durations by

$$T_{n,\delta}^d = \kappa_n^\delta - \mu_n^\delta, \qquad T_{n,\delta}^u = \mu_n^\delta - \kappa_{n-1}^\delta, \qquad n \geqslant 0\,.$$

We also define the number of down / upcrossings up to time t by

$$D_t^\delta = \inf\{n \geqslant 0 : \kappa_n^\delta \geqslant t\}\,, \qquad N_t^\delta = \sup\{n \geqslant 0 : \mu_n^\delta \leqslant t\}\,,$$

see Figure 4, as well as the quantities

$$e_\delta(t) = t + \sum_{n=1}^{\bar{N}_t^\delta} T_{n,\delta}^u\,, \quad \bar{e}_\delta(t) = t + \delta^2 \bar{N}_t^\delta\,, \quad \bar{N}_t^\delta = \sup\Big\{k \geqslant 0 : \mu_k^\delta \leqslant t + \sum_{n=1}^k T_{n,\delta}^u\Big\}\,.$$

These quantities can be interpreted as follows: we stop a 'special' clock every time the process hits the vertex O, and re-start it once the process reaches the level set $\partial\mathcal{V}^\delta$. Then $e_\delta(t)$ is the real time that has elapsed when the special clock reaches time t, $\bar{N}_t^\delta$ is the number of upcrossings completed before this happens, and $\bar{e}_\delta(t)$ is the analogous quantity to $e_\delta(t)$ when, for every upcrossing, we count δ^2 in real time.

Given $\eta, r, T, \delta > 0$, we then define a set $\mathcal{F}(\eta, r, T, \delta)$ of trajectories ω such that the following properties hold:

1. For all $s, t \leqslant T + r$ with $|t - s| \leqslant r$, one has $d_{\bar{\mathcal{G}}}(\omega(t), \omega(s)) \leqslant \eta$.

2. One has the bounds $e_\delta(T) - T \leqslant r$ and $\delta^2(2 + \bar{N}_T^\delta(\omega)) \leqslant \eta$.

3. Writing $\omega = (V, G)$ as above, for every $n \in \{0, \ldots, \bar{N}_T^\delta(\omega)\}$, one has the bounds

$$\sup_{t \in [\mu_n^\delta, \kappa_n^\delta]} |V(t) - V(\mu_n^\delta)| \leqslant \delta^2, \qquad \text{and} \qquad |V(0)| \leqslant \delta^2 \ .$$

The following lemmas are the crucial ingredients for our proof of the theorem.

Lemma 5.3. *Provided that $r > \delta^2$, every $\omega \in \mathcal{F}(\eta, r, T, \delta)$ satisfies the bound*

$$\sup_{t \in [0,T]} d_{\bar{\mathcal{G}}}(\omega(t), (\mathcal{R}_\delta \mathcal{P}_\delta \omega)(t)) \leqslant 4\eta \ . \tag{5.4}$$

Lemma 5.4. *For every $\eta > 0$, every $T > 0$, and every sequence ν_ε of probability measures on $\mathbb{R}^2$ such that $\Gamma_\varepsilon^* \nu_\varepsilon$ is tight, there exist $\delta_0 > 0$, $r \geqslant \delta_0^2$ such that for $\delta \in (0, \delta_0)$, there is $\varepsilon_0 > 0$ such that*

$$\mathbf{P}_\varepsilon^{\nu_\varepsilon}(\widehat{\Gamma}_\varepsilon(Z) \in \mathcal{F}(\eta, r, T, \delta)) \geqslant 1 - \eta \ ,$$

for every $\varepsilon \leqslant \varepsilon_0$.

Lemma 5.5. *We have*

$$\sup_{y \in \mathcal{G}} \bar{\mathbf{E}}^y |e_\delta(t) - t| = \mathcal{O}(\delta) \qquad \text{and} \qquad \sup_{y \in \mathcal{G}} \bar{\mathbf{E}}^y |\delta(\bar{N}_t^\delta - D_t^\delta)| \xrightarrow{\delta \to 0} 0 \ .$$

Lemma 5.6 ([FS00]). *For every $t > 0$, we have $\lim_{\delta \to 0} \sup_{y \in \mathcal{G}} \bar{\mathbf{E}}^y |\delta D_t^\delta - L_t| = 0$.*

Lemma 5.6 is contained in in the proof of Lemma 2.2 in [FS00], and we do not prove it here. For clarity of presentation the proofs of Lemmas 5.3–5.5 are postponed until the proof of Theorem 3.4 is complete.

The rest of the proof is a standard "triangle" argument. Fix $T > 0$ and let f be a uniformly continuous bounded functional on $\mathcal{C}([0, T]; \bar{\mathcal{G}})$. Pick any $\eta' > 0$ and choose $\eta > 0$ small enough such that

$$|f(\omega) - f(\omega')| < \eta' \qquad \text{whenever} \qquad \sup_{t \in [0,T]} d_{\bar{\mathcal{G}}}(\omega(t), \omega'(t)) < 4\eta.$$

Note that the reconstruction map $\mathcal{R}_\delta : \mathcal{X}^\infty \to \mathcal{C}(\mathbb{R}_+, \bar{\mathcal{G}})$ is continuous with the choice of the metric (5.3). Since the restriction operator $\Pi_{[0,T]} : \mathcal{C}(\mathbb{R}_+, \bar{\mathcal{G}}) \to \mathcal{C}([0, T], \bar{\mathcal{G}})$ is also continuous, Lemma 5.2 implies that

$$\mathbf{E}_\varepsilon^{\nu_\varepsilon} f(\Pi_{[0,T]} \mathcal{R}_\delta \mathcal{P}_\delta(\widehat{\Gamma}_\varepsilon(Z))) \xrightarrow{\varepsilon \to 0} \bar{\mathbf{E}}^{\hat{\nu}} f(\Pi_{[0,T]} \mathcal{R}_\delta \mathcal{P}_\delta(\Xi)) \ . \tag{5.5}$$

On the other hand, Lemma 5.3 and Lemma 5.4 imply that we can find a $\delta > 0$ such that for any small enough $\varepsilon > 0$, we have

$$|\mathbf{E}_\varepsilon^{\nu_\varepsilon} f(\Pi_{[0,T]} \widehat{\Gamma}_\varepsilon(Z)) - \mathbf{E}_\varepsilon^{\nu_\varepsilon} f(\Pi_{[0,T]} \mathcal{R}_\delta \mathcal{P}_\delta(\widehat{\Gamma}_\varepsilon(Z)))| \leqslant 2\eta \|f\|_\infty + \eta'. \tag{5.6}$$

For the limiting process Ξ, standard results on the Brownian modulus of continuity, and Lemmas 5.5–5.6 imply that, by possibly making δ smaller, we have

$$\bar{\mathbf{P}}^{\hat{\nu}}(\Xi \in \mathcal{F}(\eta, r, T, \delta)) \geqslant 1 - \eta \ .$$

Together with Lemma 5.3, this implies

$$|\bar{\mathbf{E}}^{\hat{\nu}} f(\Pi_{[0,T]}\Xi) - \bar{\mathbf{E}}^{\hat{\nu}} f(\Pi_{[0,T]}\mathcal{R}_\delta\mathcal{P}_\delta(\Xi)| \leqslant 2\eta\|f\|_\infty + \eta'. \tag{5.7}$$

Combining (5.5), (5.6), (5.7) and noting that η and η' can be made arbitrarily small gives the convergence of $\Pi_{[0,T]}\widehat{\Gamma}_\varepsilon(Z)$ in law to $\Pi_{[0,T]}\Xi$. Since $T > 0$ was also arbitrary, this finishes the proof. □

It remains to prove Lemmas 5.3–5.5.

Proof of Lemma 5.3. For fixed small $\lambda > 0$ and any $T > 0$, consider $\omega = (V, G) \in \mathcal{F}(\eta, r, T, \delta)$. We want to show that if δ and η are sufficiently small, then, writing $\omega^\delta = (V^\delta, G^\delta) := \mathcal{R}_\delta\mathcal{P}_\delta\omega$, one has $d_{\bar{\mathcal{G}}}(\omega^\delta(t), \omega(t)) \leqslant \lambda$ for $t \leqslant T$.

For this, we first build the time change

$$f_\delta(t) = e_\delta(\inf\{s \,:\, \bar{e}_\delta(s) \geqslant t\}) \,.$$

(On the range of $\bar{e}_\delta$, this equals $e_\delta(\bar{e}_\delta^{-1}(t))$.) The second property of $\mathcal{F}$ then guarantees that

$$|f_\delta(t) - t| \leqslant r \,, \qquad \forall t \leqslant T \,. \tag{5.8}$$

It also follows from the constructions of $\mathcal{R}_\delta$ and $\mathcal{P}_\delta$ that, for all t in the range of $\bar{e}_\delta$, one has

$$G^\delta(t) = G(f_\delta(t)) \,, \qquad V^\delta(t) = \sum_{n=1}^{\bar{N}_t^\delta}(V(\mu_n^\delta) - V(\kappa_{n-1}^\delta)) \,. \tag{5.9}$$

We can rewrite the second identity as

$$V^\delta(t) = V(f_\delta(t)) - V(0) - (V(f_\delta(t)) - V(\mu_{\bar{N}_t^\delta}^\delta)) - \sum_{n=1}^{\bar{N}_t^\delta}(V(\kappa_{n-1}^\delta) - V(\mu_{n-1}^\delta)) \,.$$

Since we have $f_\delta(t) \in [\mu_n^\delta, \kappa_n^\delta]$ for $n = \bar{N}_t^\delta$ by definition, we can combine this with the third property of $\mathcal{F}$, thus yielding the bound $|V^\delta(t) - V(f_\delta(t))| \leqslant (2 + \bar{N}_t^\delta)\delta^2 \leqslant \eta$. Together with the first equality in (5.9) and the first property of $\mathcal{F}$, this finally yields

$$d_{\bar{\mathcal{G}}}(\omega(t), \omega^\delta(t)) \leqslant d_{\bar{\mathcal{G}}}(\omega(t), \omega(f_\delta(t))) + |V^\delta(t) - V(f_\delta(t))| \leqslant 2\eta,$$

for all times $t \leqslant T$ belonging to the range of $\bar{e}_\delta$. It remains to consider times outside the range of $\bar{e}_\delta$, which correspond to the upcrossings. Write $t_0 < t$ for the start of the upcrossing, so that $|t - t_0| \leqslant \delta^2 < r$ by definition. Then, one has

$$d_{\bar{\mathcal{G}}}(\omega(t), \omega^\delta(t)) \leqslant d_{\bar{\mathcal{G}}}(\omega(t), \omega(t_0)) + d_{\bar{\mathcal{G}}}(\omega(t_0), \omega^\delta(t_0)) + d_{\bar{\mathcal{G}}}(\omega^\delta(t), \omega^\delta(t_0)) \leqslant 4\eta \,.$$

This is because the first and last terms are bounded by η as a consequence of the first and second properties of $\mathcal{F}$, while the second term is bounded by 2η from before. □

Since Lemma 5.5 is used in the proof of Lemma 5.4, we prove it first.

Proof of Lemma 5.5. By the strong Markov property, the $T_{i,\delta}^u$ are independent and identically distributed under $\bar{\mathbf{P}}^y$, while the $T_{i,\delta}^d$ are identically distributed, but not independent of the upcrossing durations in between. However, when conditioned on the corresponding downcrossing taking place on edge j, they have the same distribution as the hitting time of the point $\delta/\sqrt{a_j}$ by a standard Brownian motion starting at the origin.

We note that for any $\lambda > 0$ and $K \in \mathbb{N}$, Chebyshev's inequality implies

$$\bar{\mathbf{P}}^y(\bar{D}_t^\delta > K) = \bar{\mathbf{P}}^y\Big(\sum_{i=0}^K T_{i,\delta}^d < t\Big) \leqslant e^{\lambda t}\bar{\mathbf{E}}^y e^{-\lambda\sum_{i=0}^K T_{i,\delta}^d} = e^{\lambda t}\prod_{i=0}^K \bar{\mathbf{E}}^y e^{-\lambda T_{i,\delta}^d},$$

where we defined $\bar{D}_t^\delta = D_{e_\delta(t)}^\delta$. Let j be the index of the slowest edge, that is $a_j = \min_{i=1,\dots,n} a_i$. Then, by the strong Markov property,

$$\bar{\mathbf{E}}^y e^{-\lambda T_{i,\delta}^d} \leqslant \bar{\mathbf{E}}^{(j,\delta)} e^{-\lambda\tau_0^\delta} = e^{-\delta b_\lambda}, \qquad b_\lambda = \sqrt{\frac{2\lambda}{a_j}}.$$

Inserting this into the above yields

$$\bar{\mathbf{P}}^y(\bar{D}_t^\delta > K) \leqslant \exp\left(\lambda t - b_\lambda K\delta\right). \tag{5.10}$$

Next, writing $e_\delta(t) - t = \sum_{i=1}^\infty \mathbf{1}_{i \leqslant \bar{D}_t^\delta} T_{i,\delta}^u$, we obtain from the Cauchy-Schwartz inequality,

$$\bar{\mathbf{E}}^y|e_\delta(t) - t| = \bar{\mathbf{E}}^y(e_\delta(t) - t) = \sum_{i=1}^\infty \sqrt{\mathbf{P}(i \leqslant \bar{D}_t^\delta)\bar{\mathbf{E}}^y(T_{i,\delta}^u)^2} \leqslant Ce^{\lambda t/2}\delta^2\sum_{i=1}^\infty e^{-\delta b_\lambda i/2} \leqslant C(t)\delta, \tag{5.11}$$

where we used (5.10), combined with the fact that $\bar{\mathbf{E}}^y(T_{i,\delta}^u)^2 = \mathcal{O}(\delta^4)$ by the Brownian scaling.

To prove the second claim, pick an $\eta > 0$. By the monotonicity of D_t^δ in t and since $|\bar{N}_t^\delta - \bar{D}_t^\delta| \leqslant 1$,

$$\bar{\mathbf{E}}^y|\delta(\bar{N}_t^\delta - D_t^\delta)| \leqslant \delta + \delta\bar{\mathbf{E}}^y(D_{t+\eta}^\delta - D_t^\delta) + \delta\bar{\mathbf{E}}^y(\bar{D}_t^\delta \mathbf{1}_{\{e_\delta(t) - t > \eta\}}). \tag{5.12}$$

The expectation $\delta\bar{\mathbf{E}}^y(D_{t+\eta}^\delta - D_t^\delta)$ above can be estimated by comparing it to the local time. Using the Markov property at time t and Lemma 5.6, we have

$$\delta\bar{\mathbf{E}}^y(D_{t+\eta}^\delta - D_t^\delta) \leqslant \bar{\mathbf{E}}^y\bar{\mathbf{E}}^{Y(t)}L_\eta + o(1).$$

The right hand side of the above can be made arbitrarily small by choosing η small enough.

On the other hand, the Cauchy-Schwartz inequality implies that the last term on the right hand side of (5.12) can be estimated from above by

$$\delta\sqrt{\bar{\mathbf{E}}^y(\bar{D}_t^\delta)^2\,\bar{\mathbf{P}}^y(e_\delta(t) - t > \eta)} = o(1).$$

Indeed, the probability converges to zero by (5.11) and Chebyshev's inequality, while the remaining factor can be bounded using (5.10), thus concluding the proof. $\qquad\square$

Finally, we turn to Lemma 5.4. The proof relies on tightness (stated as Lemma 5.7, below) and the fact that upcrossing durations are negligible compared to the downcrossings durations (Lemma 5.5).

Lemma 5.7. *The law of $\widehat{\Gamma}_\varepsilon(Z)$ under $\mathbf{P}^{\nu_\varepsilon}$ is tight in $\mathcal{C}(\mathbb{R}_+, \bar{\mathcal{G}})$. In particular, for every $T, \eta > 0$, there is an $r > 0$ and $\varepsilon_0 > 0$ such that for $\varepsilon \in (0, \varepsilon_0]$, we have*

$$\mathbf{P}^{\nu_\varepsilon}\Big(\sup_{\substack{|t-s| \leqslant r \\ s,t \in [0,T]}} d_{\bar{\mathcal{G}}}(\widehat{\Gamma}_\varepsilon(Z(t)), \widehat{\Gamma}_\varepsilon(Z(s))) \geqslant \eta \Big) < \eta$$

Proof. The tightness of the $\mathcal{G}$-component follows from Theorem 3.2, we only have to prove the tightness of the $\mathbb{R}^2$ component. Using the strong Markov property and the fact that the displacement is bounded by $\mathcal{O}(\varepsilon)$ as long as the process remains inside a cell, tightness of the $\mathbb{R}^2$ component reduces to showing the following: for every $\eta > 0$ and $r \in (0,1)$ sufficiently small, there exists $\varepsilon_0 = \varepsilon_0(\eta, r) > 0$ such that

$$\mathbf{P}_\varepsilon^x\Big(\sup_{0 \leqslant t \leqslant r} |\varepsilon^{\frac{1-\alpha}{4}} Z_t| > \eta \Big) \leqslant r\eta , \tag{5.13}$$

for every $\varepsilon \leqslant \varepsilon_0$ and $x \in \mathcal{L} \cap \mathcal{T}_0$ (see also Theorem 18.17 in [KS07]). We prove this below.

