Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 4, 1592–1613
DOI: 10.1214/15-AIHP690

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

From averaging to homogenization in cellular flows – An exact description of the transition

Martin Hairer[a], Leonid Koralov[b] and Zsolt Pajor-Gyulai[b]

[a]*Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. E-mail: M.Hairer@Warwick.ac.uk*
[b]*Department of Mathematics, University of Maryland, College Park, MD 20742-4015, USA.*
E-mail: koralov@math.umd.edu; pgyzs@math.umd.edu

Received 4 July 2014; revised 20 April 2015; accepted 26 May 2015

Abstract. We consider a two-parameter averaging-homogenization type elliptic problem together with the stochastic representation of the solution. A limit theorem is derived for the corresponding diffusion process and a precise description of the two-parameter limit behavior for the solution of the PDE is obtained.

Résumé. Nous considérons un problème elliptique de type moyennisation / homogénisation à deux paramètres, en combinaison avec la représentation stochastique de la solution. Nous obtenons un théorème limite pour le processus de diffusion correspondant ainsi qu'une description précise du comportement limite de la solution de l'EDP.

MSC: 35B27

Keywords: Local time; Diffusion on graphs; Averaging; Homogenization

1. Introduction

Let $D_R \subseteq \mathbb{R}^2$ be obtained from a bounded smooth domain D by stretching it by a factor R. Consider the elliptic Dirichlet problem

$$\frac{1}{2}\Delta u^{\varepsilon,R} + \frac{1}{\varepsilon} v \nabla u^{\varepsilon,R} = -f\left(\frac{x}{R}\right) \quad \text{in } D_R, \qquad u^{\varepsilon,R}|_{\partial D_R} = 0, \tag{1.1}$$

where f is a bounded continuous function on D and v is a smooth incompressible periodic Hamiltonian vector field. For simplicity, assume that D contains the origin. We further assume that the stream function $H(x_1, x_2)$ such that

$$v = \nabla^{\perp} H = (-\partial_2 H, \partial_1 H),$$

is itself periodic in both variables, that is, the integral of v over the periodicity cell is zero. Let us assume for simplicity that the period is one in both directions. We will denote the cell of periodicity by $\mathcal{T}$, which can be viewed as a unit square or, alternatively, as a torus. Our main additional structural assumption is that the critical points of H are non degenerate and that there is a level set $\mathcal{L}$ of H (say $\mathcal{L} = \{x \in \mathbb{R}^2 : H(x) = 0\}$ without loss of generality) containing some of the saddle points and forming a lattice in $\mathbb{R}^2$, thus dividing the plane into bounded sets that are invariant under the flow (see Figure 1). A typical example to keep in mind is the canonical cellular flow given by $H(x_1, x_2) = \sin(x_1)\sin(x_2)$.

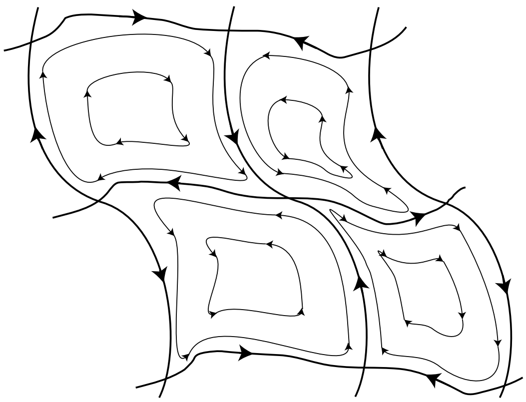

Fig. 1. A period of the cellular flow.

There are two parameters in this problem: ε measures the inverse of the strength of the vector field, while R measures the size of the domain. For fixed R (e.g. when D_R coincides with exactly one cell) and $\varepsilon \downarrow 0$, solution to (1.1) becomes constant on stream lines. Indeed, multiplying by ε and letting $\varepsilon \downarrow 0$ formally gives us $v\nabla u = 0$. The precise values of the asymptotics of the solution on each streamline are determined by an ODE corresponding to the structure of the level sets according to classical averaging results [8].

If on the other hand ε is fixed and $R \uparrow \infty$, then the asymptotic behavior of u can be obtained by homogenization (e.g. [4,13,15]), i.e., by solving an elliptic problem on D with appropriately chosen constant coefficients.

It was shown in [9] that averaging and homogenization can also be used to study the two-parameter asymptotics in certain regimes. Namely, if $R^4 \log^2 R \le c/(\varepsilon \log^2 \varepsilon)$ for some constant c as $1/\varepsilon, R \uparrow \infty$, then averaging theory applies. On the other hand, if $R^{4-\alpha} \ge 1/\varepsilon$ for some positive α, then homogenization type behavior is observed. The methods in [9] are analytic, based on investigating the asymptotic behavior of the principal Dirichlet eigenvalue of the elliptic operator, and it seems unlikely that they can be directly applied near the transition regime. To our knowledge, only numerical results were available in the intermediate cases [10,14] up until now.

In this paper, we study the two-parameter asymptotics using a probabilistic approach and we prove that the crossover from homogenization to averaging occurs when R is precisely of order $\varepsilon^{-1/4}$. In order to achieve this, we study the family of two dimensional diffusion processes associated to (1.1), namely

$$dX_t^{x,\varepsilon} = \frac{1}{\varepsilon} v\big(X_t^{x,\varepsilon}\big)\,dt + dW_t, \qquad X_0^{x,\varepsilon} = x,$$

on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where W_t is a two dimensional Brownian motion. Our goal is to obtain a limit theorem as $\varepsilon \downarrow 0$ provided that $X_t^{x,\varepsilon}$ is considered on scales of order $\varepsilon^{-1/4}$, and to identify the limiting process as a time changed Brownian motion. The time change arising in the construction of the limiting process is non-trivial and can be described as the local time of a diffusion process on a certain graph which we now explain.