Let R_1^δ, R_2^δ, etc. be independent, distributed as $\sqrt{\delta}\xi\mathcal{N}(0, Q)$. Doob's maximal inequality then shows that there exists a constant C such that

$$\mathbf{P}\Big(l^{-1/2} \max_{1 \leqslant m \leqslant l} |R_1^\delta + \cdots + R_m^\delta| > K \Big) \leqslant \frac{C\delta^5}{K^{10}} \mathbf{E}|R_1^\delta|^{10} ,$$

for all $K > 0$. Choosing $K = \frac{1}{4}\eta\sqrt{\delta/k}$, we see that for a given $\eta > 0$, there exist $k_0 \in (0,1)$ and $\delta_1 > 0$ such that

$$\mathbf{P}\Big(\max_{1 \leqslant m \leqslant k/\delta} |R_1^\delta + \cdots + R_m^\delta| > \frac{\eta}{4} \Big) \leqslant \frac{k^4 \eta}{4} , \tag{5.14}$$

whenever $k \in (0, k_0)$ and $\delta \in (0, \delta_1)$. From (5.14) and Lemma 5.2, it follows that there is $\varepsilon_1(k, \delta) > 0$ such that

$$\mathbf{P}_\varepsilon^x\Big(\max_{1 \leqslant m \leqslant k/\delta} \varepsilon^{(1-\alpha)/4} |\Delta_1^\varepsilon + \cdots + \Delta_m^\varepsilon| > \frac{\eta}{3} \Big) \leqslant \frac{k^4 \eta}{3} , \tag{5.15}$$

provided that $\varepsilon \leqslant \varepsilon_1(k, \delta)$. Note that this estimate and those below are uniform in $x \in \mathcal{L} \cap \mathcal{T}_0$. Combining (5.15) and (5.2), it now follows that there is $\varepsilon_2(k, \delta) > 0$ such that

$$\mathbf{P}_\varepsilon^x\Big(\sup_{0 \leqslant t \leqslant \kappa_{[k/\delta]}^{\delta,\varepsilon}} \varepsilon^{(1-\alpha)/4} |Z_t| > \eta/2 \Big) \leqslant \frac{k^4 \eta}{2} . \tag{5.16}$$

provided that $\varepsilon \leqslant \varepsilon_2(k, \delta)$.

By Lemma 5.6, for a given $\eta > 0$, we can find $r > 0$ and $\delta_2 = \delta_2(r) > 0$ such that, for any $\delta \leqslant \delta_2$, and $\ell = [r^{1/4}/\delta]$ we have

$$\sup_{y \in \mathcal{G}} \bar{\mathbf{P}}^y(D_r^\delta \geqslant \ell) < \sup_{y \in \mathcal{G}} \bar{\mathbf{P}}^y(L_r \geqslant r^{1/4}) + \frac{\eta r}{4} \leqslant r^2 \bar{\mathbf{E}}^O\Big(\frac{L_r}{r^{1/2}} \Big)^8 + \frac{\eta r}{4} \leqslant \frac{\eta r}{3}. \tag{5.17}$$

Here the second inequality follows from the Chebyshev inequality and the strong Markov property, while the last inequality follows from the fact that the distribution of $L_r/r^{1/2}$ under $\bar{\mathbf{P}}^O$ does not depend on $r > 0$ and has Gaussian tails. As a consequence of Theorem 3.2, there is $\varepsilon_3(r, \delta)$ such that if $\varepsilon \leqslant \varepsilon_3(r, \delta)$, and $x \in \mathcal{L}$ we have

$$\mathbf{P}_\varepsilon^x(\kappa_\ell^{\delta,\varepsilon} < r) \leqslant \bar{\mathbf{P}}^O(D_r^\delta \geqslant \ell) + \frac{\eta r}{6}, \tag{5.18}$$

and hence

$$\mathbf{P}_\varepsilon^x \left(\sup_{0 \leqslant t \leqslant r} \varepsilon^{(1-\alpha)/4} |Z_t| > \eta \right) \leqslant \mathbf{P}_\varepsilon^x (\kappa_\ell^{\delta,\varepsilon} < r) + \mathbf{P}_\varepsilon^x \left(\sup_{0 \leqslant t \leqslant \kappa_\ell^{\delta,\varepsilon}} \varepsilon^{(1-\alpha)/4} |Z_t| > \eta \right).$$

Applying (5.16), (5.17), and (5.18) with

$$k = r^{1/4}, \qquad \delta < \min(\delta_1, \delta_2) \qquad \text{and} \qquad \varepsilon < \min(\varepsilon_1(k, \delta), \varepsilon_2(k, \delta), \varepsilon_3(r, \delta)),$$

we obtain (5.13) as required. $\qquad\square$

Finally we prove Lemma 5.4.

Proof of Lemma 5.4. Fix $\eta > 0$ and $T > 0$. As a consequence of Lemma 5.7, we can find ε_0 and $r > 0$ such that the first property is satisfied with probability at least $1 - \eta$, uniformly over $\varepsilon < \varepsilon_0$. By Lemma 5.5, we then choose δ with $\delta^2 < r$ sufficiently small so that the second estimate holds. The third bound immediately follows from the definitions as soon as $\varepsilon^{\frac{1-\alpha}{4}} \leqslant \delta^2$, thus concluding the proof. $\qquad\square$

6   The averaging principle on the short time scales

We now turn to the proof of Theorem 3.2. For notational simplicity, we view Z itself as a process on the torus $\mathcal{T}$ and define set $Y_t^\varepsilon = \Gamma_\varepsilon(Z_t)$. Let $\Psi \subset \mathcal{C}_0(\mathcal{G})$ be the dense subset consisting of all compactly supported functions that are continuously differentiable on each edge.

The proof of Theorem 3.2 relies on the following two lemmas (compare with the result of Freidlin and Wentzell [FW12, Ch. 8, Lemma 3.1]).

Lemma 6.1. *Let $\mathcal{A}$ be the operator on the domain $D(\mathcal{A})$ introduced in Section 3 and $\mathcal{D} \subset D(\mathcal{A})$ be the subset consisting of all the functions f for which $\mathcal{A}f \in \Psi$. For each $f \in \mathcal{D}$, $T > 0$, we have*

$$\sup_{x \in \mathcal{T}} \left| \mathbf{E}_\varepsilon^x \left[f(Y_T^\varepsilon) - f(Y_0^\varepsilon) - \int_0^T \mathcal{A}f(Y_t^\varepsilon)\, dt \right] \right| \xrightarrow{\varepsilon \to 0} 0. \tag{6.1}$$

Lemma 6.2. *For each compact set $K \subset \mathcal{G}$, the laws of the processes $\{Y^\varepsilon\}$ under the measures $\mathbf{P}_\varepsilon^x$ for $\varepsilon \in (0, 1]$ and $x \in \Gamma_\varepsilon^{-1}(K) \subset \mathcal{T}$ are tight.*

Momentarily postponing the proof of Lemmas 6.1 and 6.2, we use them to prove Theorem 3.2.

Proof of Theorem 3.2. Lemma 6.1, the Markov property, and the time-independence of $\mathcal{A}$ ensure that any subsequential limit of Y^ε solves the martingale problem with the operator $\mathcal{A}' = \mathcal{A}|_\mathcal{D}$. By the Ito formula, Y is a solution of the martingale problem for $\mathcal{A}'$ and any initial measure μ on $\mathcal{G}$. Since $\mathcal{D}$ is a core for $\mathcal{A}$ (as one can easily verify using [EK86, Ch 1, Proposition 3.1]), Theorem 4.1 in [EK86, Ch 4] implies that Y is actually the unique solution to the martingale problem for $\mathcal{A}$ with any initial measure μ on $\mathcal{G}$. Moreover, it also follows that the subsequential limits are solutions of the martingale problem for $\mathcal{A}$ as well. Therefore any subsequential limit of Y^ε must equal Y and Lemma 6.2 and hence Prokhorov's theorem imply the convergence of Y^ε itself. $\qquad\square$

The rest of the section is devoted to the proof of Lemma 6.1, while the proofs of several auxiliary lemmas and of Lemma 6.2 are relegated to the next section. Some elements in our proof are similar to those used in [DK08, DK13], where a different extension of the original averaging principle (Chapter 8 of [FW12]) was addressed.

For $h \geqslant 0$, we will write $\mathcal{L}(h) = \{x \in \mathcal{T} : |H(x)| = h\}$, so that $\mathcal{L}_\mathcal{T} = \mathcal{L}(0)$. Take a function $\beta = \beta(\varepsilon) \in (\alpha/2, 1/2)$ such that

$$\beta(\varepsilon) - \frac{\alpha}{2} \to 0 \quad \text{and} \quad \varepsilon^{\beta(\varepsilon)-\frac{\alpha}{2}} \to 0 \quad \text{as} \quad \varepsilon \to 0. \tag{6.2}$$

Denote $\bar{\mathcal{L}} = \mathcal{L}(\varepsilon^\beta)$ and let σ be the first time when the process Z_t reaches $\mathcal{L}_\mathcal{T}$ (this coincides with β_0^ε introduced earlier) and τ be the first time when it reaches $\bar{\mathcal{L}}$. We inductively define the following two sequences of stopping times. Let $\sigma_1 = \sigma$. For $n \geqslant 1$ let τ_n be the first time following σ_n when the process reaches $\bar{\mathcal{L}}$. For $n \geqslant 2$ let σ_n be the first time following τ_{n-1} when the process reaches $\mathcal{L}_\mathcal{T}$.

We can consider the following discrete time Markov chains $\xi_n^1 = Z_{\sigma_n}$ and $\xi_n^2 = Z_{\tau_n}$ with state spaces $\mathcal{L}_\mathcal{T}$ and $\bar{\mathcal{L}}$, respectively. Let $P_1(x, dy)$ and $P_2(x, dy)$ be transition operators for the Markov chains ξ_n^1 and ξ_n^2, respectively. It was then shown in [DK08, Lem 2.3] that they are uniformly exponentially mixing in the following sense.

Lemma 6.3. *There exist constants $0 < c < 1$, $\varepsilon_0 > 0$, $n_0 > 0$, and probability measures ν and μ (which depend on ε) on $\mathcal{L}_\mathcal{T}$ and $\bar{\mathcal{L}}$, respectively, such that for $\varepsilon < \varepsilon_0$ and $n \geqslant n_0$ we have*

$$\sup_{x \in \mathcal{L}_\mathcal{T}} \|P_1^n(x, \cdot) - \nu\|_{\mathrm{TV}} \leqslant c^n, \quad \sup_{x \in \bar{\mathcal{L}}} \|P_2^n(x, \cdot) - \mu\|_{\mathrm{TV}} \leqslant c^n, \tag{6.3}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm of a measure.

We will need to control the number of excursions between c and $\mathcal{L}_\mathcal{T}$ before time T. This is our next lemma.

Lemma 6.4. *There is a constant $r > 0$ such that for all sufficiently small ε we have*

$$\sup_{x \in \bar{\mathcal{L}}} \mathbf{E}_\varepsilon^x \exp(-\sigma) \leqslant 1 - r\varepsilon^{\beta-\frac{\alpha}{2}}.$$

The proof of this lemma is given in Section 7. Using the Markov property of the process and Lemma 6.4, we get the estimate

$$\sup_{x \in \mathcal{T}} \mathbf{E}_\varepsilon^x \exp(-\sigma_n) \leqslant \sup_{x \in \bar{\mathcal{L}}} \mathbf{E}_\varepsilon^x \exp(-\sigma_{n-1}) \leqslant (\sup_{x \in \bar{\mathcal{L}}} \mathbf{E}_\varepsilon^x \exp(-\sigma))^{n-1} \leqslant (1 - r\varepsilon^{\beta-\frac{\alpha}{2}})^{n-1}. \tag{6.4}$$

The next lemma, also proved in Section 7, allows us to estimate expressions of the type (6.1) over the random intervals $[0, \tau]$ and $[0, \sigma]$.

Lemma 6.5. *For each $f \in \mathcal{D}$, we have the following asymptotic estimates*

$$\sup_{x \in \mathcal{T}} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon)\, dt \right] \right| \to 0 \qquad \text{as } \varepsilon \to 0, \tag{6.5}$$

$$\sup_{x \in \mathcal{T}} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\tau^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\tau \mathcal{A}f(Y_t^\varepsilon)\, dt \right] \right| \to 0 \qquad \text{as } \varepsilon \to 0, \tag{6.6}$$

$$\sup_{x \in \bar{\mathcal{L}}} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon)\, dt \right] \right| = o(\varepsilon^{\beta - \frac{\alpha}{2}}) \quad \text{as } \varepsilon \to 0, \tag{6.7}$$

$$\mathbf{E}_\varepsilon^\nu \left[f(Y_\tau^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\tau \mathcal{A}f(Y_t^\varepsilon)\, dt \right] = o(\varepsilon^{\beta - \frac{\alpha}{2}}) \quad \text{as } \varepsilon \to 0. \tag{6.8}$$

Here ν is the invariant measure on $\mathcal{L}_{\mathcal{T}}$ given by Lemma 6.3

We prove Lemma 6.1 by splitting the time interval $[0, T]$ into subsequent upcrossing and downcrossing periods. The first downcrossing from the general starting point is special and the contribution to (6.1) is estimated using (6.5). The estimate (6.4) gives us sufficient control on the growth rate of the number of upcrossing-downcrossings, so that by the stronger estimates (6.7), (6.8), we can show that the contribution from these time intervals to (6.1) is also negligible. In order to be able to use (6.8), we will use Lemma 6.3 to argue that after many such crossings, the section of the process on the separatrix can be approximated by its stationary counterpart. Finally, estimate (6.6) is used to show that the error thus introduced is negligible for small ε.

Proof of Lemma 6.1. Let $f \in \mathcal{D}$, $T > 0$, and $\eta > 0$ be fixed. We would like to show that the absolute value of the left hand side of (6.1) is less than η for all sufficiently small positive ε.

First, we replace the time interval $[0, T]$ by a larger one, $[0, \widetilde{\sigma}]$, where $\widetilde{\sigma}$ is the first of the stopping times σ_n that is greater than or equal to T, that is

$$\widetilde{\sigma} = \sigma_{N+1}, \qquad N = \max\{n : \sigma_n < T\}.$$

Using the Markov property of the process, the difference can be rewritten as

$$\left| \mathbf{E}_\varepsilon^x \left[f(Y_{\widetilde{\sigma}}^\varepsilon) - f(Y_0^\varepsilon) - \int_0^{\widetilde{\sigma}} \mathcal{A}f(Y_t^\varepsilon) dt \right] - \mathbf{E}_\varepsilon^x \left[f(Y_T^\varepsilon) - f(Y_0^\varepsilon) - \int_0^T \mathcal{A}f(Y_t^\varepsilon) dt \right] \right|$$

$$= \left| \mathbf{E}_\varepsilon^x \mathbf{E}_\varepsilon^{Z_T} \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon) dt \right] \right|.$$

Using (6.5) we can ensure that the right hand side of the above is smaller than $\frac{\eta}{5}$ for all sufficiently small ε. Therefore, it remains to show that

$$\left| \mathbf{E}_\varepsilon^x \left[f(Y_{\widetilde{\sigma}}^\varepsilon) - f(Y_0^\varepsilon) - \int_0^{\widetilde{\sigma}} \mathcal{A}f(Y_t^\varepsilon) dt \right] \right| < \frac{4\eta}{5}$$

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

$$\mathbf{E}_\varepsilon^x \left[f(Y_{\widetilde{\sigma}}^\varepsilon) - f(Y_0^\varepsilon) - \int_0^{\widetilde{\sigma}} \mathcal{A}f(Y_t^\varepsilon) dt \right] = \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon) dt \right]$$

$$+ \mathbf{E}_\varepsilon^x \left(\sum_{n=1}^N \mathbf{E}_\varepsilon^{Z_{\sigma_n}} \left[f(Y_\tau^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\tau \mathcal{A}f(Y_t^\varepsilon) dt \right] \right) \tag{6.9}$$

$$+ \mathbf{E}_\varepsilon^x \left(\sum_{n=1}^N \mathbf{E}_\varepsilon^{Z_{\tau_n}} \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon) dt \right] \right), \tag{6.10}$$

provided that the sums in the right hand side converge absolutely (which follows from the arguments below). Due to (6.5), the absolute value of the first term on the right hand side of this equality can be made smaller than $\frac{\eta}{5}$ for all sufficiently small ε. Therefore, it remains to estimate the two sums.