It is well known that there is a graph G naturally associated to the structure of the level sets of H (see Figure 2), by setting $G = \mathcal{T}/\sim$ for the equivalence relation $\sim$ such that two points are equivalent if and only if they belong to the same connected level set of H. With this definition, G is a "real graph" in the sense that it is homeomorphic to finitely many copies of $[0,1]$ glued together at their extremities. As above, let $\mathcal{L} = \{x \in \mathbb{R}^2 : H(x) = 0\}$ be the connected level set of H that contains a periodic array of saddle points, and denote the corresponding level set on the torus by $\mathcal{L}_\mathcal{T}$. Let A_i, $i = 1, \ldots, n$, be the saddle points of H in $\mathcal{L}_\mathcal{T}$. Then $\mathcal{L}$ (or $\mathcal{L}_\mathcal{T}$) is the union of heteroclinic orbits connecting the A_i's and will be referred to as the separatrix. For notational simplicity, we assume that there are no homoclinic orbits, i.e., ones that connect a saddle to itself. Also, let U_i, $i = 1, \ldots, n$, be the connected components of $\mathcal{T} \setminus \mathcal{L}_\mathcal{T}$. (There is no particular connection between the numbering of the U_i's and that of the A_i's, although by Euler's theorem there is actually the same number of them.) For convenience, we also assume that there are no saddle points of H inside any U_i. The graph G then consists of an interior vertex O, which is the image of $\mathcal{L}_\mathcal{T}$ in G, and of n edges connecting O to the exterior vertices corresponding to the extrema of H. Every other point on an edge corresponds to the appropriate connected component of a level curve of $|H|$. Accordingly, $|H|$ will serve as a local

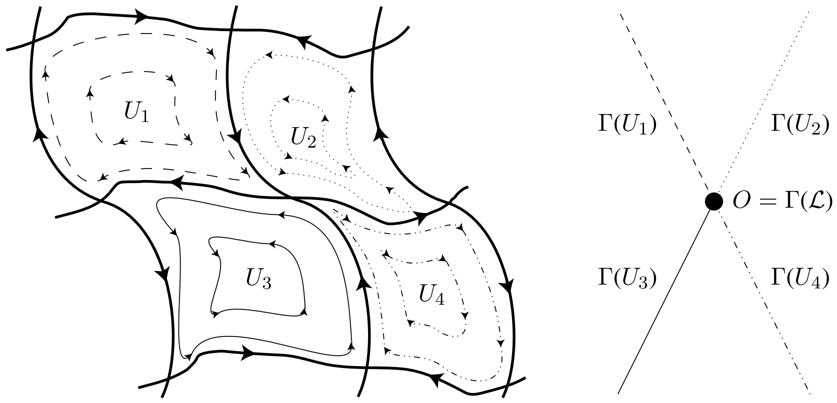

Fig. 2. The graph corresponding to the structure of the level sets of H on $\mathcal{T}$.

coordinate on each edge I_i, which gives G a natural metric structure. This is also called the Reeb graph of H in the topology literature.

From now on, we will therefore identify G with a subset of $\mathbb{N} \times \mathbb{R}_+$ by writing (i, z) for the connected component of $\{x : |H(x)| = z\}$ lying in U_i. We make the abuse of notation of identifying all the points $(i, 0)$ with each other, and we write

$$\Gamma : \mathcal{T} \to G, \qquad \Gamma(x) = \bigl(i, \bigl|H(x)\bigr|\bigr) \quad \text{if } x \in \overline{U}_i,$$

for the canonical map. (Although the sets $\overline{U}_i$ overlap, this is well defined since we have identified the points $(i, 0)$.) Note that Γ takes each U_i into an edge I_i of the graph and the extrema inside each U_i are mapped to the corresponding exterior vertices. Naturally, Γ can be extended periodically to the entire plane.

It was shown in [8, Chapter 8] that the non-Markovian processes $\Gamma(X_t^{x,\varepsilon})$ converge in distribution, as $\varepsilon \downarrow 0$, to a diffusion on G. Let us describe this limiting process briefly. On the ith edge of the graph, the process is a diffusion with generator

$$\mathcal{A}_i = \frac{a(i,z)^2}{2} \frac{d^2}{dz^2} + b(i,z) \frac{d}{dz},$$

where the coefficients $a(i, z)$, $b(i, z)$ can be computed explicitly from H. The behavior of the process at the interior vertex O can also be described in terms of H. More precisely, for a set of constants $\alpha_i > 0$ with $\sum_{i=1}^n \alpha_i = 1$, we can define an operator A on the domain $D(A)$ that consists of the functions F that satisfy:

(a) $F \in \mathcal{C}(G)$ and furthermore $F \in \mathcal{C}^2(I_i)$ for each edge i,
(b) $\mathcal{A}_i F(z)$, $z \in I_i$, which is defined on the union of the interiors of all the edges, can be extended to a continuous function on G,
(c) $\sum_{i=1}^n \alpha_i D_i F(O) = 0$, where $D_i F(O)$ is the one-sided interior derivative of F along the edge I_i.

We then define the operator A by $AF|_{I_i} = \mathcal{A}_i F|_{I_i}$. Below, we are going to write $y = (i, z)$ to refer to a point on G. As shown in [7], A generates a Fellerian Markov family Y_t^y on G. With these notations at hand, the measures on $\mathcal{C}([0,\infty); G)$ induced by the processes $\Gamma(X_t^{x,\varepsilon})$ converge weakly to the one induced by the process $Y_t^{\Gamma(x)}$, provided that the constants $\{\alpha_i\}_{i=1}^n$ are suitably chosen.

Note that the classical Freidlin–Wentzell theory requires $H(x) \to \infty$ as $|x| \to \infty$. Nevertheless, adapting the results for the compact setting on $\mathcal{T}$ is trivial.

Definition 1.1. *The local time of Y^{y_0} is the unique nonnegative random field*

$$L^{y_0} = \bigl\{L_t^{y_0}(y) : (t, y) \in [0,\infty) \times G\bigr\}$$

such that the following hold:

1. *The mapping* $(t, y) \to L_t^{y_0}(y)$ *is measurable and* $L_t^{y_0}(y)$ *is adapted.*
2. *For each* $y \in G$, *the mapping* $t \to L_t^{y_0}(y)$ *is non-decreasing and constant on each open interval where* $Y_t^{y_0} \neq y$.
3. *For every Borel measurable* $f : G \to [0, \infty)$, *we have*

$$\int_0^t f\big(Y_s^{y_0}\big)a^2\big(Y_s^{y_0}\big)\,ds = 2\int_G f(y)L_t^{y_0}(y)\,dy \quad a.s.$$

4. $L_t^{y_0}(y)$ *is a.s. jointly continuous in* t *and* y *for* $y \neq O$, *while*

$$L_t^{y_0}(O) = \sum_{i=1}^n \lim_{y\to O, y\in I_i} L_t^{y_0}(y).$$

The existence and uniqueness of local time for diffusions on the real line is relatively well studied. These standard results, together with a straightforward modification of Lemma 2.2 of [6], give the existence and uniqueness for the local time on the graph. Note (see e.g. [8]) that for processes on G arising from the averaging of a Hamiltonian system, $a^{-2}(\cdot)$ is locally integrable near the interior vertex, which is sufficient for the method of [6] to work.