Let us start with the second sum (6.10). Note that

$$\mathbf{P}_\varepsilon^x(\sigma_n < T) = \mathbf{P}_\varepsilon^x(e^{-\sigma_n} > e^{-T}) \leqslant e^T(1 - r\varepsilon^{\beta - \frac{\alpha}{2}})^{n-1},$$

where the last inequality follows from (6.4) and Chebyshev's inequality. Taking the sum in n, we obtain

$$\mathbf{E}_\varepsilon^x N \leqslant \sum_{n=1}^\infty e^T \left(1 - r\varepsilon^{\beta - \frac{\alpha}{2}}\right)^{n-1} \leqslant \frac{K}{\varepsilon^{\beta - \frac{\alpha}{2}}},$$

where the constant K depends on T and r. By Lemma 6.5, we can find $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ we have

$$\sup_{x \in \mathcal{L}} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A}f(Y_t^\varepsilon)dt \right] \right| \leqslant \frac{\eta\varepsilon^{\beta - \frac{\alpha}{2}}}{5K}.$$

Multiplying these two bounds it follows that, for $\varepsilon < \varepsilon_0$, the term (6.10) is bounded by $\eta/5$.

Next, to estimate the term (6.9), we first note that (6.8) and the above argument shows

$$\left| \mathbf{E}_\varepsilon^x \sum_{n=1}^N \mathbf{E}_\varepsilon^\nu \left[f(Y_\tau^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\tau \mathcal{A}f(Y_t^\varepsilon)dt \right] \right| \leqslant \frac{\eta}{5} . \tag{6.11}$$

The left hand side of this inequality, however, is not quite the term (6.9), since the inner expectation is with respect to the invariant measure ν rather than individual points. (This limitation is due to (6.8).) Thus, in view of (6.11), to estimate (6.9) we only need to bound

$$\left| \mathbf{E}_\varepsilon^x \sum_{n=1}^N (F(Z_{\sigma_n}) - \overline{F}) \right|,$$

where

$$F(x) \stackrel{\text{def}}{=} \mathbf{E}_\varepsilon^x \left[f(Y_\tau^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\tau \mathcal{A}f(Y_t^\varepsilon)\, dt \right] \qquad \text{and} \qquad \overline{F} \stackrel{\text{def}}{=} \int_{\mathcal{L}_\mathcal{T}} F\, d\nu .$$

Observe

$$\left| \mathbf{E}_\varepsilon^x \sum_{n=1}^N (F(Z_{\sigma_n}) - \overline{F}) \right| \leqslant \left| \sum_{n=1}^\infty \mathbf{E}_\varepsilon^x (F(Z_{\sigma_n}) - \overline{F}) \right| + \left| \mathbf{E}_\varepsilon^x \sum_{n > N} (F(Z_{\sigma_n}) - \overline{F}) \right|$$

$$= \left| \sum_{n=1}^\infty \mathbf{E}_\varepsilon^x (F(Z_{\sigma_n}) - \overline{F}) \right| + \left| \mathbf{E}_\varepsilon^x \mathbf{E}_\varepsilon^{Z(\tau_N)} \sum_{n=1}^\infty (F(Z_{\sigma_n}) - \overline{F}) \right|$$

$$\leqslant 2 \sup_{x \in \mathcal{L}_\mathcal{T}} |F(x)| \sum_{n=1}^\infty \sup_{x \in \mathcal{L}_\mathcal{T}} \|P_1^{n-1}(x, \cdot) - \nu\|_{\text{TV}},$$

which is smaller than $\frac{\eta}{5}$ for all sufficiently small ε due to (6.3) and (6.6). Consequently the term (6.9) is bounded by $2\eta/5$ when ε is sufficiently small.

Combining the above estimates, we see that the absolute value of the left hand side of (6.1) is less than η for all sufficiently small positive ε. This completes the proof of Lemma 6.1. $\qquad\square$

7   Proofs of lemmas used in Section 6

In this section we prove Lemmas 6.2, 6.4, and 6.5. We start with estimates on the transition times and transition probabilities between different level sets of H. Recall that $\mathcal{L}(h) = \{x \in \mathcal{T} : |H(x)| = h\}$ and define $\mathcal{L}_i(h) = \mathcal{L}(h) \cap \overline{U}_i$. For $h \geqslant 0$, let

$$\bar{\tau}_h = \bar{\tau}(h) \overset{\text{def}}{=} \inf\{t \geqslant 0 : Z_t \in \mathcal{L}(h)\} \,,$$

so in particular, $\bar{\tau}_0 = \sigma$ and $\bar{\tau}(\varepsilon^\beta) = \tau$. For $0 \leqslant h_1 \leqslant h_2$, let $U(h_1, h_2) = \{x \in \mathcal{T} : h_1 \leqslant |H(x)| \leqslant h_2\}$ and $U_i(h_1, h_2) = U(h_1, h_2) \cap \overline{U}_i$.

In what follows, we take a more detailed look at the behaviour of Z_t near the separatrix. Let $z_t = z_t^\varepsilon(x)$ be the deterministic process

$$dz_t = \alpha \varepsilon^{\alpha-1} |\log \varepsilon|\, v(z_t)\, dt, \quad z_0 = x.$$

This is the same as the process Z_t under $\mathbf{P}_\varepsilon^x$, but with the stochastic term removed. Let $T^\varepsilon = T^\varepsilon(x)$ be the time it takes for the process z_t^ε, starting at x, to make one rotation along the level set, i.e., $T^\varepsilon(x) = \inf\{t > 0 : z_t = x\}$.

Lemma 7.1. *Suppose that $\lambda_1(\varepsilon), \lambda_2(\varepsilon)$ are such that $\varepsilon^{\lambda_1(\varepsilon)}, \varepsilon^{\lambda_2(\varepsilon)} = \mathcal{O}(\varepsilon^{\alpha/2})$ and $\lambda_1(\varepsilon) \leqslant \lambda_2(\varepsilon) < 1/2 - c$ for some $c > 0$.*

(a) There are positive constants c_1 and c_2 such that $c_1 \varepsilon^{1-\alpha} \leqslant T^\varepsilon(x) \leqslant c_2 \varepsilon^{1-\alpha}$ for all sufficiently small ε and all $x \in U(\varepsilon^{\lambda_2}, \varepsilon^{\lambda_1})$. Moreover, there are constants $\bar{c}_i > 0$ such that if $\alpha' > 0$ and $\lambda_1'(\varepsilon), \lambda_2'(\varepsilon) \to \alpha'/2$ as $\varepsilon \to 0$, then $\varepsilon^{-(1-\alpha)} T^\varepsilon(x) \to \bar{c}_i \alpha'/\alpha$ as $\varepsilon \to 0$ uniformly in $x \in U_i(\varepsilon^{\lambda_2'}, \varepsilon^{\lambda_1'})$.

(b) For each $\delta > 0$, $R > 0$, and all sufficiently small ε we have

$$\mathbf{P}_\varepsilon^x \Big(\sup_{t \leqslant T(x)} |H(Z_t) - H(x)| > \varepsilon^{\frac{1}{2} - \delta} \Big) < \varepsilon^R \ \ \text{for all} \ \ x \in U(\varepsilon^{\lambda_2(\varepsilon)}, \varepsilon^{\lambda_1(\varepsilon)}). \tag{7.1}$$

(c) For each $\delta > 0$, $R > 0$, and all sufficiently small ε we have

$$\mathbf{P}_\varepsilon^x \Big(\sup_{t \leqslant T(x)} |Z_t - z_t^\varepsilon(x)| > \varepsilon^{\frac{1}{2} - \lambda_2(\varepsilon) - \delta} \Big) < \varepsilon^R \ \ \text{for all} \ \ x \in U(\varepsilon^{\lambda_2(\varepsilon)}, \varepsilon^{\lambda_1(\varepsilon)}).$$

Statement (a) of the above lemma follows from a direct computation for the deterministic process, and the fact that

$$\bar{c}_i = \lim_{\varepsilon \to 0} \frac{T_i(\varepsilon^{1/2})}{|\log \varepsilon|},$$

where the limit is the same arising in (3.3). Statement (b) of the above lemma is basically a large deviation estimate on probability of the stochastic process to move transversal to the level-sets of H. Statement (c) is a large deviation estimate plus the fact that purely deterministic flow may separate points by an amount $\varepsilon^{-\lambda_2 - \delta}$, if both points are outside the boundary layer $|H(x)| \leqslant \varepsilon^{\lambda_2}$. Heuristic explanation for this deterministic separation is as follows. The rotation time T is of the same order as the time needed to pass the neighbourhood of a saddle and the largest separation also occurs near saddles. Let us analyse the linearised system $\dot{x} = x$, $\dot{y} = -y$. For the linearised system the particle trajectories and separation

between particles behave as the x-component given by $x = x_0 e^t$. The time to pass the neighbourhood of a saddle if found from $x_0 e^t = \mathcal{O}(1)$. Thus the rotation time $T = -C \log x_0$, and separation after one rotation is $\mathcal{O}(e^T) = c/x_0 \leqslant \mathcal{O}(\varepsilon^{-\lambda_2})$. In order to claim the separation in the nonlinear system occurs at the same rate, one needs to use the normal forms argument (see e.g. [AMB11]), because eigenvalues of the linearised system are resonant. Therefore we only obtain an estimate $\varepsilon^{-\lambda_2 - \delta}$ with the linear approximation argument. The rigorous proof of Lemma 7.1 is identical to that of Lemma 3.3 in [DK08], and we do not repeat it here.

Lemma 7.2. *Suppose that $\lambda_1(\varepsilon) \leqslant \lambda_2(\varepsilon)$ are such that $\varepsilon^{\lambda_1(\varepsilon)}, \varepsilon^{\lambda_2(\varepsilon)} = \mathcal{O}(\varepsilon^{\alpha/2})$, $\lambda_1(\varepsilon), \lambda_2(\varepsilon) < 1/2 - \kappa$ for some $\kappa > 0$, and $(\lambda_2(\varepsilon) - \lambda_1(\varepsilon))|\log \varepsilon| \to \infty$ as $\varepsilon \to 0$. Then*

(a) When $\lambda_1(\varepsilon) - \lambda_2(\varepsilon) \to 0$ as $\varepsilon \to 0$, we have the following upper bound on the expected exit time from a channel.

$$\sup_{x \in U(0, \varepsilon^{\lambda_2(\varepsilon)})} \mathbf{E}_\varepsilon^x (\bar{\tau}(\varepsilon^{\lambda_1(\varepsilon)}) \wedge \bar{\tau}_0) = \mathcal{O}(\varepsilon^{\lambda_2(\varepsilon) + \lambda_1(\varepsilon) - \alpha}) \text{ as } \varepsilon \to 0.$$

(b) If $\lambda_2(\varepsilon) \leqslant 2\lambda_1(\varepsilon) - c$ for some $c > 0$, the asymptotic behaviour of the exit probabilities is given by

$$\mathbf{P}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1(\varepsilon)}) < \bar{\tau}_0) \sim \varepsilon^{\lambda_2(\varepsilon) - \lambda_1(\varepsilon)} \text{ as } \varepsilon \to 0 \text{ uniformly in } x \in \mathcal{L}(\varepsilon^{\lambda_2(\varepsilon)}).$$

(c) The asymptotic expected exit time from a two sided channel satisfies

$$\sup_{x \in U(0, \varepsilon^{\lambda_1(\varepsilon)})} \mathbf{E}_\varepsilon^x \bar{\tau}(\varepsilon^{\lambda_1(\varepsilon)}) = \mathcal{O}(\varepsilon^{2\lambda_1(\varepsilon) - \alpha}) \text{ as } \varepsilon \to 0.$$

(d) There is a constant $c > 0$ such that

$$\sup_{x \in \mathcal{L}(\varepsilon^{\lambda_2(\varepsilon)})} \mathbf{E}_\varepsilon^x (\bar{\tau}(\varepsilon^{\lambda_1(\varepsilon)}) \wedge \bar{\tau}_0) > c\varepsilon^{\lambda_2(\varepsilon) + \lambda_1(\varepsilon) - \alpha}.$$

Proof. Without loss of generality, we may assume that the initial point x belongs to U_i, where U_i has the property that $H(x) \geqslant 0$ for all $x \in U_i$. Recall from Lemma 4.2 (after applying an appropriate time change) in [Kor04] that

$$\sup_{x \in U_i(0, h)} \mathbf{E}_\varepsilon^x \bar{\tau}(h) \leqslant c\varepsilon^{-\alpha} h^2 , \tag{7.2}$$

for some $c > 0$, which implies the third statement. Also, by [Kor04, Lem. 4.3],

$$\mathbf{P}_\varepsilon^x(\bar{\tau}(h) < \bar{\tau}_0) = \frac{H(x)}{h} + \mathcal{O}(h|\log \varepsilon|), \tag{7.3}$$

which implies the second statement as $\lambda_2(\varepsilon) < 2\lambda_1(\varepsilon) - c$. (Lemma 4.3 in [Kor04] can be improved to the extent where the assumption $\lambda_2(\varepsilon) < 2\lambda_1(\varepsilon) - c$ is not necessary, but we don't need it here.)

It remains to prove the first and fourth statements. From now on, we write λ_1 instead of $\lambda_1(\varepsilon)$, and similarly for λ_2. We also introduce $\lambda_3 = \frac{1}{2} - \kappa$ for a small number $\kappa > 0$ and set $u^\varepsilon(x) = \mathbf{E}_\varepsilon^x(\tau_{\text{ex}})$. Note that u^ε satisfies the boundary value problem

$$L^\varepsilon u^\varepsilon(x) = -1, \qquad x \in U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}),$$

$$u^\varepsilon(x) = 0, \qquad x \in \partial U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}),$$

where

$$L^\varepsilon = \frac{\alpha \varepsilon^\alpha |\log \varepsilon|}{2} \Delta + \frac{\alpha |\log \varepsilon|}{\varepsilon^{1-\alpha}} \nabla^\perp H \cdot \nabla$$

is the generator of the process Z_t under $\mathbf{P}^x_\varepsilon$. Let also $\tilde{u}^\varepsilon(x) = \varepsilon^{-\alpha}(H(x) - \varepsilon^{\lambda_3})(\varepsilon^{\lambda_1} - H(x))$. It is not hard to see that

$$\begin{aligned} L^\varepsilon \tilde{u}^\varepsilon(x) &= -\alpha |\log \varepsilon||\nabla H(x)|^2 + h^\varepsilon(x), & x \in U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}), \\ \tilde{u}^\varepsilon(x) &= 0, & x \in \partial U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}), \end{aligned}$$

where

$$h^\varepsilon(x) = \frac{\alpha |\log \varepsilon|}{2} \Delta H(x)(\varepsilon^{\lambda_1} + \varepsilon^{\lambda_3} - 2H(x)) = \mathcal{O}(\varepsilon^{\lambda_1} |\log \varepsilon|),$$

uniformly in $x \in U_i(0, \varepsilon^{\lambda_1})$.

Let $c > 0$ be a constant to be specified later and note that

$$\begin{aligned} L^\varepsilon(u^\varepsilon - c\tilde{u}^\varepsilon)(x) &= c\alpha |\log \varepsilon||\nabla H(x)|^2 - 1 - ch^\varepsilon(x), & x \in U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}), \\ u^\varepsilon(x) - c\tilde{u}^\varepsilon(x) &= 0, & x \in \partial U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1}). \end{aligned}$$

Writing $\tau_{\mathrm{ex}} = \bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}(\varepsilon^{\lambda_3})$ for the first exit time from the region $U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1})$, it follows from the Feynman-Kac formula that

$$u^\varepsilon(x) - c\tilde{u}^\varepsilon(x) = \mathbf{E}^x_\varepsilon \int_0^{\tau_{\mathrm{ex}}} \left(1 - \alpha c|\log \varepsilon||\nabla H(Z_t)|^2 + ch^\varepsilon(Z_t)\right) dt.$$

We will show that for all sufficiently small $\kappa > 0$, by choosing $c > 0$ large (small), the right hand side can be made positive (negative) for all $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$ for small enough ε, which then implies that there is a constant $c > 0$ such that

$$\frac{1}{c}\varepsilon^{-\alpha}(H(x) - \varepsilon^{\lambda_3})(\varepsilon^{\lambda_1} - H(x)) \leqslant \mathbf{E}^x_\varepsilon(\tau_{\mathrm{ex}}) \leqslant c\varepsilon^{-\alpha}(H(x) - \varepsilon^{\lambda_1})(\varepsilon^{\lambda_1} - H(x)), \tag{7.4}$$

for all $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$.

To do this, it clearly suffices to show that there are constants $0 < A < B$ such that, for small enough ε, the quantity

$$I(x, \tau_{\mathrm{ex}}, \varepsilon) := \mathbf{E}^x_\varepsilon \int_0^{\tau_{\mathrm{ex}}} |\log \varepsilon||\nabla H(Z_t)|^2 dt,$$

satisfies

$$\frac{I(x, \tau_{\mathrm{ex}}, \varepsilon)}{\mathbf{E}^x_\varepsilon(\tau_{\mathrm{ex}})} \in [A, B], \qquad \forall x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2}). \tag{7.5}$$

Let us rewrite $I(x, \tau_{\mathrm{ex}}, \varepsilon)$ by breaking the domain of integration into intervals corresponding to individual rotations of the unperturbed process z^ε_t. We inductively define the stopping times

$$\hat{T}^\varepsilon_0 = 0, \qquad \hat{T}^\varepsilon_{n+1} = \hat{T}^\varepsilon_n + T^\varepsilon(Z(\hat{T}^\varepsilon_n)), \qquad T^\varepsilon_n = \hat{T}^\varepsilon_n \wedge \bar{\tau}_0 \wedge \bar{\tau}(2\varepsilon^{\alpha/2}),$$

and note that by part (b) of Lemma 7.1, we have that for every R and small enough ε,

$$\mathbf{P}^x_\varepsilon(T^\varepsilon_1 \neq T^\varepsilon(x)) \leqslant \mathbf{P}^x_\varepsilon\left(\sup_{0 \leqslant t \leqslant T^\varepsilon(x)} |H(Z_t) - H(x)| > \varepsilon^{\lambda_3}\right) \leqslant \varepsilon^R \tag{7.6}$$

for $x \in U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1})$. Setting

$$\bar{T}^\varepsilon = \min\{T_n^\varepsilon : T_n^\varepsilon \geqslant \tau_{\text{ex}}\} \,,$$

we replace the exit time in $I(x, \tau_{\text{ex}}, \varepsilon)$ by $\bar{T}^\varepsilon$ in the upper limit of the integration. We will show later that the error introduced in this way is of order $\mathcal{O}(\varepsilon^{1-\alpha})$. We have the identity

$$
\begin{aligned}
I(x, \bar{T}^\varepsilon, \varepsilon) &= \sum_{n=0}^\infty \mathbf{E}_\varepsilon^x \left[\mathbf{1}_{\{T_n^\varepsilon < \bar{T}^\varepsilon\}} \mathbf{E}_\varepsilon^{Z(T_n^\varepsilon)} \int_0^{T_1^\varepsilon} |\log \varepsilon| |\nabla H(Z_t)|^2 dt \right] \\
&= \sum_{n=0}^\infty \mathbf{E}_\varepsilon^x \left(\mathbf{1}_{\{T_n^\varepsilon < \bar{T}^\varepsilon\}} I(Z(T_n^\varepsilon), T_1^\varepsilon, \varepsilon) \right) .
\end{aligned}
\tag{7.7}
$$

By (7.6) and part (c) of Lemma 7.1, it is not hard to see that for any $y \in U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1})$,

$$I(y, T_1^\varepsilon, \varepsilon) = I(y, T^\varepsilon(y), \varepsilon) + o(\varepsilon^{1-\alpha}). \tag{7.8}$$

Since $y \in U_i$, one can write this as

$$I(y, T^\varepsilon(y), \varepsilon) = \frac{\varepsilon^{1-\alpha}}{\alpha} \oint_{\mathcal{L}_i(H(x))} |\nabla H| \, dl = \frac{\varepsilon^{1-\alpha}}{\alpha} \oint_{\partial U_i} |\nabla H| \, dl + o(\varepsilon^{1-\alpha}). \tag{7.9}$$

Putting together (7.7), (7.8), (7.9), we conclude that there exists a constant c such that

$$I(x, \bar{T}^\varepsilon, \varepsilon) = \mathbf{E}_\varepsilon^x \left(\sum_{n=0}^\infty \mathbf{1}_{\{T_n^\varepsilon < \bar{T}^\varepsilon\}} \left(c\varepsilon^{1-\alpha} + o(\varepsilon^{1-\alpha}) \right) \right) .$$

By part (a) of Lemma 7.1, there exist constants $0 < a < b$ such that, for ε small enough and on the event $\{T_n^\varepsilon < \bar{T}^\varepsilon\}$, one has $\varepsilon^{\alpha-1}(T_{n+1}^\varepsilon - T_n^\varepsilon) \in (a, b)$ almost surely. Therefore, there exists a closed interval $J \subset \mathbb{R}_+ \setminus \{0\}$, a sequence of random variables $c_n(\varepsilon) \in J$ and an element $c(\varepsilon) \in J$ such that

$$I(x, \bar{T}^\varepsilon, \varepsilon) = \mathbf{E}_\varepsilon^x \left(\sum_{n=0}^\infty \mathbf{1}_{\{T_n^\varepsilon < \bar{T}^\varepsilon\}} (c_n(\varepsilon) + o(\varepsilon^{1-\alpha})) [T_{n+1}^\varepsilon - T_n^\varepsilon] \right) = (c(\varepsilon) + o(1)) \mathbf{E}_\varepsilon^x \bar{T}^\varepsilon \,.$$

In order to obtain (7.5), we would like to replace $\bar{T}^\varepsilon$ by τ_{ex}. First note that

$$\mathbf{E}_\varepsilon^x(\bar{T}^\varepsilon - \tau_{\text{ex}}) = \mathcal{O}(\varepsilon^{1-\alpha}). \tag{7.10}$$

It is clear from the analysis in (7.7), (7.8), and (7.9) and part (b) of Lemma 7.1 that this is indeed a small error term compared to each of the stopping times under the expectation. Indeed, by the conditions on λ_1, λ_2, it follows that $\varepsilon^{\lambda_2 - \lambda_1} \to 0$ as $\varepsilon \to 0$. Pick a $\delta > 0$ such that $\varepsilon^{\lambda_2} + \varepsilon^{1/2-\delta} < \varepsilon^{\lambda_1}$ for sufficiently small ε. Also choose $0 < \bar{\kappa} < \kappa \wedge \delta$ and note that in order to get out of $U_i(\varepsilon^{\lambda_3}, \varepsilon^{\lambda_1})$ with starting point in $U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$ within $\varepsilon^{-\bar{\kappa}}$ rotations, $H(Z_t)$ needs to change by more than $\varepsilon^{1/2+\bar{\kappa}-(\kappa\wedge\delta)}$ during at least one of these rotations and thus, writing Rot for the number of rotations before the exit time, one obtains from part (b) of Lemma 7.1

$$\mathbf{P}_\varepsilon^x(\text{Rot} \leqslant \varepsilon^{-\bar{\kappa}}) < \varepsilon^{-\bar{\kappa}} \varepsilon^R$$

for every $R > 0$. This implies the primitive a priori lower bound on the expected exit time

$$\mathbf{E}_\varepsilon^x(\tau_{\text{ex}}) > c\varepsilon^{1-\alpha} \mathbf{E}_\varepsilon^x(\text{Rot}) > c\varepsilon^{1-\alpha} \varepsilon^{-\bar{\kappa}}(1 - \varepsilon^{R-\bar{\kappa}}),$$

which, when combining it with (7.10), shows that

$$\mathbf{E}_\varepsilon^x(\bar{T}^\varepsilon - \tau_{\text{ex}}) = o(1)\mathbf{E}_\varepsilon^x(\tau_{\text{ex}}).$$

Similarly, replacing $\bar{T}^\varepsilon$ by τ_{ex} in the integral produces an error term of order $\mathcal{O}(\varepsilon^{1-\alpha}|\log \varepsilon|)$, which is much smaller than $\mathbf{E}_\varepsilon^x(\tau_{\text{ex}})$ by exactly the same argument.

This way we have proved

$$I(x, \varepsilon) = c_2(\varepsilon)\mathbf{E}_\varepsilon^x(\tau_{\text{ex}}),$$

where $c_2(\varepsilon) \in [A_2, B_2]$ for some positive constants A_2, B_2 and all $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$, and therefore by (7.4),

$$\frac{1}{c}H(x)\varepsilon^{\lambda_1-\alpha} \leqslant \mathbf{E}_\varepsilon^x(\tau_{\text{ex}}) \leqslant cH(x)\varepsilon^{\lambda_1-\alpha} \tag{7.11}$$

for all $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$. This immediately implies the fourth statement of the lemma. Indeed, if $H(x) = \varepsilon^{\lambda_2}$, then

$$\mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0) > \mathbf{E}_\varepsilon^x(\tau_{\text{ex}}) > c^{-1}\varepsilon^{\lambda_1+\lambda_2-\alpha}.$$

Similarly, for $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2})$ we have the upper bound

$$\mathbf{E}_\varepsilon^x(\tau_{\text{ex}}) = \mathcal{O}(\varepsilon^{\lambda_1+\lambda_2-\alpha}).$$

To prove the first statement of the lemma, note that

$$\mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0) \leqslant \mathbf{E}_\varepsilon^x(\tau_{\text{ex}}) + \sup_{x \in \mathcal{L}_i(\varepsilon^{\lambda_3})} \mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0).$$

The first term is of the right order by (7.11). To treat the second one, let $\kappa_1 > \kappa$. By (7.2),

$$\sup_{x \in \mathcal{L}_i(\varepsilon^{\lambda_3})} \mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0) \leqslant c\varepsilon^{1-2\kappa_1-\alpha} + \sup_{x \in \mathcal{L}_i(\varepsilon^{\lambda_3})} \mathbf{P}_\varepsilon^x(\bar{\tau}(\varepsilon^{1/2-\kappa_1}) < \bar{\tau}_0) \sup_{x \in \mathcal{L}_i(\varepsilon^{1/2-\kappa_1})} \mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0).$$
$$\tag{7.12}$$

By (7.3), the probability appearing on the right hand side is less than $\varepsilon^{\kappa_1-\kappa} + c\varepsilon^{1/2-\kappa_1}|\log \varepsilon|$. By this and the third statement of the lemma, we have

$$\sup_{x \in \mathcal{L}_i(\varepsilon^{1/2-\kappa})} \mathbf{E}_\varepsilon^x(\bar{\tau}(\varepsilon^{\lambda_1}) \wedge \bar{\tau}_0) < c(\varepsilon^{1-2\kappa_1-\alpha} + \varepsilon^{2\lambda_1-\alpha+\kappa_1-\kappa} + \varepsilon^{\frac{1}{2}-\kappa_1+2\lambda_1-\alpha}|\log \varepsilon|).$$

It is easy to check by choosing κ and κ_1 small enough and recalling that in this case $\lambda_2(\varepsilon) - \lambda_1(\varepsilon)$ goes to zero as $\varepsilon \to 0$, all three of these terms are $\mathcal{O}(\varepsilon^{\lambda_1+\lambda_2-\alpha})$, which finishes the proof for $x \in U_i(2\varepsilon^{\lambda_3}, \varepsilon^{\lambda_2(\varepsilon)})$.

Finally if the starting point is somewhere in $U_i(0, 2\varepsilon^{\lambda_3})$, then we first wait until the process gets out of this set. By (7.2), the expected value of how long this takes is $\mathcal{O}(\varepsilon^{2\lambda_3-\alpha})$ which does not change the conclusion. $\qquad\square$

Remark 7.3. *It is possible to prove the first statement of the previous lemma without the assumption that* $\lambda_1(\varepsilon) - \lambda_2(\varepsilon) \to 0$ *as* $\varepsilon \to 0$ *by iterating* (7.12). *We don't, however, need this here.*

Before we proceed, we introduce a little more notation. For $x \in \mathbb{R}^2 \setminus \mathcal{L}$, write $U_{i(x)}$ for the U_i containing $\pi(x)$ and define $H_x^\varepsilon \colon \mathbb{R}^2 \to \mathbb{R}$ by

$$H_x^\varepsilon(y) = \begin{cases} \varepsilon^{-\frac{\alpha}{2}}|H(y)| & \text{if } \pi(y) \in U_{i(x)}, \\ -\varepsilon^{-\frac{\alpha}{2}}|H(y)| & \text{otherwise.} \end{cases}$$

The reason for introducing H_x^ε is that it provides us with a "signed" version of the distance $d_{\mathcal{G}}$ on $\mathcal{G}$, i.e., $|H_x^\varepsilon(y) - H_x^\varepsilon(x)| = d_{\mathcal{G}}(\Gamma_\varepsilon(x), \Gamma_\varepsilon(y))$.

Lemma 7.4. *Suppose that* $\beta_1 = \beta_1(\varepsilon)$ *and* $\beta_2 = \beta_2(\varepsilon)$ *are such that* $0 < \beta_2 < \beta_1 < 1/2$ *and* $\beta_1(\varepsilon), \beta_2(\varepsilon) \to \alpha/2$ *as* $\varepsilon \to 0$. *For each* $f \in \mathcal{D}$,

$$\sup_{x \in U_i(\varepsilon^{\beta_1}, \varepsilon^{\beta_2})} \left| \mathbf{E}_\varepsilon^x \left[f(Y_{T(x)}^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^{T(x)} \mathcal{A}f(Y_t^\varepsilon) dt \right] \right| = o(\varepsilon^{1-\alpha}) \quad as \ \varepsilon \to 0. \tag{7.13}$$

Moreover,

$$\sup_{x \in U_i(\varepsilon^{\beta_1}, \varepsilon^{\beta_2})} |\mathbf{E}_\varepsilon^x(H_x^\varepsilon(Z_{T(x)}) - H_x^\varepsilon(x))| = o(\varepsilon^{1-\alpha}) \quad as \ \varepsilon \to 0. \tag{7.14}$$

$$\sup_{x \in U_i(\varepsilon^{\beta_1}, \varepsilon^{\beta_2})} \left| \mathbf{E}_\varepsilon^x[(H_x^\varepsilon(Z_{T(x)}) - H_x^\varepsilon(x))^2 - T(x)a_i] \right| = o(\varepsilon^{1-\frac{\alpha}{2}}) \quad as \ \varepsilon \to 0. \tag{7.15}$$

Proof. Let $T'(x) = \min(T(x), \sigma)$. By parts (a) and (b) of Lemma 7.1, for every R,

$$\mathbf{P}_\varepsilon^x(T(x) \neq T'(x)) = o(\varepsilon^R) = o(\varepsilon^{1-\alpha}) \, ,$$
$$\mathbf{E}_\varepsilon^x[T(x) - T'(x)] = \mathbf{E}_\varepsilon^x \mathbf{1}_{\{\sigma < T(x)\}}[T(x) - \sigma] \leqslant T(x)\mathbf{P}_\varepsilon^x(\sigma < T(x)) = \mathcal{O}(\varepsilon^{1-\alpha})o(\varepsilon^R) = o(\varepsilon^{1-\alpha}) \, ,$$

uniformly in $x \in U_i(\varepsilon^{\beta_1}, \varepsilon^{\beta_2})$. Due to this and the boundedness of f and $\mathcal{A}f$, it is sufficient to prove (7.13) with $T(x)$ replaced by $T'(x)$. By Itô's formula,

$$\mathbf{E}_\varepsilon^x \left[f(Y_{T'(x)}^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^{T'(x)} \mathcal{A}f(Y_t^\varepsilon) dt \right]$$

$$= \frac{\alpha|\log\varepsilon|}{2} \mathbf{E}_\varepsilon^x \int_0^{T'(x)} \left(|\nabla H(Z_t)|^2 f''(Y_t^\varepsilon) + \varepsilon^{\alpha/2}\Delta H(Z_t)f'(Y_t^\varepsilon) \right) dt - \mathbf{E}_\varepsilon^x \int_0^{T'(x)} \mathcal{A}f(Y_t^\varepsilon) dt.$$

The contribution from the second term in the first integral on the right hand side can be ignored due to Part (a) of Lemma 7.1 and the presence of the factor $\varepsilon^{\alpha/2}$. To deal with the first term, we observe that

$$\frac{\alpha|\log\varepsilon|}{2} \mathbf{E}_\varepsilon^x \int_0^{T'(x)} (|\nabla H(Z_t)|^2 f''(Y_t^\varepsilon) - |\nabla H(z_t^\varepsilon(x))|^2 f''(\Gamma_\varepsilon(x))) \, dt$$

$$= \mathcal{O}(|\log\varepsilon|\varepsilon^{\frac{1}{2}-\beta_1-\delta}T(x)) = o(\varepsilon^{1-\alpha}) \, ,$$

as $\varepsilon \to 0$, provided that δ is sufficiently small. This is where we use that H is C^2. We also used the fact that f''' is bounded (since $\mathcal{A}f \in \Psi$) and parts (b) and (c) of Lemma 7.1 for the first equality. We also used part (a) of Lemma 7.1 for the second one. Thus the first term can be replaced by

$$\frac{\alpha|\log\varepsilon|}{2} f''(\Gamma_\varepsilon(x))\mathbf{E}_\varepsilon^x \int_0^{T'(x)} |\nabla H(z_t^\varepsilon(x))|^2 dt = \frac{\alpha|\log\varepsilon|}{2} f''(\Gamma_\varepsilon(x)) \int_0^{T(x)} |\nabla H(z_t^\varepsilon(x))|^2 dt + o(\varepsilon^{1-\alpha})$$

$$= \frac{\varepsilon^{1-\alpha}}{2} f''(\Gamma_\varepsilon(x)) \oint_{\mathcal{L}(|H(x)|)\cap U_i} |\nabla H| \, dl + o(\varepsilon^{1-\alpha}) = \frac{\varepsilon^{1-\alpha}}{2} f''(\Gamma_\varepsilon(x)) \oint_{\partial U_i} |\nabla H| \, dl + o(\varepsilon^{1-\alpha}),$$

as $\varepsilon \to 0$. The last term is treated similarly:

$$\mathbf{E}_\varepsilon^x \int_0^{T'(x)} \mathcal{A}f(Y_t^\varepsilon) dt = \frac{a_i}{2} T(x)f''(\Gamma_\varepsilon(x)) + o(\varepsilon^{1-\alpha}) = \frac{\varepsilon^{1-\alpha}}{2} f''(\Gamma_\varepsilon(x)) \oint_{\partial U_i} |\nabla H| \, dl + o(\varepsilon^{1-\alpha}),$$

as $\varepsilon \to 0$, where the last equality is due to the definition of a_i (which is well defined by part (a) of Lemma 7.1). Note that in the definition of a_i in (3.3), $T_i(x)$ is the period of the unperturbed x_t^x while here $T(x)$ is the period of the process z_t^ε. Collecting all the terms, we obtain

$$\mathbf{E}_\varepsilon^x \left[f(Y_{T'(x)}^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^{T'(x)} \mathcal{A}f(Y_t^\varepsilon) \, dt \right] = o(\varepsilon^{1-\alpha}), \text{ as } \varepsilon \to 0,$$

as required for the proof of (7.13).