The main result of this paper is the following. For a positive definite symmetric matrix Q, let $\tilde{W}_t^Q$ be a two dimensional Brownian motion with covariance matrix Q. Assume that the families of processes Y_t^y, and $\tilde{W}_t^Q$ are independent. Also consider the process $\tilde{W}^Q_{L_t^y}$, where $L_t^y = L_t^y(O)$ is the local time of Y_t^y at the interior vertex.

Theorem 1.2. *There exists a strictly positive definite matrix* Q *such that the law of the process* $\varepsilon^{1/4}X_t^{x,\varepsilon}$ *converges, as* $\varepsilon \downarrow 0$, *to that of* $\tilde{W}^Q_{L_t^{\Gamma(x)}}$.

Remark 1.3. *One might also consider the process* $X_t^{x,\varepsilon}$ *on slightly shorter timescales, i.e.,* $t \sim \varepsilon^\alpha|\log\varepsilon|$ *with* $\alpha \in (0, 1)$ *. At first glance, this may appear uninteresting since, for a generic starting point* x, *this would simply lead to a fast rotation on the level set* $\{y : H(y) = H(x)\}$. *However, if we consider a starting point on (or sufficiently close to) the separatrix, one expects to see a non-trivial limiting process also at these shorter scales. It is natural to conjecture that this process, after spatial re-scaling by* $\varepsilon^{\frac{1-\alpha}{4}}$ *, is given by* $\tilde{W}^Q_{L_t}$, *where* L_t *is the local time of a Brownian motion at the vertex of a star-shaped graph. A similar process, called* $FK(1/2)$, *already arose as the scaling limit for heavy-tailed trap models in* [1]. *The authors are planning to return to this question in a forthcoming publication. We also mention that the* $\alpha = 1$ *case has been considered in* [2].

It is well known that the solution of (1.1) can be represented as

$$u^{\varepsilon,R}(x) = \mathbf{E}\int_0^{\tau_{\partial D_R}(X_\cdot^{x,\varepsilon})} f\big(X_s^{x,\varepsilon}/R\big)\,ds,$$

where $\tau_{\partial D_R}(\omega)$ is the first hitting of the boundary of D_R by the trajectory $\omega \in \mathcal{C}([0, T]; \mathbb{R}^2)$. The essence of the averaging and transition regimes can be captured by the mechanism of the exit of the process X_t^ε from D_R (see [9]).

In the averaging regime, the process $X_t^{x,\varepsilon}$ revolves many times roughly along the flow lines within one cell, but once the separatrix is reached, the process exits D_R quickly (as intuitively follows from the typical fluctuation of the limiting Brownian motion after one notices that the local time immediately becomes non-zero after the process reaches the boundary).

On the other hand, in the homogenization regime, the interiors of many cells are visited before the process exits D_R, and there is enough time for the process $L_t^{\Gamma(x)}$ to start growing nearly linearly in t, and therefore an overall Brownian behavior to set in. The mean exit time becomes infinite in the limit.

In the intermediate transition regime, the time required to leave D_R remains finite and is of the same order as the local time, although $L_t^{\Gamma(x)}$ is not directly proportional to t in this regime. We will apply Theorem 1.2 in order to obtain the following asymptotic results for the solution of equation (1.1). The precise statement of our results from a PDE perspective can be summarized by the following theorem.

Theorem 1.4. *Let $\varepsilon \downarrow 0$ and $R = R(\varepsilon) \uparrow \infty$ in* (1.1).

1. (*Averaging regime*) *If $R\varepsilon^{1/4} \downarrow 0$, then*

$$u^{\varepsilon,R}(x) \to f(0) \cdot \mathbf{E}\overline{\tau}_0\big(Y_\cdot^{\Gamma(x)}\big),$$

where $\overline{\tau}_0$ is the first time when a process on G hits the interior vertex.

2. (*Transition regime*) *If $R\varepsilon^{1/4} \to C \in (0,\infty)$, then*

$$u^{\varepsilon,R}(x) \to \mathbf{E}\int_0^{\tau_{\partial D}} f\big(\tilde{W}^{Q/C^2}_{L_t^{\Gamma(x)}}\big)\,dt,$$

with Q as in Theorem 1.2, *where $\tau_{\partial D}$ is the first time the process $\tilde{W}^{Q/C^2}_{L_t^{\Gamma(x)}}$ hits the boundary of D.*

3. (*Homogenization regime*) *There is a constant $c > 0$ such that if $R\varepsilon^{1/4} \uparrow \infty$, then*

$$\varepsilon^{-1/2}R^{-2}u^{\varepsilon,R}(x) \to \mathbf{E}\int_0^{\tau_{\partial D}} f\big(\tilde{W}_t^{cQ}\big)\,dt, \tag{1.2}$$

where $\tilde{W}_t^{cQ}$ is a Brownian motion with covariance cQ and $\tau_{\partial D}$ is the first time the process $\tilde{W}_t^{cQ}$ hits the boundary of D.

Remark 1.5. *Note that there is no x dependence on the right-hand side of* (1.2). *If we scale the problem back to the original domain D and then normalize appropriately, the above result gives us that the limit is the solution of a constant coefficient Dirichlet problem on D evaluated at the origin. To get the values of this solution at another point x, we must apply the result to the shifted domain $D - x$. This way we can prove that*

$$\big(\varepsilon^{1/2}R^2\big)^{-1}u^{\varepsilon,R}(Rx) \to \mathbf{E}\int_0^{\tau_{\partial D}} f\big(x + \tilde{W}_t^{cQ}\big)\,dt \quad \text{as } \varepsilon \downarrow 0, R \uparrow \infty,$$

which contains the classical homogenization result. Here $\tau_{\partial D}$ is the first time when the process $x + \tilde{W}_t^{cQ}$ hits the boundary of D.

Remark 1.6. *Although it is not an aim of the present paper, Theorem* 1.2 *can also be used to derive asymptotics for PDEs with periodic right-hand side and for parabolic problems* (*using the well known probabilistic representations*). *These techniques are suitable for investigating equations with non-zero boundary data as well.*

This paper is organized as follows. In Section 2, we derive a limit theorem to describe the displacement that occurs when the process leaves the interior of a cell and comes close to the separatrix. This, combined with a Lévy-type downcrossing representation of the local time at the interior vertex, will help us prove Theorem 1.2 in Section 3. Section 4 is dedicated to the proof of Theorem 1.4.

2. Displacement when the process is near the separatrix

In this section we study the behavior of the process when it is close to the separatrix. The process spends most of the time in the interiors of the cells where no cell changes are possible. However, when the process leaves the cell interior, rapid displacement occurs along the separatrix. We will show what happens during one excursion, i.e., between the time when the process hits the separatrix and the time when it goes back to the interior of the domain (the exact meaning of the latter will be explained below).

First, we need some notations. For any two saddle points, introduce $\gamma(A_i, A_j)$ as the set of points in $\mathcal{L}_\mathcal{T}$ that get taken to A_j by the flow $\dot{x} = v(x)$ and to A_i by the flow $\dot{x} = -v(x)$. Since we assumed that the separatrices do not form loops, we always have $\gamma(A_i, A_i) = \varnothing$.

Let $V^\delta = \{x \in \mathbb{R}^2 : |H(x)| \le \delta\}$. For δ sufficiently small, we can make a continuous coordinate change $(x_1, x_2) \to (H, \theta)$ in $V^\delta \cap \overline{U}_k$. Here θ takes values in $[0, \int_{\partial U_k} |\nabla H|\, dl]$, with the endpoints of the interval identified, and satisfies the conditions

1. Its gradient satisfies $|\nabla\theta| = |\nabla H|$ on the curve $\gamma(A_i, A_j)$.
2. The function θ is constant on curves perpendicular to the level sets of H.

Note that this defines θ uniquely, up to the choice of origin $\theta = 0$ and the direction in which θ increases. Note also that, as a consequence of the smoothness of H, θ is smooth in a neighborhood of $\gamma(A_i, A_j)$ for each A_i, A_j such that $\gamma(A_i, A_j) \subset \overline{U}_k$. Using these new coordinates, we can define what it means for the process to pass a saddle point. Namely, let

$$B(A_i, U_k) = \{x \in V^\delta \cap \overline{U}_k : \theta(x) = \theta(A_i)\}, \qquad B(A_i) = \bigcup_{k: A_i \in \partial U_k} B(A_i, U_k).$$

Observe that $B(A_i, U_k)$ is a curve in U_k transversal to the flow with an endpoint being the saddle point A_i. Let $\pi : \mathbb{R}^2 \to \mathcal{T}$ be the quotient map from the plane to the torus and, for simplicity, let us denote $\pi(V^\delta)$ by V^δ again. Introduce the stopping times $\alpha_0^{x,\delta,\varepsilon} = 0$, $\beta_0^{x,\delta,\varepsilon} = \inf\{t \ge 0 : X_t^{x,\varepsilon} \in \mathcal{L}\}$ and recursively define $\alpha_n^{x,\delta,\varepsilon}$ and $\beta_n^{x,\delta,\varepsilon}$ as follows. Given $\beta_{n-1}^{x,\delta,\varepsilon}$, find i and j such that $\pi(X^{x,\varepsilon}_{\beta_{n-1}^{x,\delta,\varepsilon}}) \in \gamma(A_i, A_j)$. Then we define

$$\alpha_n^{x,\delta,\varepsilon} = \inf\left\{t \ge \beta_{n-1}^{x,\delta,\varepsilon} : \pi(X_t^{x,\varepsilon}) \in \bigcup_{k \ne i} B(A_k) \cup \partial V^\delta\right\},$$

$$\beta_n^{x,\delta,\varepsilon} = \inf\{t \ge \alpha_n^{x,\delta,\varepsilon} : X_t^{x,\varepsilon} \in \mathcal{L}\}.$$

In other words, $\alpha_n^{x,\delta,\varepsilon}$ is the first time after $\beta_{n-1}^{x,\delta,\varepsilon}$ that the process either hits ∂V^δ, or goes past a saddle point different from the one behind $X^{x,\varepsilon}_{\beta_{n-1}^{x,\delta,\varepsilon}}$.

We introduce another pair of sequences of stopping times corresponding to successive visits to $\mathcal{L}$ and ∂V^δ. Namely, let $\mu_0^{x,\delta,\varepsilon} = 0$, $\sigma_0^{x,\delta,\varepsilon} = \beta_0^{x,\delta,\varepsilon}$, and recursively define

$$\mu_n^{x,\delta,\varepsilon} = \inf\{t \ge \sigma_{n-1}^{x,\delta,\varepsilon} : X_t^{x,\varepsilon} \in \partial V^\delta\}, \qquad \sigma_n^{x,\delta,\varepsilon} = \inf\{t \ge \mu_n^{x,\delta,\varepsilon} : X_t^{x,\varepsilon} \in \mathcal{L}\}.$$

Let

$$S_n^{x,\delta,\varepsilon} = X^{x,\varepsilon}_{\sigma_n^{x,\delta,\varepsilon}} - X^{x,\varepsilon}_{\sigma_{n-1}^{x,\delta,\varepsilon}}, \quad n \ge 1, \qquad T_n^{x,\delta,\varepsilon} = \sigma_n^{x,\delta,\varepsilon} - \mu_n^{x,\delta,\varepsilon}, \quad n \ge 0,$$

be the displacement between successive visits to $\mathcal{L}$ and the time spent on the nth downcrossing of V^δ, respectively. We will use the following notion of uniform weak convergence for probability measures in the sequel.

Definition 2.1. *Given two families of random variables $f^{x,\varepsilon}$ and g^x with values in a metric space M and indexed by a parameter x, we will say that $f^{x,\varepsilon}$ converge to g^x in distribution uniformly in x if*

$$\mathbf{E}\varphi(f^{x,\varepsilon}) \to \mathbf{E}\varphi(g^x),$$

as $\varepsilon \to 0$, uniformly in x for each $\varphi \in \mathcal{C}_b(M)$.

Let $\eta^{x,\delta,\varepsilon}$ be the random vector with values in $\{1, \dots, n\}$ defined by

$$\eta^{x,\delta,\varepsilon} = i \quad \text{if } X^{x,\varepsilon}_{\mu_1^{x,\delta,\varepsilon}} \in U_i, i = 1, \dots, n,$$

i.e., $\eta^{x,\delta,\varepsilon} = i$ if the process ends up in U_i after the first upcrossing of V^δ. The main result of this section is

Theorem 2.2. *There are a 2×2 non-degenerate matrix Q, a vector $(p_1, \dots, p_n)$, and functions $a(\delta), b_1(\delta), \dots, b_n(\delta)$ that go to zero as $\delta \to 0$, such that*

$$\left(\varepsilon^{1/4} S_1^{x,\delta,\varepsilon}, \eta^{x,\delta,\varepsilon}\right) \to \left(\sqrt{\delta}\left(1 + a(\delta)\right)\sqrt{\xi} N(0, Q), \eta^\delta\right) \tag{2.1}$$

in distribution as $\varepsilon \downarrow 0$, *uniformly in* $x \in \mathcal{L}$ *for all sufficiently small* $\delta > 0$, *where* ξ *is an exponential random variable with parameter one*, N *is a two dimensional normal with covariance matrix* Q, *independent of* ξ, *and* η^δ *is a random vector with values in* $\{1, \dots, n\}$ *independent of* ξ *and* N *such that* $\mathbf{P}(\eta^\delta = i) = p_i + b_i(\delta)$.