Formulas (7.14) and (7.15) are proved similarly. For (7.14), we apply the same arguments to a smooth bounded function f such that $f(h) = h - \Gamma_\varepsilon(x)$ in a neighbourhood of $\Gamma_\varepsilon(x)$. (Here, we identify the half line in $\mathcal{G}$ containing $\Gamma_\varepsilon(x)$ with $\mathbb{R}^+$ with the origin at O.) For (7.15), we apply the above arguments to a smooth bounded function f such that $f(h) = (h - \Gamma_\varepsilon(x))^2$ in a neighbourhood of $\Gamma_\varepsilon(x)$. $\qquad\square$

The following version of Donsker's theorem will be useful. Suppose that $\{\mathcal{F}_n\}_{n\geqslant 0}$ is an increasing sequence of σ-algebras, that $\tilde{u}_n$ and $\tilde{\xi}_n$ are two families of real-valued random variables measurable with respect to $\mathcal{F}_n$ and $\mathcal{F}_{n+1}$ respectively, and let $\gamma > 0$ be a (small) parameter. Assume that the $\tilde{u}_n$ are positive and that there exists a deterministic $\bar{u} \in (0, \gamma)$ and such that $|\tilde{u}_n - \bar{u}| \leqslant \gamma\bar{u}$ almost surely for all n. The quantities $\tilde{u}_n$ play the role of time steps for a random walk with spatial steps $\tilde{\xi}_n$. Define the partial sums $S_n = \tilde{u}_0 + \cdots + \tilde{u}_{n-1}$. Given an interval $[a, b]$ with $0 \in (a, b)$, define $\widetilde{R}_0 = 0$ and $\widetilde{R}_{n+1} = \widetilde{R}_n + \tilde{\xi}_n \mathbf{1}_{\tilde{R}_n \in (a,b)}$, $n \geqslant 0$. Assume that

(a) For each n, $|\mathbf{E}(\tilde{\xi}_n | \mathcal{F}_n)| \leqslant \gamma\bar{u}$ almost surely.

(b) For each n, $|\mathbf{E}(\tilde{\xi}_n^2 | \mathcal{F}_n) - \bar{u}| \leqslant \gamma\bar{u}$ almost surely.

(c) For each $\lambda > 0$ and all n, $\mathbf{P}(|\tilde{\xi}_n| > \lambda | \mathcal{F}_n) \leqslant \gamma\bar{u}$, almost surely.

Let $n(t) = \max(n : S_n \leqslant t)$ and define the continuous time process R_t by

$$R_t = \widetilde{R}_{n(t)} + \frac{t - S_{n(t)}}{S_{n(t)+1} - S_{n(t)}} \tilde{\xi}_{n(t)}, \quad t \geqslant 0 \,.$$

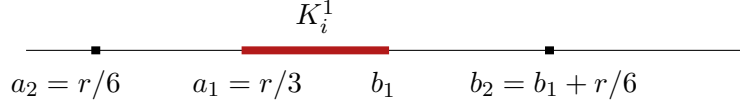We now claim the above process is close to stopped Brownian motion.

Lemma 7.5. *Let $\tilde{W}^{[a,b]}$ be a standard Brownian motion, stopped when it leaves $[a, b]$. For every $\delta > 0$ and every continuous $F \colon C(\mathbb{R}^+, \mathbb{R}) \to \mathbb{R}$ there exists γ_0 such that, for every $\gamma \leqslant \gamma_0$, one has*

$$|\mathbf{E}F(R) - \mathbf{E}F(\tilde{W}^{[a,b]})| \leqslant \delta \,.$$

Moreover, for arbitrary positive t_0, C, η, and r, there are $\delta \in (0, 1)$ and $\gamma_0 > 0$ such that, for every $\gamma \leqslant \gamma_0$ and every a, b with $-C \leqslant a < 0 < b \leqslant C$, one has

$$\sup_{t\in[0,t_0]} \mathbf{P}\left(\sup_{s\in[t,t+\delta]} |R_s - R_t| > r \right) \leqslant \delta\eta \,. \tag{7.16}$$

Proof. A standard proof of Donsker's theorem (see [EK86, Chap. 7.4] for the first statement and [Bil99, Chap. 8] for the second one) can be easily adapted to our situation. $\qquad\square$

$$K_i^1$$

$$a_2 = r/6 \qquad a_1 = r/3 \qquad b_1 \qquad b_2 = b_1 + r/6$$

Figure 5: Intervals $K_i^1 = [a_1, b_1]$ and $K_i^2 = [a_2, b_2]$.

Proof of Lemma 6.2. Recall that we identify one of the edges I_i of the graph $\mathcal{G}$ with the semi-axis $\mathbb{R}_+$. Define $K_i^1 = [a_1, b_1] \subset I_i$, where $a_1 = r/3$, $b_1 > a_1$. Let $K_i^2 = [a_2, b_2]$, where $a_2 = r/6$, $b_2 = b_1 + r/6$. For $x \in (\Gamma_\varepsilon)^{-1}(K_i^2)$, define inductively

$$y = \Gamma_\varepsilon(x), \qquad u_0^\varepsilon = T^\varepsilon(x), \qquad u_n^\varepsilon = T^\varepsilon(Z_{S_n^\varepsilon}), \qquad \xi_n^\varepsilon = H_x^\varepsilon(Z(S_{n+1}^\varepsilon)) - H_x^\varepsilon(Z(S_n^\varepsilon)).$$

Then Lemma 7.4 (relations (7.14) and (7.15)), parts (a) and (b) of Lemma 7.1, and the Markov property of the process imply that conditions (a)–(c) preceding Lemma 7.5 are met for

$$\widetilde{R}_n^{y,\varepsilon} = a_i^{-1/2}(H_x^\varepsilon(Z_{S_n^\varepsilon}) - H_x^\varepsilon(x)), \qquad \tilde{\xi}_n = a_i^{-1/2}\xi_n^\varepsilon, \qquad \tilde{u}_n = u_n^\varepsilon,$$

with some constant $\gamma = \gamma^\varepsilon$ converging to 0 as $\varepsilon \to 0$. Therefore, by Lemma 7.5 applied to the segment K_i^2 (suitably centred and rescaled) and part (b) of Lemma 7.1, for arbitrary positive η and r, there are $\delta > 0$ and $\varepsilon_0 > 0$ such that

$$\sup_{x \in \Gamma_\varepsilon^{-1}(K_i^1)} \mathbf{P}_\varepsilon^x\left(\sup_{s \in [0,\delta]} d_\mathcal{G}(Y_s^\varepsilon, \Gamma_\varepsilon(x)) > \frac{r}{3}\right) = \sup_{x \in \Gamma_\varepsilon^{-1}(K_i^1)} \mathbf{P}_\varepsilon^x\left(\sup_{s \in [0,\delta]} |H_x^\varepsilon(Z_s) - H_x^\varepsilon(x)| > \frac{r}{3}\right) \leqslant \delta\eta$$

for all $\varepsilon \leqslant \varepsilon_0$, where $d_\mathcal{G}$ stands for distance on $\mathcal{G}$. (There is no need to introduce the stopping time associated with exiting the segment K_i^2 since the starting point belongs to a smaller segment K_i^1.) Let $K_{r/3} \subset \mathcal{G}$ be the set of points whose distance from O does not exceed $r/3$. Using the strong Markov property for the process Z_t, i.e., stopping it when $\varepsilon^{-\alpha/2}|H(Z_t)| = r/3$, we obtain

$$\sup_{x \in \Gamma_\varepsilon^{-1}(K_{r/3})} \mathbf{P}_\varepsilon^x\left(\sup_{s \in [0,\delta]} d_\mathcal{G}(Y_s^\varepsilon, \Gamma_\varepsilon(x)) > r\right) \leqslant \sup_{x \in \Gamma_\varepsilon^{-1}(\bigcup_{i=1}^n K_i^1)} \mathbf{P}_\varepsilon^x\left(\sup_{s \in [0,\delta]} d_\mathcal{G}(Y_s^\varepsilon, \Gamma_\varepsilon(x)) > \frac{r}{3}\right) \leqslant \delta\eta.$$

For any compact set $K \subset \mathcal{G}$, we can ensure $K \subseteq \bigcup_{i=1}^n K_i^1 \cup K_{r/3}$ by choosing b_1 sufficiently large. Then

$$\sup_{x \in \Gamma_\varepsilon^{-1}(K)} \mathbf{P}_\varepsilon^x\left(\sup_{s \in [0,\delta]} d_\mathcal{G}(Y_s^\varepsilon, \Gamma_\varepsilon(x)) > r\right) \leqslant \delta\eta,$$

which implies the statement by the Markov property. $\qquad\square$

Proof of Lemma 6.4. For any $\beta < \alpha/2$ conditioning and the strong Markov property imply

$$\mathbf{E}_\varepsilon^x(e^{-\sigma}) \leqslant 1 - \mathbf{P}_\varepsilon^x(\bar{\tau}(\varepsilon^{\alpha/2}) < \bar{\tau}_0) + \mathbf{P}_\varepsilon^x(\bar{\tau}(\varepsilon^{\alpha/2}) < \bar{\tau}_0) \sup_{y \in \mathcal{L}(\varepsilon^{\alpha/2})} \mathbf{E}_\varepsilon^y(e^{-\sigma}), \tag{7.17}$$

for any $x \in \bar{\mathcal{L}} = \mathcal{L}(\varepsilon^\beta)$. It follows from Lemma 7.5 and Lemma 7.1(b) that there is a constant $r > 0$ independent of ε such that

$$\sup_{y \in \mathcal{L}(\varepsilon^{\alpha/2})} \mathbf{E}_\varepsilon^y(e^{-\sigma}) \leqslant \sup_{y \in \mathcal{L}(\varepsilon^{\alpha/2})} \mathbf{E}_\varepsilon^y \exp[-(\bar{\tau}(\varepsilon^{\alpha/2}/2) \wedge \bar{\tau}(2\varepsilon^{\alpha/2}))] \leqslant 1 - 2r,$$

for all sufficiently small ε. Using this in (7.17), we get

$$\mathbf{E}_\varepsilon^x(e^{-\sigma}) \leqslant 1 - 2r\mathbf{P}_\varepsilon^x(\bar\tau(\varepsilon^{\alpha/2}) < \bar\tau_0) \leqslant 1 - r\varepsilon^{\beta - \frac{\alpha}{2}},$$

for any $x \in \bar{\mathcal{L}}$. The second inequality above follows from the second statement of Lemma 7.2. $\qquad\square$

Proof of Lemma 6.5. First, let us show that (6.5) holds. As before, we define inductively

$$u_0^\varepsilon = T^\varepsilon(x), \qquad u_n^\varepsilon = T^\varepsilon(Z_{S_n^\varepsilon}), \qquad S_n^\varepsilon = u_0^\varepsilon + \cdots + u_{n-1}^\varepsilon.$$

Recall that $\bar\tau_h = \bar\tau(h)$ is the first time when Z_t reaches $\mathcal{L}(h)$. By the third statement of Lemma 7.2, there is a constant c such that

$$\sup_{x \in U(0, \varepsilon^{\alpha/2}r)} \mathbf{E}_\varepsilon^x \bar\tau(r\varepsilon^{\alpha/2}) \leqslant cr^2, \tag{7.18}$$

while by (7.3) we have

$$\sup_{x \in U(0, \varepsilon^{\alpha/2}r)} \mathbf{P}_\varepsilon^x\left(\bar\tau(2r\varepsilon^{\alpha/2}) < \bar\tau(\varepsilon^\beta)\right) \leqslant \sup_{x \in \mathcal{L}(r\varepsilon^{\alpha/2})} \mathbf{P}_\varepsilon^x\left(\bar\tau(2r\varepsilon^{\alpha/2}) < \bar\tau_0\right) \tag{7.19}$$

$$\leqslant \frac{1}{2} + \mathcal{O}(\varepsilon^{\alpha/2}|\log\varepsilon|) \leqslant \frac{2}{3}$$

for all sufficiently small ε. We claim that for each $r > 0$,

$$\sup_{x \in U(\varepsilon^\beta, r\varepsilon^{\alpha/2})} \left| \mathbf{E}_\varepsilon^x \left[f\left(Y_{\bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})}^\varepsilon\right) - f(\Gamma_\varepsilon(x)) - \int_0^{\bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})} \mathcal{A}f(Y_t^\varepsilon)\,dt \right] \right| \xrightarrow{\varepsilon \to 0} 0. \tag{7.20}$$

Indeed, let $\widetilde{n} = \min\{n : S_n^\varepsilon \geqslant \bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})\}$. Then, due to (7.1), it is sufficient to show that

$$\sup_{x \in U(\varepsilon^\beta, r\varepsilon^{\alpha/2})} \left| \mathbf{E}_\varepsilon^x \left[f(Y_{S_{\widetilde{n}}^\varepsilon}^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^{S_{\widetilde{n}}^\varepsilon} \mathcal{A}f(Y_t^\varepsilon)\,dt \right] \right| \xrightarrow{\varepsilon \to 0} 0. \tag{7.21}$$

By part (a) of Lemma 7.1 and (7.18), there is $\bar{c} > 0$ such that

$$\mathbf{E}_\varepsilon^x \widetilde{n} = \sum_{n=0}^\infty \mathbf{P}_\varepsilon^x\left(S_n^\varepsilon < \bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})\right) \leqslant \sum_{n=0}^\infty \mathbf{P}_\varepsilon^x\left(n \leqslant \frac{\bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})}{\bar{c}\varepsilon^{1-\alpha}}\right)$$

$$\leqslant \frac{\mathbf{E}_\varepsilon^x \bar\tau(\varepsilon^\beta) \wedge \bar\tau(r\varepsilon^{\alpha/2})}{\bar{c}\varepsilon^{1-\alpha}} \leqslant \frac{\mathbf{E}_\varepsilon^x \bar\tau(r\varepsilon^{\alpha/2})}{\bar{c}\varepsilon^{1-\alpha}} \leqslant cr^2\varepsilon^{\alpha-1}$$

for some constant c and all sufficiently small ε. Now the validity of (7.21) follows from (7.13). By Lemma 7.2 and (6.2),

$$\sup_{x \in U(0, \varepsilon^\beta)} \mathbf{E}_\varepsilon^x(\bar\tau(0) \wedge \bar\tau(r\varepsilon^{\alpha/2})) = \mathcal{O}(\varepsilon^{\beta - \frac{\alpha}{2}}),$$

and

$$\sup_{x \in U(0, \varepsilon^\beta)} \mathbf{P}_\varepsilon^x(\bar\tau(0) > \bar\tau(r\varepsilon^{\alpha/2})) \leqslant \sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{P}_\varepsilon^x(\bar\tau(0) > \bar\tau(r\varepsilon^{\alpha/2})) = \mathcal{O}(\varepsilon^{\beta - \frac{\alpha}{2}}), \tag{7.22}$$

as $\varepsilon \to 0$. Therefore, since f' and $\mathcal{A}f$ are bounded,

$$\sup_{x \in U(0, \varepsilon^\beta)} \left| \mathbf{E}_\varepsilon^x \left[f\left(Y_{\bar\tau(0) \wedge \bar\tau(r\varepsilon^{\alpha/2})}^\varepsilon\right) - f(\Gamma_\varepsilon(x)) - \int_0^{\bar\tau(0) \wedge \bar\tau(r\varepsilon^{\alpha/2})} \mathcal{A}f(Y_t^\varepsilon)\,dt \right] \right| = \mathcal{O}(\varepsilon^{\beta - \frac{\alpha}{2}}), \tag{7.23}$$

as $\varepsilon \to 0$.