Before proving Theorem 2.2, let us briefly discuss one implication. Let $\bar{T}^y$ be the time it takes the limiting process Y_t^y on the graph to reach the vertex O.

Lemma 2.3. *For fixed* m *and* δ, *the random vectors*

$$\left(T_0^{x,\delta,\varepsilon}, \varepsilon^{1/4} S_1^{x,\delta,\varepsilon}, T_1^{x,\delta,\varepsilon}, \dots, T_{m-1}^{x,\delta,\varepsilon}, \varepsilon^{1/4} S_m^{x,\delta,\varepsilon}\right)$$

converge, *as* $\varepsilon \downarrow 0$, *to a random vector with independent components*. *The limiting distribution for each of the components* $\varepsilon^{1/4} S_1^{x,\delta,\varepsilon}, \dots, \varepsilon^{1/4} S_m^{x,\delta,\varepsilon}$ *is given by Theorem* 2.2, *i.e.*, *it is equal to the distribution of* $\sqrt{\delta}(1 + a(\delta))\sqrt{\xi} N(0, Q)$. *The limiting distribution of* $T_0^{x,\delta,\varepsilon}$ *is the distribution of* $\bar{T}^{\Gamma(x)}$. *The limiting distribution for each of the components* $T_1^{x,\delta,\varepsilon}, \dots, T_{m-1}^{x,\delta,\varepsilon}$ *is equal to the distribution of* $\bar{T}^\zeta$, *where* ζ *is a random initial point for the process on the graph*, *chosen to be at distance* δ *from the vertex* O, *in such a way that* ζ *belongs to the* i*th edge with probability* $p_i + b_i(\delta)$.

Proof. By the averaging principle [7], $T_0^{x,\delta,\varepsilon} \to \bar{T}^{\Gamma(x)}$ in distribution uniformly in $x \in \mathcal{T}$. The convergence of other components of the random vector to their respective limits follows from Theorem 2.2. The independence of the components of the limiting vector immediately follows from the strong Markov property of the process $X_t^{x,\varepsilon}$ and the fact that the convergence in Theorem 2.2 is uniform with respect to x. □

We will prove Theorem 2.2 by proving a more abstract lemma on Markov chains with a small probability of termination at each step, and demonstrating that the conditions of the lemma are satisfied in the situation of Theorem 2.2.

Let M be a metric space that can be written as a disjoint union

$$M = X \sqcup C_1 \sqcup \dots \sqcup C_n,$$

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

(1) $p_0(x, X) = 1$ for all $x \in M$ and $p_\varepsilon(x, X) = 1$ for all $x \in M \setminus X$.
(2) $p_0(x, dy)$ is weakly Feller, that is the map $x \mapsto \int_M f(y) p_0(x, dy)$ belongs to $\mathcal{C}_b(M)$ if $f \in \mathcal{C}_b(M)$.
(3) There exist bounded continuous functions $h_1, \dots, h_n : X \to [0, \infty)$ such that

$$\varepsilon^{-\frac{1}{2}} p_\varepsilon(x, C_i) \to h_i(x) \quad \text{uniformly in } x \in K \text{ if } K \subseteq X \text{ is compact},$$

while $\sup_{x \in X} |\varepsilon^{-\frac{1}{2}} p_\varepsilon(x, C_i)| \le c$ for some positive constant c. We also have

$$J(x) := h_1(x) + \dots + h_n(x) > 0 \quad \text{for } x \in X.$$

(4) $p_\varepsilon(x, dy)$ converges weakly to $p_0(x, dy)$ as $\varepsilon \to 0$, uniformly in $x \in K$ if $K \subseteq X$ is compact.
(5) The transition functions satisfy a strong Doeblin condition uniformly in ε. Namely, there exist a probability measure η on X, a constant $a > 0$, and an integer $m > 0$ such that

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

It then follows that for every ε, there is a unique invariant measure $\lambda^\varepsilon(dy)$ on 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

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

(6) The function g is such that $\int_M g \, d\lambda^\varepsilon = 0$ for each $\varepsilon \in [0, \varepsilon_0]$.

Lemma 2.4. *Suppose that assumptions* (1)–(6) *above are satisfied and let $Z_k^{x,\varepsilon}$ be the Markov chain on M starting at x, with transition function p_ε. Let $\tau = \tau(x,\varepsilon)$ be the first time when the chain reaches the set $C = C_1 \sqcup \cdots \sqcup C_n$. Let $e(Z_k^{x,\varepsilon}) = i$ if $Z_k^{x,\varepsilon} \in C_i$. Then*

$$\left(\varepsilon^{\frac{1}{4}}\left(g\left(Z_1^{x,\varepsilon}\right) + \cdots + g\left(Z_\tau^{x,\varepsilon}\right)\right), e\left(Z_\tau^{x,\varepsilon}\right)\right) \to (F_1, F_2) \tag{2.2}$$

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

$$\left(g\left(Z_1^{x,0}\right) + \cdots + g\left(Z_k^{x,0}\right)\right)/\sqrt{k} \to N(0,\bar{Q}) \quad \textit{in distribution as } k \to \infty.$$

The random variable F_2 satisfies $\mathbf{P}(F_2 = i) = \int_X h_i\,d\lambda^0 / \int_X J\,d\lambda^0$, $i = 1,\dots,n$.

Before we proceed with the proof of Lemma 2.4, let us show that it does indeed imply Theorem 2.2.