Now take r sufficiently large so that $f(\Gamma_\varepsilon(x)) = 0$ whenever $x \notin U(0, r\varepsilon^{\alpha/2})$, which is possible since f has compact support. By the strong Markov property,

$$I_0 \overset{\text{def}}{=} \sup_{x \in U(\varepsilon^\beta, r\varepsilon^{\alpha/2})} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) \, dt \right] \right|$$

$$\leqslant \sup_{x \in U(\varepsilon^\beta, r\varepsilon^{\alpha/2})} \left| \mathbf{E}_\varepsilon^x \left[f\left(Y_{\bar\tau(\varepsilon^\beta) \wedge \bar\tau(2r\varepsilon^{\alpha/2})}^\varepsilon \right) - f(\Gamma_\varepsilon(x)) - \int_0^{\bar\tau(\varepsilon^\beta) \wedge \bar\tau(2r\varepsilon^{\alpha/2})} \mathcal{A} f(Y_t^\varepsilon) \, dt \right] \right|$$

$$+ \sup_{x \in U(\varepsilon^\beta, r\varepsilon^{\alpha/2})} \mathbf{P}_\varepsilon^x \left(\bar\tau(2r\varepsilon^{\alpha/2}) < \bar\tau(\varepsilon^\beta) \right) \sup_{x \in \mathcal{L}(2r\varepsilon^{\alpha/2})} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) \, dt \right] \right|$$

$$+ \sup_{x \in \mathcal{L}(\varepsilon^\beta)} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(\Gamma_\varepsilon(x)) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) \, dt \right] \right| \overset{\text{def}}{=} I_1 + I_2 + I_3.$$

By (7.19), $I_2 \leqslant \frac{2}{3} I_0$ for all sufficiently small ε. By (7.22) and (7.23), $I_3 \leqslant \frac{1}{6} I_0 + \mathcal{O}(\varepsilon^{\beta - \frac{\alpha}{2}})$ as $\varepsilon \to 0$. Therefore, $I_0 \leqslant 6 I_1 + \mathcal{O}(\varepsilon^{\beta - \frac{\alpha}{2}})$, which implies that $I_0 \to 0$ due to (7.20) used with $2r$ instead of r. Finally, (6.5) follows by combining this with (7.23) and using the strong Markov property.

The proof of (6.6) is nearly identical, and we, therefore, omit its proof.

Estimate (7.23) is not quite sufficient to obtain (6.7). Instead, we introduce $\beta'(\varepsilon)$ to be chosen later, such that $1/2 > \beta'(\varepsilon) > \beta(\varepsilon)$, and express the supremum in question as

$$J = \sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{E}_\varepsilon^x \left[f\left(Y_{\bar\tau(r\varepsilon^{\alpha/2}) \wedge \bar\tau(\varepsilon^{\beta'})}^\varepsilon \right) - f(\Gamma_\varepsilon(x)) - \int_0^{\bar\tau(r\varepsilon^{\alpha/2}) \wedge \bar\tau(\varepsilon^{\beta'})} \mathcal{A} f(Y_t^\varepsilon) \, dt \right]$$

$$+ \sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{E}_\varepsilon^x \left(\mathbf{1}_{\{\bar\tau(r\varepsilon^{\alpha/2}) < \bar\tau(\varepsilon^{\beta'})\}} \mathbf{E}_\varepsilon^{Z(\bar\tau(r\varepsilon^{\alpha/2}))} \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) dt \right] \right)$$

$$+ \sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{E}_\varepsilon^x \left(\mathbf{1}_{\{\bar\tau(r\varepsilon^{\alpha/2}) > \bar\tau(\varepsilon^{\beta'})\}} \mathbf{E}_\varepsilon^{Z(\bar\tau(\varepsilon^{\beta'}))} \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) \, dt \right] \right)$$

$$= J_1 + J_2 + J_3.$$

Note that (7.22) implies that

$$\sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{P}_\varepsilon^x(\bar\tau(r\varepsilon^{\alpha/2}) < \bar\tau(\varepsilon^{\beta'})) \leqslant \sup_{x \in \bar{\mathcal{L}}} \mathbf{P}_\varepsilon^x(\bar\tau(r\varepsilon^{\alpha/2}) < \bar\tau(0)) = \mathcal{O}(\varepsilon^{\beta - \alpha/2}).$$

Together with (6.5), this implies $J_2 = o(\varepsilon^{\beta - \alpha/2})$. On the other hand, we have as before that

$$\sup_{x \in \mathcal{L}(\varepsilon^{\beta'})} \mathbf{P}_\varepsilon^x(\bar\tau(0) > \bar\tau(\beta)) = \mathcal{O}(\varepsilon^{\beta' - \beta}),$$

as $\varepsilon \to 0$. Similarly to (7.23),

$$\sup_{x \in \mathcal{L}(\varepsilon^{\beta'})} \left| \mathbf{E}_\varepsilon^x \left[f\left(Y_{\bar\tau(0) \wedge \bar\tau(\beta)}^\varepsilon \right) - f(\Gamma_\varepsilon(x)) - \int_0^{\bar\tau(0) \wedge \bar\tau(\beta)} \mathcal{A} f(Y_t^\varepsilon) dt \right] \right| = \mathcal{O}(\varepsilon^{\beta' - \frac{\alpha}{2}}),$$

as $\varepsilon \to 0$, and therefore by

$$|J_3| \leqslant \sup_{x \in \mathcal{L}(\varepsilon^{\beta'})} \left| \mathbf{E}_\varepsilon^x \left[f(Y_\sigma^\varepsilon) - f(Y_0^\varepsilon) - \int_0^\sigma \mathcal{A} f(Y_t^\varepsilon) dt \right] \right| \leqslant \frac{|J|}{2} + \mathcal{O}(\varepsilon^{\beta' - \frac{\alpha}{2}}),$$

where we want that $\varepsilon^{\beta'-\beta} \to 0$, which can be achieved by choosing β' appropriately.

Consequently, $|J| \leqslant 2J_1 + o(\varepsilon^{\beta-\alpha/2})$ and it remains to show that $|J_1| = o(\varepsilon^{\beta-\alpha/2})$. This can be proved by first showing that

$$
J_1 = o\left(\sup_{x \in \mathcal{L}(\varepsilon^\beta)} \mathbf{E}_\varepsilon^x \bar{\tau}(r\varepsilon^{\alpha/2}) \wedge \bar{\tau}(\varepsilon^{\beta'}) \right) ,
$$

by breaking up the time interval into individual rotations, using (7.13) and then applying Lemma 7.2(a). Since this is only a slight modification of the machinery we used above, we omit the details.

Finally, the left hand side of (6.8) can be written using a Taylor-expansion as

$$
\varepsilon^{\beta-\alpha/2} \sum_{i=1}^{n} D_i f(0) \mathbf{P}_\varepsilon^\nu(Z_\tau \in U_i) + o(\varepsilon^{\beta-\alpha/2}) + \|\mathcal{A}f\|_\infty \mathbf{E}_\varepsilon^x \tau. \tag{7.24}
$$

Recall, that by (78) in [DK08], there exists a constant $c > 0$ such that

$$
\mathbf{P}_\varepsilon^\nu(Z_\tau \in U_i) = \mu(\bar{\mathcal{L}} \cap U_i) = c(1 + o(1)) \int_{\partial U_i} |\nabla H| \, dl \xrightarrow{\varepsilon \to 0} cq_i.
$$

Using this, the assumption $f \in \mathcal{D}$ and the third statement in Lemma 7.2, it follows that (7.24) is $o(\varepsilon^{\beta-\alpha/2})$, which proves (6.8). □

Appendix A   Proof of Proposition 5.1

In this appendix we sketch the proof of our limit theorem regarding the displacement of Z_t during an upcrossing. The verification of this result is based on an abstract lemma that was proved in [HKP14].

Before we state the lemma we introduce its setup. Let $\mathcal{M}$ be a metric space that can be written as a disjoint union

$$
\mathcal{M} = \mathcal{X} \sqcup C_1 \sqcup \ldots \sqcup C_n ,
$$

where the sets C_i are closed. Assume also that $\mathcal{X}$ is a σ-locally compact separable subspace, i.e., locally compact that is the union of countably many compact subspaces. Let $p_\varepsilon(x, dy)$, $0 \leqslant \varepsilon \leqslant \varepsilon^0$, be a family of transition probabilities on $\mathcal{M}$ and let $g \in \mathcal{C}_b(\mathcal{M}, \mathbb{R}^2)$. Later, $p_\varepsilon(x, dy)$ will come up as transition probabilities of a certain discrete time process associated to Z_t. We assume that the following properties hold:

1. $p_0(x, \mathcal{X}) = 1$ for all $x \in \mathcal{M}$ and $p_\varepsilon(x, \mathcal{X}) = 1$ for all $x \in \mathcal{M} \setminus \mathcal{X}$.

2. $p_0(x, dy)$ is weakly Feller, that is the map $x \mapsto \int_\mathcal{M} f(y) p_0(x, dy)$ belongs to $\mathcal{C}_b(\mathcal{M})$ if $f \in \mathcal{C}_b(\mathcal{M})$.

3. (Small escape probability from $\mathcal{X}$.) There exist bounded continuous functions $h_1, \ldots, h_n : \mathcal{X} \to \mathbb{R}_+$ such that
$$
\varepsilon^{-\frac{1-\alpha}{2}} p_\varepsilon(x, C_i) \to h_i(x) ,
$$
uniformly over $x \in K$ for $K \subseteq \mathcal{X}$ compact, while $\sup_{x \in \mathcal{X}} |\varepsilon^{-\frac{1-\alpha}{2}} p_\varepsilon(x, C_i)| \leqslant c$ for some positive constant c. We also have
$$
J(x) \stackrel{\text{def}}{=} h_1(x) + \cdots + h_n(x) > 0 \text{ for } x \in \mathcal{X}.
$$

4. $p_\varepsilon(x, dy)$ converges weakly to $p_0(x, dy)$ as $\varepsilon \to 0$, uniformly in $x \in K$ if $K \subseteq \mathcal{X}$ is compact.

5. (Doeblin condition) The transition functions satisfy a strong Doeblin condition uniformly in ε. Namely, there exist a probability measure η on $\mathcal{X}$, a constant $a > 0$, and an integer $m > 0$ such that

$$p_\varepsilon^m(x, A) \geqslant a\, \eta(A) \quad \text{for } x \in \mathcal{M}, \ A \in \mathcal{B}(\mathcal{X}), \ \varepsilon \in [0, \varepsilon_0].$$

It then follows that for every ε, there is a unique invariant measure $\lambda^\varepsilon(dy)$ on $\mathcal{M}$ for $p_\varepsilon(x, dy)$, and the associated Markov chain is uniformly exponentially mixing, i.e., there are $\Lambda > 0, c > 0$, such that

$$|p_\varepsilon^k(x, A) - \lambda^\varepsilon(A)| \leqslant c\, e^{-\Lambda k} \quad \text{for all } x \in \mathcal{M}, \ A \in \mathcal{B}(\mathcal{M}), \ \varepsilon \in [0, \varepsilon_0].$$

6. The function g is such that $\int_\mathcal{M} g\, d\lambda^\varepsilon = 0$ for each $\varepsilon \in [0, \varepsilon_0]$.

Lemma A.1. *(Lemma 2.4 in [HKP14]) Suppose that Properties 1–6 above are satisfied and let $R_k^{x,\varepsilon}$ be the Markov chain on $\mathcal{M}$ starting at x, with transition function p_ε. Let $\tau = \tau(x, \varepsilon)$ be the first time when the chain reaches the set $C = C_1 \sqcup \ldots \sqcup C_n$. Let $e(R_k^{x,\varepsilon}) = i$ if $R_k^{x,\varepsilon} \in C_i$. Then, as $\varepsilon \to 0$,*

$$\left(\varepsilon^{\frac{1-\alpha}{4}} (g(R_1^{x,\varepsilon}) + \cdots + g(R_\tau^{x,\varepsilon})), e(R_\tau^{x,\varepsilon}) \right) \to (F_1, F_2)$$

in distribution, uniformly in $x \in \mathcal{X}$, where F_1 takes values in $\mathbb{R}^2$, F_2 takes values in $\{1, \ldots, n\}$, and F_1 and F_2 are independent. The random variable F_1 is distributed as $(\xi / \int_\mathcal{X} J d\lambda^0)^{\frac{1}{2}} N(0, \bar{Q})$, where ξ is exponential with parameter one independent of $N(0, \bar{Q})$ and $\bar{Q}$ is the matrix such that

$$(g(R_1^{x,0}) + \cdots + g(R_k^{x,0}))/\sqrt{k} \to N(0, \bar{Q})$$

in distribution as $k \to \infty$. The random variable F_2 satisfies

$$p_i = \mathbf{P}(F_2 = i) = \frac{\int_\mathcal{X} h_i\, d\lambda^0}{\int_\mathcal{X} J\, d\lambda^0}, \ i = 1, \ldots, n. \tag{A.1}$$

Let us now show that Lemma A.1 is applicable to a certain Markov chain associated with Z_t. Our objective is to define this Markov chain and verify that it satisfies Properties (1)-(6). We start by explaining what it means for a process Z_t to pass a saddle point. Consider "the projection on the separatrix" mapping $\rho : \mathcal{V}^{\delta,\varepsilon} \to \mathcal{L}$ with

$$\mathcal{V}^{\delta,\varepsilon} = \{x \in \mathbb{R}^2 : |H(x)| \leqslant \delta\varepsilon^{\alpha/2}\},$$

and given by

$$\rho(x) = \mathcal{L} \cap \overline{\{\Phi_t^x : t \in \mathbb{R}\}}, \qquad \dot{\Phi}_t^x = \nabla H(\Phi_t^x), \quad \Phi_0^x = x.$$

For sufficiently small δ and ε this ρ is uniquely defined because the closure of the orbit of the gradient flow Φ_t^x does indeed intersect the separatrix $\mathcal{L}$ at exactly one point. We will say that Z passes a saddle point A_i if its trajectory intersects the curves

$$B(A_i) = \{x \in \mathcal{V}^{\delta,\varepsilon}, \ \pi(\rho(x)) = A_i\},$$

where $\pi : \mathbb{R}^2 \to \mathcal{T}$ is the quotient map from the plane to the torus. Set $\gamma_0^\varepsilon = \beta_0^\varepsilon = 0$, and then recursively

$$\gamma_n^\varepsilon = \inf\left\{ t \geqslant \beta_{n-1}^\varepsilon : Z_t \in \left(\bigcup_{k \neq i} B(A_k) \right) \bigcup \partial\mathcal{V}^{\delta,\varepsilon} \right\}, \quad \beta_n^\varepsilon = \inf\{t \geqslant \gamma_n^\varepsilon : Z_t \in \mathcal{L}\}, \tag{A.2}$$

provided that $\pi(Z(\beta_{n-1}^\varepsilon)) \in \{A_i\} \cup \bigcup_j E(A_i \to A_j)$. Here, $E(A_i \to A_j)$ denotes the heteroclinic connection emanating from the saddle A_i and ending at A_j, or the empty set if no such connection exists. It means the stopping time γ_n^ε clocks the first time after β_{n-1}^ε that the process either hits $\partial\mathcal{V}^{\delta,\varepsilon}$, or goes past a saddle point different from the one behind $Z(\beta_{n-1}^\varepsilon)$. Recall that we assumed that there are no homoclinic orbits, and therefore the definition of γ_n^ε makes sense.

Define an auxiliary metric space $\bar{\mathcal{M}} = \mathcal{L} \sqcup \partial\mathcal{V}^{\delta,\varepsilon}$. Let us define a family of transition functions $\bar{p}_\varepsilon(x, dy)$ on $\bar{\mathcal{M}}$. For $x \in \mathcal{L}$, we define $\bar{p}_\varepsilon(x, dy)$ as the distribution of $Z_{\bar\tau}$ (under $\mathbf{P}_\varepsilon^x$) with the random transition time $\bar\tau = \mu_1^\varepsilon \wedge \beta_1^\varepsilon$, where β_1^ε and μ_1^ε are defined in (A.2) and (5.1), respectively. In other words, it is the measure induced by the process stopped when it either reaches the boundary of $\mathcal{V}^{\delta,\varepsilon}$ or reaches the separatrix after passing by a saddle point. For $x \in \partial\mathcal{V}^{\delta,\varepsilon}$, we define $\bar{p}_\varepsilon(x, dy)$ as the distribution of $Z_{\bar\tau}$ (under $\mathbf{P}_\varepsilon^x$) with $\bar\tau = \beta_1^\varepsilon$, i.e., the measure induced by the process stopped when it first reaches the separatrix. We write $\bar{R}_k^{x,\varepsilon}$ for the corresponding Markov chain starting at $x \in \bar{\mathcal{M}}$.

Note that $\bar{\mathcal{M}}$ depends on ε since it contains $\partial\mathcal{V}^{\delta,\varepsilon}$, and we would like to get rid of this dependence in order to use the abstract setup of Lemma A.1. The projection on the separatrix mapping ρ (when lifted to the torus $\mathcal{T}$) defines a natural homeomorphism between $\pi(\partial\mathcal{V}^{\delta,\varepsilon}) \cap U_i$ and a circle $S_i \subset \pi(\mathcal{L})$ with circumference $\int_{\partial U_i} |\nabla H|\, dl$. We will assume that the circles S_i and S_j are disjoint for $i \neq j$, and denote this homeomorphism by ρ^ε.

While we introduced $\bar{\mathcal{M}}$ as a subset of $\mathbb{R}^2$, it is going to be more convenient to keep track of $\pi(\bar{R}_k^{x,\varepsilon})$ and the latest displacement separately. Let N be a bounded measurable set in $\mathbb{R}^2$ such that $\mathbb{R}^2 = \bigsqcup_{z \in \mathbb{Z}^2}(N + z)$. For $x \in \mathbb{R}^2$, let $[x] \in \mathbb{Z}^2$ be such that $x \in N + [x]$. We can choose N in such a way that $[x]$ is constant on each connected component of $\pi^{-1}(U_i)$, i.e., N consists of a finite number of cells and parts of their boundaries. It is then natural to define a metric space $\mathcal{M}$ independent of ε by
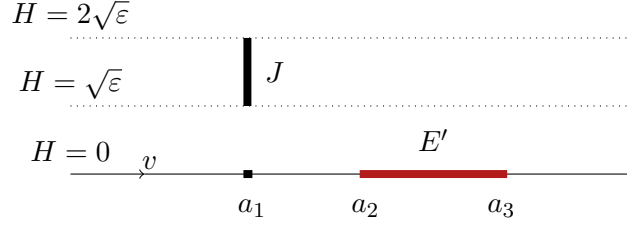
$$\mathcal{M} = (\pi(\mathcal{L}) \sqcup S_1 \sqcup \cdots \sqcup S_n) \times \mathbb{Z}^2 =: \mathcal{M}_1 \times \mathbb{Z}^2\,,$$

which is indeed of the above type by setting $\mathcal{X} = \pi(\mathcal{L}) \times \mathbb{Z}^2$ and $C_i = S_i \times \mathbb{Z}^2$. Let $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ be given by $\varphi(x) = (\rho^\varepsilon(\pi(x)), [x])$ for $x \notin \mathcal{L}$, and $\varphi(x) = (\pi(x), [x])$ for $x \in \mathcal{L}$. We will write $\varphi_1 : \bar{\mathcal{M}} \to \mathcal{M}_1$ and $\varphi_2 : \bar{\mathcal{M}} \to \mathbb{Z}^2$ for the first and second components of φ, respectively. This allows us to define transition probabilities p_ε on $\mathcal{M}$ by setting $p_\varepsilon((x_0, k), \cdot)$ to be the law of $\varphi(Z_{\bar\tau})$ under $\mathbf{P}_\varepsilon^{\bar{x}_0}$ with $\bar{x}_0$ the only element of N with $\pi(\bar{x}_0) = x_0$. Similarly to before, we write $R_k^{x,\varepsilon}$ for the Markov chain starting at $x \in \mathcal{M}$ with transition probabilities p_ε.

We finally define the function g appearing in Property (6) by, for $x = (q, \xi) \in \mathcal{M}$, setting $g((q, \xi)) = \xi \in \mathbb{Z}^2$. This continuous function measures the displacement during the last step if the chain is viewed as a process on $\mathbb{R}^2$, where only the integer parts of the initial and end points are counted.

The Markov chain is now defined, and it remains to verify that Properties (1)-(6) are satisfied. Here we will adopt the approach [Kor04]. The thrust of [Kor04] is the asymptotic analysis of the behaviour of the process Z_t in an ε-neighbourhood of the separatrix. A Markov chain on separatrices, similar to our $R_k^{x,\varepsilon}$, also arises there. In [Kor04] the analysis, however, was done in a scaling slightly different from ours. More specifically, in [Kor04] the width of the separatrix region is restricted to be of order ε^{α_1} with some $\alpha_1 \in (1/4, 1/2)$, while our width is $\delta\varepsilon^{\alpha/2}$, $\alpha > 0$. It is fairly straightforward to verify that ε^{α_1} may be replaced by $\delta\varepsilon^{\alpha/2}$, i.e., the imposed restriction $\alpha > 1/2$ is not needed and can be replaced by $\alpha > 0$. We will, therefore, simply quote here the corresponding statements from [Kor04] and explain their modifications necessary to imply our Properties (1)-(6).

Properties (1), (2) and (4). The existence of the limit of the transition functions p_ε in the sense of Property (4) was justified in [Kor04, Lem. 3.1]. This limit is denoted by p_0. An explicit formula for the density of p_0 was also provided in [Kor04, Eq. (9)], which implies that Property (2) is satisfied. Observe

Figure 6: Construction of J.

that the probability of β_1^ε being less than μ_1^ε tends to one as $\varepsilon \to 0$, uniformly in $x \in \mathcal{L}$ by [Kor04, Eq. (26)]. This implies Property (1).

Let us sketch the proof of the Doeblin condition (Property (5)). It suffices to show that there exists $E' \subset \mathcal{L}_0$, a connected segment of the same heteroclinic orbit of H, and there are a constant $c > 0$ and an integer $m > 0$ such that

$$\bar{p}_\varepsilon^m(x, E') \geqslant c \, |E'| \quad \text{for all } x \in \mathcal{T}_0 \cap \bar{\mathcal{M}}, \ \varepsilon \in [0, \varepsilon_0], \tag{A.3}$$

where $|E'|$ is the arclength of E'. Since $\bar{p}_\varepsilon(x, \mathcal{T}_0 \cap \mathcal{L}) = 1$, if $x \in \mathcal{T}_0 \cap \partial \mathcal{V}^{\delta, \varepsilon}$, it suffices to show the last estimate immediately holds for all $x \in \mathcal{T}_0 \cap \mathcal{L}$. For $x \in \mathcal{T}_0 \cap \mathcal{L}$ we can obtain (A.3) once we show that there is a set $J \subset \mathbb{R}^2$, that may depend on ε, such that it has the following two properties. Firstly, there is $m > 0$ such that

$$\mathbf{P}_\varepsilon^x \left(Z_t \in J \ \text{ for some } \gamma_m^\varepsilon < t < \beta_m^\varepsilon \right) > c > 0, \forall x \in \mathcal{T}_0 \cap \mathcal{L} \ .$$

Secondly,

$$\mathbf{P}_\varepsilon^x \left(Z_t \in E' \ \text{ for some } t > 0 \right) > c > 0, \forall x \in J.$$

We construct J as follows. Suppose a_2 and a_3 are the endpoints of E', and a_1 lies on the same heteroclinic orbit of H so that the points are ordered in the direction of the flow v, as depicted on Figure 6. Let J be a piece of the curve in $\mathcal{V}^{\delta, \varepsilon}$, that is mapped by ρ to a_1: $J = \{x \in \mathcal{V}^{\delta, \varepsilon} \bigcap \pi^{-1}(U_k) : \sqrt{\varepsilon} \leqslant |H(x)| \leqslant 2\sqrt{\varepsilon}, \rho(x) = a_1\}$ for some k.

Roughly speaking, the first property means that the process has a positive chance of going to a particular curve at a distance $\sqrt{\varepsilon}$ from the separatrix, transversal to the flow lines, prior to passing by m saddle points. This is not surprising since the ratio of the parallel advection to the diffusion is of order $1/\varepsilon$. The proof follows along the same lines as the proof of [Kor04, Lem. 3.1]. Similarly, the second property is true, because if the process Z starts on J, then it will take $\mathcal{O}(\varepsilon)$ time for the flow v to carry Z past the segment E', but this is sufficient for diffusion move the distance $\mathcal{O}(\sqrt{\varepsilon})$, and to reach the separatrix. The latter argument is the same as the derivation of (63) in [DK08]. Thus the Doeblin condition for $R_k^{x, \varepsilon}$ holds.

With our definition of g,

$$\int_{\mathcal{M}} g(x) \, d\lambda^\varepsilon(x) \left(\int_{\mathcal{M}} \mathbf{E}\bar{\tau} \, d\lambda^\varepsilon(x) \right)^{-1} = \lim_{t \to \infty} (\mathbf{E} Z_t / t),$$

where $\bar{\tau} = \mu_1^\varepsilon \wedge \beta_1^\varepsilon$ is the random transition time for our Markov chain, and the right hand side is the effective drift for the original process starting from an arbitrary point x. Note that for some $C(\varepsilon) > 0$,

$$\lim_{t \to \infty} \frac{\mathbf{E} Z_t}{t} = C(\varepsilon) \int_{\mathcal{T}} v(x) \, dx = 0,$$

which implies Property (6).

Small escape probability from $\mathcal{X}$ (Property (3)) follows from [Kor04, Lem. 4.1 & 4.3]. Indeed, the first lemma describes the asymptotics of the distribution of $H(Z_{\gamma_1^\varepsilon})$, while the second describes the probability of the process starting at x to exit the boundary layer before reaching the separatrix, assuming that $H(x)$ is fixed. The two lemmas, combined with the Markov property of the process, imply Property (3). Moreover, our functions $h_i(x) = h_i^\delta(x)$ depend on δ and can be identified as

$$h_i^\delta(x) = \lim_{\varepsilon \to 0} \varepsilon^{-\frac{1-\alpha}{2}} \mathbf{P}\Big(\text{ the process starting at } Z_{\gamma_1^\varepsilon} \text{ reaches } \partial \mathcal{V}^{\delta,\varepsilon} \cap \pi^{-1}(U_i) \text{ before reaching } \mathcal{L}\Big). \quad \text{(A.4)}$$

Proof of Proposition 5.1. From [Kor04, Lemma 4.1 and Lemma 4.3], it follows that $h_i^\delta(x)$ in (A.4) satisfy

$$\int_X h_i^\delta(x)\,d\lambda^0(x) = \delta^{-1}\bar{p}_i, \quad i = 1, \ldots, n,$$

for some $\bar{p}_i > 0$. Now Lemma A.1 implies that covariance matrix Q and probabilities p_i in (A.1) must satisfy

$$Q = \bar{Q}/(\bar{p}_1 + \cdots + \bar{p}_n), \qquad p_i = \bar{p}_i/(\bar{p}_1 + \cdots + \bar{p}_n).$$

The non-degeneracy of Q was demonstrated in [HKP14] in the case when $\alpha = 0$. The proof does not require modifications in the current case.

It is clear that Lemma A.1 with g and C_i introduced above implies the first statement of Proposition 5.1. The second statement of Proposition 5.1 requires a slight strengthening of the abstract Lemma A.1 without major modifications in the proof (see [HKP14]). $\qquad \square$

Appendix B A formal asymptotic expansion

We devote this appendix to a heuristic derivation of the system (4.2) through a formal asymptotic expansion. Let δ be a small parameter such that

$$\delta \xrightarrow{\varepsilon \to 0} 0, \qquad \text{and} \qquad \frac{\varepsilon}{\delta} \xrightarrow{\varepsilon \to 0} 0.$$

The typical situation arising throughout this paper is for $\delta = \varepsilon^{1-\alpha}$. This specific choice, however, is unnecessary for the formal asymptotics presented here.

Define $\varphi = \varphi^{\varepsilon,\delta}$ by

$$\varphi(z, t) = \tilde{\theta}\Big(z, \frac{t \log(\delta/\varepsilon)}{\delta}\Big)$$

and observe

$$\frac{1}{\log(\delta/\varepsilon)}\partial_t \varphi = \frac{1}{\delta}v \cdot \nabla\varphi + \frac{\varepsilon}{2\delta}\Delta\varphi.$$

For notational convenience, we suppress the ε and δ-dependence of φ and other quantities below. We perform an asymptotic expansion for φ by writing

$$\varphi(z, t) = \varphi_0(x, z, t) + \gamma\varphi_1(x, z, t) + \mathcal{O}(\gamma^2).$$

where z is the "fast variable", $\gamma = \delta^{1/4}$ and $x = \gamma z$. Using

$$\nabla = \gamma\nabla_x + \nabla_z \qquad \text{and} \qquad \Delta = \gamma^2\Delta_x + 2\gamma\nabla_x \cdot \nabla_z + \Delta_z,$$

we compute

$$\frac{1}{\log(\delta/\varepsilon)}\partial_t\varphi = \frac{\gamma}{\delta}v(z)\cdot\nabla_x\varphi_0 + \frac{1}{\delta}v(z)\cdot\nabla_z\varphi_0 + \frac{\varepsilon}{2\delta}\Delta\varphi_0$$

$$+ \gamma\Big(\frac{\gamma}{\delta}v(z)\cdot\nabla_x\varphi_1 + \frac{1}{\delta}v(z)\cdot\nabla_z\varphi_1 + \frac{\gamma^2\varepsilon}{2\delta}\Delta_x\varphi_1 + \frac{\gamma\varepsilon}{\delta}\nabla_x\cdot\nabla_z\varphi_1 + \frac{\varepsilon}{2\delta}\Delta_z\varphi_1\Big) + \mathcal{O}(\gamma^2). \quad \text{(B.1)}$$

Since there is only one term of order $\mathcal{O}(1/\delta)$, it must vanish. This gives

$$v(z)\cdot\nabla_z\varphi_0 = 0, \tag{B.2}$$

and hence, the dependence of φ_0 on z is only through $\Gamma_\varepsilon(z)$, the projection of z onto $\mathcal{G}$.[7] In view of our rescaling, we expect

$$\theta(x,y,t) = \varphi_0(x,z,t) + o(1), \qquad \text{where } y \in \mathcal{G} \text{ and } y = \Gamma_\varepsilon(z).$$

Consequently, in order to justify (4.2), we will formally obtain equations for the function θ defined by

$$\theta(x,y,t) \overset{\text{def}}{=} \lim_{\varepsilon\to 0}\varphi_0(x,z,t), \qquad \text{where } y \in \mathcal{G} \text{ and } y = \Gamma_\varepsilon(z).$$

We begin by balancing the $\mathcal{O}(\gamma/\delta)$ terms by choosing φ_1 to be the solution of

$$\frac{\varepsilon}{2}\Delta_z\varphi_1 + v(z)\cdot\nabla_z\varphi_1 = -v(z)\cdot\nabla_x\varphi_0, \tag{B.3}$$

with periodic boundary conditions in z. In order to do this we would need to verify the compatibility condition

$$\int_U v(z)\cdot\nabla_x\varphi_0\,dz = 0.$$

Indeed, let U be any connected component of the support of φ_0. Then, for fixed x and t, equation (B.2) implies that for all $z \in U$, we have $\nabla_x\varphi_0(x,z,t) = F \circ H(z)$ for some function F. Consequently,

$$\int_U v(z)\cdot\nabla_x\varphi_0\,dz = \int_U v(z)\cdot F(H(z))\,dz = \int_U \nabla_z \times (F \circ H)\,dz = \int_{\partial U} F \circ H \cdot dl = 0, \quad \text{(B.4)}$$

since $F(H)$ is constant on ∂U. This ensures that the compatibility condition for (B.3) is satisfied.

In order to express φ_1 more conveniently, define the corrector $\chi = (\chi_1, \chi_2)$ to be the solution of the normalised cell problem

$$\frac{\varepsilon}{2}\Delta_z\chi_i + v(z)\cdot\nabla_z\chi_i = -v_i(z)\Big(\frac{\partial_{x_i}\varphi_0}{\langle\partial_{x_i}\varphi_0\rangle}\Big), \tag{B.5}$$

with periodic in z boundary conditions. Here $\langle f\rangle$ denotes the average of f with respect to the fast variable z. We remark that the "standard corrector", denoted by $\bar\chi = (\bar\chi_1, \bar\chi_2)$, is usually chosen to be the solution of the cell problem

$$\frac{\varepsilon}{2}\Delta_z\bar\chi_i + v(z)\cdot\nabla_z\bar\chi_i = -v_i(z), \tag{B.6}$$

Our corrector χ has an extra term depending on the fast variable, however this dependence is only through the projection onto the Reeb graph.

---

[7]Here, analogous to (3.2), $\Gamma_\varepsilon(z) = (i, (\delta/\varepsilon)^{1/2}|H(z)|) \in \mathcal{G}$ when $z \in U_i$.

With this notation, observe

$$\varphi_1 = \chi \cdot \nabla_x \langle \varphi_0 \rangle.$$

Balancing terms in φ yields

$$\frac{1}{\log(\delta/\varepsilon)} \partial_t \varphi_0 = \frac{\varepsilon}{2\delta} \Delta\varphi_0 + \frac{\gamma^2}{\delta} v(z) \cdot \nabla_x \varphi_1 + o(1)$$

$$= \frac{\varepsilon}{2\delta} \Delta\varphi_0 + \frac{\gamma^2}{\delta} v(z) \cdot \nabla_x \Big(\chi \cdot \nabla_x \langle \varphi_0 \rangle \Big) + o(1). \tag{B.7}$$

We will deduce (4.2a) and (4.2b) from this by multiplying by an appropriate test function and integrating.