Proof of Theorem 2.2. Let $\mathcal{L}_0 = \mathcal{L} \setminus \{A \in \mathbb{R}^2 : \pi(A) \in \{A_i, i = 1,\dots,n\}\}$. Define $\bar{M} = \mathcal{L}_0 \sqcup \partial V^\delta$. Let us define a family of transition functions $\bar{p}_\varepsilon(x,dy)$ on $\bar{M}$. For $x \in \mathcal{L}_0$, we define $\bar{p}_\varepsilon(x,dy)$ as the distribution of $X_{\overline{\tau}^x}^{x,\varepsilon}$ with $\overline{\tau}^x = \mu_1^{x,\delta,\varepsilon} \wedge \beta_1^{x,\delta,\varepsilon}$. In other words, it is the measure induced by the process stopped when it either reaches the boundary of V^δ or reaches the separatrix after passing by a saddle point different from the one behind x. For $x \in \partial V^\delta$, let $\bar{p}_\varepsilon(x,dy)$ coincide with the distribution of $X_{\overline{\tau}^x}^{x,\varepsilon}$ with $\overline{\tau}^x = \beta_0^{x,\delta,\varepsilon}$, i.e., the measure induced by the process stopped when it reaches the separatrix. Since almost every trajectory of $X_t^{x,\varepsilon}$ that starts outside of the set of saddle points does not contain saddle points, $\bar{p}_\varepsilon$ is indeed a stochastic transition function. Let $\bar{Z}_k^{x,\varepsilon}$ be the corresponding Markov chain starting at $x \in \bar{M}$.

While we introduced $\bar{M}$ as a subset of $\mathbb{R}^2$, it is going to be more convenient to keep track of $\pi(\bar{Z}_k^{x,\varepsilon})$ and the latest displacement separately. Let $\varphi : \bar{M} \to M := \pi(\bar{M}) \times \mathbb{Z}^2$ map $x \in \bar{M}$ into $(\pi(x), ([x_1],[x_2]))$ ($[x_1]$ and $[x_2]$ are the integer parts of the first and second coordinates of x). Define the Markov chains $Z_k^{x,\varepsilon}$, $Z_k^{\pi(x),\varepsilon}$ on M via

$$Z_0^{x,\varepsilon} = x, \qquad Z_0^{\pi(x),\varepsilon} = (\pi(x), 0), \qquad Z_k^{x,\varepsilon} = Z_k^{\pi(x),\varepsilon} = \left(\varphi_1\left(\bar{Z}_k^{x,\varepsilon}\right), \varphi_2\left(\bar{Z}_k^{x,\varepsilon}\right) - \varphi_2\left(\bar{Z}_{k-1}^{x,\varepsilon}\right)\right), \quad k \geq 1.$$

With a slight abuse of notation, we treat x both as an element of $\bar{M}$ and an element of M. Let $X = \pi(\mathcal{L}_0) \times \mathbb{Z}^2 = (\mathcal{L}_{\mathcal{T}} \setminus \{A_1,\dots,A_n\}) \times \mathbb{Z}^2$ and $C_i = (\pi(\partial V^\delta) \cap U_i) \times \mathbb{Z}^2$. Thus $M = X \sqcup C_1 \sqcup \cdots \sqcup C_n$ as required. The transition functions $p_\varepsilon(x,dy)$ are defined as the transition functions for the Markov chain $Z_k^{x,\varepsilon}$.

For $x = (q,\xi) \in M$, define $g((q,\xi)) = \xi \in \mathbb{Z}^2$, which corresponds to 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. From the definition of the stopping times $\beta_k^{x,\delta,\varepsilon}$, it follows that $\varphi_2(\bar{Z}_k^{x,\varepsilon}) - \varphi_2(\bar{Z}_{k-1}^{x,\varepsilon})$ can only take a finite number of values (roughly speaking, the process $X_t^{x,\varepsilon}$ makes transitions from one periodicity cell to a neighboring one or to itself between the times $\beta_k^{x,\delta,\varepsilon}$ and $\beta_{k+1}^{x,\delta,\varepsilon}$). Therefore, $g(Z_k^{\pi(x),\varepsilon})$ is bounded almost surely, uniformly in x and k. Also, the function g is continuous in the product topology of $\pi(\bar{M}) \times \mathbb{Z}^2$.

The paper [11] contains some detailed results on the behavior of the process $X_t^{x,\varepsilon}$ near the separatrix. The main idea behind those results is that the process can be considered in (H,θ) coordinates in the vicinity of $\mathcal{L}$. In those coordinates, after an appropriate re-scaling, the limiting process (as $\varepsilon \to 0$) is easily identified. Note that in [11], the width of the separatrix region is of order ε^{α_1} with some $\alpha_1 \in (1/4, 1/2)$, while here it is of width δ. The results we are about to refer to can all be easily seen to hold with ε^{α_1} replaced by δ, our current case being simpler.

The existence of the limit of the transition functions p_ε in the sense of assumption (4) was justified in [11, Lemma 3.1]. This limit is denoted by p_0. An explicit formula for the density of p_0 was also provided ([11, formula (9)]), which implies that assumption (2) is satisfied. Observe that the probability of $\beta_1^{x,\delta,\varepsilon}$ being less than $\mu_1^{x,\delta,\varepsilon}$ tends to one as $\varepsilon \downarrow 0$ uniformly in $x \in \mathcal{L}$ by [11, formula (26)]. This implies property (1).

Let us sketch the proof of the Doeblin condition (5). Fix $a_1, a_2, a_3 \in \gamma(A_i, A_j) \subset \overline{U}_k$ with some A_i, A_j, and U_k. The points are ordered in the direction of the flow v. Let γ' be the part of $\gamma(A_i, A_j)$ that lies between a_2 and a_3. Let

$J = \{(H,\theta) \in V^\delta \cap U_k : \sqrt{\varepsilon} \le H \le 2\sqrt{\varepsilon}, \theta = \theta(a_1)\}$. We can assume that a_1, a_2, and a_3 are chosen in such a way that φ_2 is constant on $J \cup \gamma'$. It is not difficult to show that there is $m > 0$ such that

$$\mathbf{P}\big(\varphi_2\big(X_t^{x,\varepsilon}\big) = \varphi_2(x), X_t^{x,\varepsilon} \in J \text{ for some } \alpha_m^{x,\delta,\varepsilon} < t < \beta_m^{x,\delta,\varepsilon}\big) > c > 0$$

for all $x \in \mathcal{L}$. Roughly speaking, this statement 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 motion consists of advection with speed of order $1/\varepsilon$ and diffusion of order one. The proof follows along the same lines as the proof of Lemma 3.1. in [11]. Now the distribution of $X^{x,\varepsilon}_{\beta_0^{x,\delta,\varepsilon}}$ has a component with density strictly bounded from below on γ', uniformly in $x \in J$, as follows from (63) in [5]. This implies the Doeblin condition for $Z_k^{x,\varepsilon}$.