To obtain (4.2a), let $\phi = \phi(z)$ be a test function that is compactly supported in $\mathbb{R}^2 - \mathcal{L}$, and only depends on z through the projection $\Gamma_\varepsilon(z)$ (i.e., $v(z) \cdot \nabla_z \phi = 0$). Multiplying (B.7) by ϕ and integrating gives

$$\int_{z \in \mathcal{T}} \Big(\frac{\partial_t \varphi_0}{\log(\delta/\varepsilon)} - \frac{\varepsilon}{2\delta} \Delta\varphi_0 \Big) \phi \, dz = \frac{\gamma^2}{\delta} \int_{z \in \mathcal{T}} v(z) \cdot \nabla_x \Big(\chi \cdot \nabla_x \langle \varphi_0 \rangle \Big) \phi(x, z) \, dz. \tag{B.8}$$

Note that away from the separatrix the corrector χ has oscillations of order ε on connected components of level sets of H (see for instance [Chi79, NPR05]). Hence

$$v(z) \cdot \nabla_z \chi = \mathcal{O}(\varepsilon), \qquad \text{and} \qquad v(z) \cdot \nabla_z \phi = 0.$$

Consequently (following (B.4)) the integral on the right of (B.8) vanishes as $\varepsilon \to 0$. Restricting our attention to U_i, writing (B.8) in terms of y and using the co-area formula gives

$$0 = \int_{U_i} \Big(\frac{\partial_t \varphi_0}{\log(\delta/\varepsilon)} - \frac{\varepsilon}{2\delta} \Delta\varphi_0 \Big) \phi \, dz + o(1) = \int_{U_i} \Big(\frac{\partial_t \varphi_0}{\log(\delta/\varepsilon)} - \frac{\varepsilon}{2\delta} \Delta_z \varphi_0 \Big) \phi \, dz + o(1)$$

$$= \int_{y_i=0}^{(\frac{\delta}{\varepsilon})^{1/2} \max_{U_i} |H|} \int_{U_i \cap \{|H|=(\frac{\varepsilon}{\delta})^{1/2} y_i\}} \Big(\frac{\partial_t \varphi_0}{\log(\delta/\varepsilon)|\nabla H|} - \frac{|\nabla H|}{2} \partial_{y_i}^2 \varphi_0$$

$$- \Big(\frac{\varepsilon}{\delta} \Big)^{1/2} \frac{\Delta H}{2|\nabla H|} \partial_{y_i} \varphi_0 \Big) \phi \, dl \, dy_i + o(1)$$

$$\xrightarrow{\varepsilon \to 0} \int_{y \in I_i} \Big(\frac{q_i}{a_i} \partial_t \theta - \frac{q_i}{2} D_i^2 \theta \Big) \phi(y) \, dy.$$

This is exactly the weak form of (4.2a).

For the gluing condition (4.2b) define the boundary layer $\mathcal{V}_\varepsilon$ to be a small neighbourhood of $\mathcal{L}$ where the effects of the diffusion and convection balance. Explicitly, set

$$\mathcal{V}_\varepsilon \overset{\text{def}}{=} \Big\{ z \in \mathcal{T} : |H(z)| < \sqrt{N\varepsilon}, \Big\},$$

where $N = N(\varepsilon) \to \infty$ arbitrarily slowly as $\varepsilon \to 0$. Note that the volume of the boundary layer satisfies $|\mathcal{V}_\varepsilon| = \mathcal{O}((N\varepsilon)^{1/2} |\log\varepsilon|)$ asymptotically as $\varepsilon \to 0$. Multiplying (B.7) by $(\delta/N\varepsilon)^{1/2}$ and integrating over $\mathcal{V}_\varepsilon$ gives

$$o(1) = \Big(\frac{\delta}{N\varepsilon} \Big)^{1/2} \int_{\mathcal{V}_\varepsilon} \Big(\frac{\partial_t \varphi_0}{\log(\delta/\varepsilon)} - \frac{\varepsilon}{2\delta} \Delta\varphi_0 - \frac{\gamma^2}{\delta} v(z) \cdot \nabla_x \Big(\chi \cdot \nabla_x \langle \varphi_0 \rangle \Big) \Big) dz$$

$$= -\frac{1}{2} \Big(\frac{\varepsilon}{N\delta} \Big)^{1/2} \int_{\partial\mathcal{V}_\varepsilon} \nabla_z \varphi_0 \cdot \hat{n} \, dl - \frac{\gamma^2}{\sqrt{N\varepsilon\delta}} \sum_{i,j=1}^{2} \Big(\int_{\mathcal{V}_\varepsilon} v_i \chi_j \, dz \Big) \partial_{x_i} \partial_{x_j} \langle \varphi_0 \rangle + o(1).$$

This forces

$$\left(\frac{\varepsilon}{\delta}\right)^{1/2} \int_{\partial \mathcal{V}_\varepsilon} \nabla_z \varphi_0 \cdot \hat{n} \, dl + \frac{\gamma^2}{\sqrt{\delta}} \sum_{i,j=1}^{2} \left(\frac{2}{\sqrt{\varepsilon}} \int_{\mathcal{V}_\varepsilon} v_i \chi_j \, dz\right) \partial_{x_i} \partial_{x_j} \langle \varphi_0 \rangle = o(1). \tag{B.9}$$

Clearly

$$\left(\frac{\varepsilon}{\delta}\right)^{1/2} \int_{\partial \mathcal{V}_\varepsilon} \nabla_z \varphi_0 \cdot \hat{n} \, dl = \left(\frac{\varepsilon}{\delta}\right)^{1/2} \sum_{i=1}^{M} \int_{\partial \mathcal{V}_\varepsilon \cap U_i} \nabla_z \varphi_0 \cdot \hat{n} \, dl \xrightarrow{\varepsilon \to 0} \sum_{i=1}^{M} q_i D_i \theta.$$

For the second term in (B.9) define the 2×2 matrix Q by

$$Q_{i,j} = \left(\sum_{k=1}^{M} q_k\right)^{-1} \lim_{\varepsilon \to 0} \frac{1}{\sqrt{\varepsilon}} \int_{\mathcal{V}_\varepsilon} \left(v_i \chi_j + v_j \chi_i\right) dz. \tag{B.10}$$

If we replace χ with $\bar{\chi}$, then the limit on the right is well studied. Indeed, using (B.6) we see

$$\frac{1}{\sqrt{\varepsilon}} \int_{\mathcal{V}_\varepsilon} \left(v_i \bar{\chi}_j + v_j \bar{\chi}_i\right) dz = \sqrt{\varepsilon} \int_{\mathcal{V}_\varepsilon} \nabla \bar{\chi}_i \cdot \nabla \bar{\chi}_j \, dz - \frac{\sqrt{\varepsilon}}{2} \int_{\partial \mathcal{V}_\varepsilon} \left(\bar{\chi}_i \nabla \bar{\chi}_j + \bar{\chi}_j \nabla \bar{\chi}_i\right) \cdot \hat{n} \, dl.$$

It is well known that $\nabla \bar{\chi}$ is $\mathcal{O}(\varepsilon^{-1/2})$ in an $\sqrt{\varepsilon}$ neighbourhood of the boundary, and $\mathcal{O}(1)$ elsewhere. Consequently, we expect the right hand side of the above to converge as $\varepsilon \to 0$ (see for instance [FP94]).

In our situation, the extra term $\partial_{x_i} \varphi_0 / \langle \partial_{x_i} \varphi_0 \rangle$ only depends on the fast variable z through its projection $\Gamma_\varepsilon(z)$. Hence the asymptotic behaviour of χ is similar to that of $\bar{\chi}$, and consequently the limit in (B.10) should exists. Thus, equation (B.9) yields

$$\sum_{k=1}^{M} q_k D_k^y \theta + \left(\sum_{k=1}^{M} q_k\right)[Q : \nabla_x^2] \varphi_0 = 0,$$

since $\gamma^2 = \delta$. This is exactly (4.2b), as desired.

Remark B.1. *Choosing $\delta = 2\varepsilon$, the above method also provides an effective equation for $\tilde{\theta}$ on time scales of order $1/\varepsilon$, as considered in [HKP14]. In this regime the analogue of equation (4.2a) is now on a finite graph (the unscaled Reeb graph of H), and the generator $\mathcal{A}$ has non-constant (singular) coefficients.*

References

[ACV16]   M. ALLEN, L. CAFFARELLI, and A. VASSEUR. A Parabolic Problem with a Fractional Time Derivative. *Arch. Ration. Mech. Anal.* **221**, no. 2, (2016), 603–630. `doi:10.1007/s00205-016-0969-z`.

[AMB11]   S. A. ALMADA MONTER and Y. BAKHTIN.   Normal forms approach to diffusion near hyperbolic equilibria.   *Nonlinearity* **24**, no. 6, (2011), 1883–1907.   `arXiv:1006.3000v1`. `doi:10.1088/0951-7715/24/6/011`.

[BAČ07]   G. BEN AROUS and J. ČERNÝ. Scaling limit for trap models on $\mathbb{Z}^d$. *Ann. Probab.* **35**, no. 6, (2007), 2356–2384. `arXiv:math/0606719`. `doi:10.1214/009117907000000024`.

[Bak11]   Y. BAKHTIN. Noisy heteroclinic networks. *Probab. Theory Related Fields* **150**, no. 1-2, (2011), 1–42. `arXiv:0712.3952`. `doi:10.1007/s00440-010-0264-0`.

[Bil99]   P. BILLINGSLEY. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second ed., 1999, x+277. `doi:10.1002/9780470316962`.

[BLP78]   A. BENSOUSSAN, J.-L. LIONS, and G. PAPANICOLAOU. *Asymptotic analysis for periodic structures*, vol. 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1978, xxiv+700.

[Bou92]   J. P. BOUCHAUD. Weak ergodicity breaking and aging in disordered systems. *J. Phys. I France* **2**, no. 9, (1992), 1705–1713. `doi:10.1051/jp1:1992238`.

[Chi79]   S. CHILDRESS. Alpha-effect in flux ropes and sheets. *Phys. Earth Planet Inter.* **20**, (1979), 172–180.

[CS89]   S. CHILDRESS and A. M. SOWARD. Scalar transport and alpha-effect for a family of cat's-eye flows. *J. Fluid Mech.* **205**, (1989), 99–133. `doi:10.1017/S0022112089001965`.

[Die10]   K. DIETHELM. *The analysis of fractional differential equations*, vol. 2004 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2010, viii+247. `doi:10.1007/978-3-642-14574-2`.

[DK08]   D. DOLGOPYAT and L. KORALOV. Averaging of Hamiltonian flows with an ergodic component. *Ann. Probab.* **36**, no. 6, (2008), 1999–2049. `arXiv:0901.2776`. `doi:10.1214/07-AOP372`.

[DK13]   D. DOLGOPYAT and L. KORALOV. Averaging of incompressible flows on two-dimensional surfaces. *J. Amer. Math. Soc.* **26**, no. 2, (2013), 427–449. `doi:10.1090/S0894-0347-2012-00755-3`.

[EK86]   S. N. ETHIER and T. G. KURTZ. *Markov processes: characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986, x+534. `doi:10.1002/9780470316658`.

[Fan02]   A. FANNJIANG. Time scales in homogenization of periodic flows with vanishing molecular diffusion. *J. Differential Equations* **179**, no. 2, (2002), 433–455. `doi:10.1006/jdeq.2001.4039`.

[FP94]   A. FANNJIANG and G. PAPANICOLAOU. Convection enhanced diffusion for periodic flows. *SIAM J. Appl. Math.* **54**, no. 2, (1994), 333–408. `doi:10.1137/S0036139992236785`.

[Fre64]   M. I. FREĬDLIN. The Dirichlet problem for an equation with periodic coefficients depending on a small parameter. *Teor. Verojatnost. i Primenen.* **9**, (1964), 133–139.

[FS00]   M. FREIDLIN and S.-J. SHEU. Diffusion processes on graphs: stochastic differential equations, large deviation principle. *Probability Theory and Related Fields* **116**, no. 2, (2000), 181–220. `doi:10.1007/PL00008726`.

[FW93]   M. I. FREIDLIN and A. D. WENTZELL. Diffusion processes on graphs and the averaging principle. *Ann. Probab.* **21**, no. 4, (1993), 2215–2245. `doi:10.1214/aop/1176989018`.

[FW12]   M. I. FREIDLIN and A. D. WENTZELL. *Random perturbations of dynamical systems*, vol. 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third ed., 2012, xxviii+458. Translated from the 1979 Russian original by Joseph Szücs. `doi:10.1007/978-3-642-25847-3`.

[HKP14]   M. HAIRER, L. KORALOV, and Z. PAJOR-GYULAI. From averaging to homogenization in cellular flows - an exact description of the transition. *ArXiv e-prints* (2014). To appear in Ann. Inst. H. Poincaré. `arXiv:1407.0982`.

[IKNR14]   G. IYER, T. KOMOROWSKI, A. NOVIKOV, and L. RYZHIK. From homogenization to averaging in cellular flows. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **31**, no. 5, (2014), 957–983. `arXiv:1108.0074`. `doi:10.1016/j.anihpc.2013.06.003`.

[IN16]   G. IYER and A. NOVIKOV. Anomalous diffusion in fast cellular flows at intermediate time scales. *Probab. Theory Related Fields* **164**, no. 3-4, (2016), 707–740. `arXiv:1406.3881`. `doi:10.1007/s00440-015-0617-9`.

[Kif81]   Y. KIFER. The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. *Israel J. Math* **40**, no. 1, (1981), 74–96.

[Kor04]   L. KORALOV. Random perturbations of 2-dimensional Hamiltonian flows. *Probab. Theory Related Fields* **129**, no. 1, (2004), 37–62. `doi:10.1007/s00440-003-0320-0`.

[KS07]   L. KORALOV and Y. SINAI. *Theory of Probability and Random Processes*. Universitext. Springer Berlin Heidelberg, 2007.

[Lej06]   A. LEJAY. On the constructions of the skew Brownian motion. *Probab. Surv.* **3**, (2006), 413–466. `doi:10.1214/154957807000000013`.

[Man68]   P. MANDL. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague; Springer-Verlag New York Inc., New York, 1968, xx+192.

[MS04]   M. M. MEERSCHAERT and H.-P. SCHEFFLER. Limit theorems for continuous-time random dom walks with infinite mean waiting times. *J. Appl. Probab.* **41**, no. 3, (2004), 623–638. `doi:10.1239/jap/1091543414`.

[MS08]   M. M. MEERSCHAERT and H.-P. SCHEFFLER. Triangular array limits for continuous time random walks. *Stochastic Process. Appl.* **118**, no. 9, (2008), 1606–1633. `doi:10.1016/j.spa.2007.10.005`.

[MS12]   M. M. MEERSCHAERT and A. SIKORSKII. *Stochastic models for fractional calculus*, vol. 43 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2012, x+291.

[Ngu89]   G. NGUETSENG. A general convergence result for a functional related to the theory of homogenization. *SIAM J. Math. Anal.* **20**, no. 3, (1989), 608–623. `doi:10.1137/0520043`.

[NPR05]   A. NOVIKOV, G. PAPANICOLAOU, and L. RYZHIK. Boundary layers for cellular flows at high Péclet numbers. *Comm. Pure Appl. Math.* **58**, no. 7, (2005), 867–922. `doi:10.1002/cpa.20058`.

[Oll94]   S. OLLA. *Lectures on Homogenization of Diffusion Processes in Random Fields*. Publications de l'Ecole Doctorale de l'Ecole Polytechnique. 1994.

[PGS15]   Z. PAJOR-GYULAI and M. SALINS. On dynamical systems perturbed by a null-recurrent motion: The general case. *ArXiv e-prints* (2015). `arXiv:1508.05346`.

[PS08]   G. A. PAVLIOTIS and A. M. STUART. *Multiscale methods*, vol. 53 of *Texts in Applied Mathematics*. Springer, New York, 2008, xviii+307. `doi:10.1007/978-0-387-73829-1`.

[RBDH87]   M. N. ROSENBLUTH, H. L. BERK, I. DOXAS, and W. HORTON. Effective diffusion in laminar convective flows. *Phys. Fluids* **30**, (1987), 2636–2647.

[Ree46]   G. REEB. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *C. R. Acad. Sci. Paris* **222**, (1946), 847–849.

[RW00a]   L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000, xx+386. Foundations, Reprint of the second (1994) edition. `doi:10.1017/CBO9781107590120`.

[RW00b]   L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000, xiv+480. Itô calculus, Reprint of the second (1994) edition. `doi:10.1017/CBO9781107590120`.

[RY99]   D. REVUZ and M. YOR. *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenchaften A series of comprehensive studies in mathematics. Springer, 1999.

[YJ91]   W. R. YOUNG and S. JONES. Shear dispersion. *Physics of Fluids A: Fluid Dynamics (1989-1993)* **3**, no. 5, (1991), 1087–1101.

[You88]   W. R. YOUNG. Arrested shear dispersion and other models of anomalous diffusion. *J. Fluid Mech.* **193**, (1988), 129–149. `doi:10.1017/S0022112088002083`.

[YPP89]   W. YOUNG, A. PUMIR, and Y. POMEAU. Anomalous diffusion of tracer in convection rolls. *Phys. Fluids A* **1**, no. 3, (1989), 462–469. `doi:10.1063/1.857415`.