With our definition of g,

$$\int_M g(x)\,d\lambda^\varepsilon(x)\left(\int_M \mathbf{E}\overline{\tau}^x\,d\lambda^\varepsilon(x)\right)^{-1} = \lim_{t\to\infty}\big(\mathbf{E}X_t^{x,\varepsilon}/t\big),$$

where $\overline{\tau}^x$ 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 $\lim_{t\to\infty}(\mathbf{E}X_t^{x,\varepsilon}/t) = \varepsilon^{-1}\int_{\mathcal{T}} v(x)\,dx = 0$, which implies property (6).

Property (3) follows from [11, Lemmas 4.1 and 4.3]. Indeed, the former lemma describes the asymptotics of the distribution of $H(X^{x,\varepsilon}_{\alpha_1^{x,\delta,\varepsilon}})$, while the latter 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 strong Markov property of the process, imply Property (3). The functions $h_i(x) = h_i^\delta(x)$ depend on δ and can be identified as

$$h_i^\delta(x) = \lim_{\varepsilon\to 0}\varepsilon^{-1/2}\mathbf{P}\quad\big(\text{the process starting at } X^{x,\varepsilon}_{\alpha_1^{x,\delta,\varepsilon}} \text{ reaches } \partial V^\delta \cap U_i \text{ before reaching } \mathcal{L}\big).$$

From [11, Lemmas 4.1 and 4.3] (with δ now playing the role of ε^{α_1}) it follows that

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

where $\bar{p}_i > 0$ and $\bar{b}_i(\delta) \to 0$ as $\delta \to 0$. Now Lemma 2.4 implies that Theorem 2.2 holds with

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

Finally, let us show that $\bar{Q}$ is non-degenerate. Assuming by contradiction that this is not the case, there is a unit vector $e \in \mathbb{R}^2$ such that the function $\bar{g} = \langle e, g\rangle : X \to \mathbb{R}$ has the property that

$$\big(\bar{g}\big(Z_1^{x,0}\big) + \cdots + \bar{g}\big(Z_k^{x,0}\big)\big)/\sqrt{k} \to 0, \tag{2.3}$$

in distribution as $k \to \infty$. It follows from $\int_X \bar{g}\,d\lambda_0 = 0$ and from exponential mixing that the sum

$$G(x) = \sum_{k=0}^{\infty}\mathbf{E}\bar{g}\big(Z_k^{x,0}\big)$$

converges in $L^2(X, \lambda^0)$.

Let z_k denote the process which is $Z_k^{x,0}$ started from the invariant distribution λ_0. It follows from [3, Theorem 11], that under our assumption (2.3),

$$0 = \mathbf{E}G^2(z_k) - \mathbf{E}\big(\big[\mathbf{E}G\big(Z_1^{x,\varepsilon}\big)\big]|_{x=z_k}\big)^2. \tag{2.4}$$

By the definition of G, we have the identity

$$\bar{g}(z_k) = U_{k+1} + G(z_k) - G(z_{k+1}), \tag{2.5}$$

where $U_{k+1} = U(z_k, z_{k+1}) = G(z_{k+1}) - [\mathbf{E}G(Z_1^{x,\varepsilon})]|_{x=z_k}$. It is straightforward to see that

$$\mathbf{E}U_{k+1}^2 = \mathbf{E}G^2(z_{k+1}) - \mathbf{E}\big(\big[\mathbf{E}G\big(Z_1^{x,\varepsilon}\big)\big]|_{x=z_k}\big)^2 = 0$$

by (2.4). This implies that $U_{k+1} = 0$ almost everywhere with respect to λ_0. Combining this fact with $k = 0$ and (2.5), we get that

$$\bar{g}(x) = G(x) - G\big(Z_1^{x,0}\big),$$

almost surely for λ^0-almost all x. Recall that $x \in X$ can be written as $x = (q, \xi)$, where $q \in \pi(\mathcal{L}_0)$ and $\xi \in \mathbb{Z}^2$. Since $Z_1^{x,0}$ does not depend on ξ, while $\bar{g}(x) = \langle e, \xi\rangle$, we can write $G(x) = \tilde{G}(q) + \langle e, \xi\rangle$ for some function $\tilde{G}$. Thus

$$\tilde{G}(q) = \tilde{G}\big(\big(Z_1^{x,0}\big)^1\big) + \big\langle e, \big(Z_1^{x,0}\big)^2\big\rangle, \tag{2.6}$$

where $(Z_1^{x,0})^1 \in \pi(\mathcal{L}_0)$ and $(Z_1^{x,0})^2 \in \mathbb{Z}^2$. Thus for λ^0-almost all x, we have $\tilde{G}(q) = \tilde{G}((Z_1^{x,0})^1)$ almost surely on the event $\langle e, (Z_1^{x,0})^2\rangle = 0$. Let $\overline{\lambda}_0$ denote the projection of λ^0 onto $\pi(\mathcal{L}_0)$. An explicit expression for the density of p_0 (found in formula (9) of [11]) implies that $(Z_k^{x,0})^1$, $k \ge 1$, has density with respect to the Lebesgue measure on $\pi(\mathcal{L}_0)$, and the density is bounded from below for sufficiently large k. Therefore $\bar{\lambda}_0$ is equivalent with the Lebesgue measure and the distribution of $(Z_1^{x,0})^1$ is absolutely continuous with respect to $\overline{\lambda}^0$ for each x. Therefore, by the strong Markov property, $\tilde{G}(q) = \tilde{G}((Z_k^{x,0})^1)$ almost surely on the event $\langle e, (Z_1^{x,0})^2\rangle = \cdots = \langle e, (Z_k^{x,0})^2\rangle = 0$, for λ^0-almost all x. For sufficiently large k, the (sub-probability) distribution of $(Z_k^{x,0})^1$ restricted to this event has a positive density with respect to $\overline{\lambda}^0$. (The latter statement is a consequence of the geometry of the flow. Roughly speaking, given two points on the separatrix that belong to the same cell of periodicity, the process $\bar{Z}_k^{x,0}$ can go with positive probability from the first point to an arbitrary neighborhood of the second point without leaving the cell of periodicity.) Therefore, $\tilde{G}$ is λ_0-almost everywhere constant. By (2.6), this implies that $\langle e, (Z_1^{x,0})^2\rangle = 0$ for λ_0-almost all x. Again by the strong Markov property, $\langle e, (Z_k^{x,0})^2\rangle = 0$ for λ_0-almost all x for each k. Observe, however, that the process $\bar{Z}_k^{x,0}$ starting at an arbitrary point x on the separatrix, has a positive probability of going to any other cell of periodicity if k is sufficiently large. This yields a contradiction, and thus $\bar{Q}$ is non-degenerate. □

Now let us turn to the proof of Lemma 2.4. Let

$$\Omega = \big\{\omega = (x, x_1, \dots, x_k; i) : k \ge 0, x, x_1, \dots, x_k \in X, i \in \{1, \dots, n\}\big\}$$

be the space of sequences that start at $x \in X$ and end when the sequence enters $C = C_1 \sqcup \cdots \sqcup C_n$, at which point only the index of the set that the sequence enters is taken into account. The Markov chain $Z_k^{x,\varepsilon}$ together with the stopping time τ determine a probability measure μ_ε on Ω, namely,

$$\mu_\varepsilon(x, A_1, \dots, A_k; i) = \int_{A_1} \cdots \int_{A_k} p_\varepsilon(x, dx_1) p_\varepsilon(x_1, dx_2) \cdots p_\varepsilon(x_{k-1}, dx_k) p_\varepsilon(x_k, C_i),$$

where $A_1, \dots, A_k \in \mathcal{B}(X)$. We introduce another probability measure on Ω via

$$\begin{aligned} &\nu_\varepsilon(x, A_1, \dots, A_k, i) \\ &\quad = \int_{A_1} \cdots \int_{A_k} e^{-\sqrt{\varepsilon}(J(x)+\cdots+J(x_{k-1}))} \frac{p_\varepsilon(x, dx_1)}{p_\varepsilon(x, X)} \cdots \frac{p_\varepsilon(x_{k-1}, dx_k)}{p_\varepsilon(x_{k-1}, X)} \frac{(1 - e^{-\sqrt{\varepsilon}J(x_k)}) h_i(x_k)}{J(x_k)}. \end{aligned}$$

In other words, we consider a Markov chain $\tilde{Z}_k^{x,\varepsilon}$ on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ with state space X and with transition function $\tilde{p}_\varepsilon(x, dy) = p_\varepsilon(x, dy)/p_\varepsilon(x, X)$. We can adjoin the states $\{1, \dots, n\}$ to the space X and assume that at each step, conditioned on $\tilde{Z}_k^{x,\varepsilon}$, the process may get killed by entering a terminal state i with probability $(1 - e^{-\sqrt{\varepsilon}J(\tilde{Z}_k^{x,\varepsilon})}) \frac{h_i(\tilde{Z}_k^{x,\varepsilon})}{J(\tilde{Z}_k^{x,\varepsilon})}$, $i = 1, \dots, n$. Let σ be the number of steps after which the process is killed. To clarify our notations, let us stress that $\tilde{Z}_k^{x,\varepsilon}$ is a conservative Markov chain, and the killing is expressed through the presence of the random variable σ defined on the same probability space. Then $\nu_\varepsilon(x, A_1, \dots, A_k, i)$ is the probability that the chain starting at x visits the sets $A_1, \dots, A_k$ and then enters the terminal state i. With a slight abuse of notation we can view σ as a random variable on Ω as well.

We will prove in Lemma 2.6, that we can replace the measure μ^ε with ν^ε in a certain sense. First, however, we need to derive a few properties of $\tilde{Z}^{x,\varepsilon}$. Note that it inherits the strong Doeblin property, which holds uniformly in ε, i.e.,

$$\tilde{p}_\varepsilon^m(x, A) \geq a\eta(A) \quad \text{for } x \in X, A \in \mathcal{B}(X), \varepsilon \in [0, \varepsilon_0].$$

This implies the uniform exponential mixing, i.e., there are $\Lambda > 0, c > 0$, such that

$$\left|\tilde{p}_\varepsilon^k(x, A) - \tilde{\lambda}^\varepsilon(A)\right| \leq ce^{-\Lambda k} \quad \text{for all } x \in X, A \in \mathcal{B}(X), \varepsilon \in [0, \varepsilon_0],$$

where $\tilde{p}_\varepsilon$ is the transition function for the chain and $\tilde{\lambda}^\varepsilon$ is the invariant measure associated with the transition function $\tilde{p}_\varepsilon(x, A)$.

Lemma 2.5. *Let $g \in \mathcal{C}_b(M, \mathcal{R})$ satisfy assumption* (6). *For each $\alpha > 0$, we have*

$$\left|\int_X g\, d\tilde{\lambda}^\varepsilon\right| \leq C\varepsilon^{1/2-\alpha} \tag{2.7}$$

for some constant C and each $\varepsilon \in [0, \varepsilon_0]$.

Proof. By the exponential mixing,

$$\left|\int_X g(y)\tilde{p}_\varepsilon^k(x, dy) - \int_X g(y)\tilde{\lambda}^\varepsilon(dy)\right| + \left|\int_M g(y)p_\varepsilon^k(x, dy) - \int_M g(y)\lambda^\varepsilon(dy)\right| \leq c_1 e^{-\Lambda k}$$

for $x \in X, \varepsilon \in (0, \varepsilon_0]$. It is also easy to see by induction that

$$\left|\int_X g(y)\tilde{p}_\varepsilon^k(x, dy) - \int_M g(y)p_\varepsilon^k(x, dy)\right| \leq c_2\sqrt{\varepsilon}k. \tag{2.8}$$

Now we can take $k = [\varepsilon^{-\alpha}]$ in these two inequalities, proving (2.7) since $\int_M g(y)\lambda^\varepsilon(dy) = 0$. □

The last two inequalities of the above proof with g replaced by an arbitrary bounded continuous function f imply that

$$\int_X f(y)\tilde{\lambda}^\varepsilon(dy) - \int_M f(y)\lambda^\varepsilon(dy) \to 0 \quad \text{as } \varepsilon \downarrow 0.$$

We also know that $\lambda^\varepsilon(M \setminus X) \to 0$ and $\lambda^\varepsilon \Rightarrow \lambda^0$ as $\varepsilon \downarrow 0$, as immediately follows from the properties of p_ε (the latter statement can be also found in Lemma 2.1 in [11]). Therefore,

$$\int_X f(y)\tilde{\lambda}^\varepsilon(dy) - \int_X f(y)\lambda^0(dy) \to 0 \quad \text{as } \varepsilon \downarrow 0,$$

that is $\tilde{\lambda}_\varepsilon \Rightarrow \lambda_0$ as $\varepsilon \downarrow 0$.

Lemma 2.6. *For every $\delta > 0$ there is $\varepsilon' > 0$ such that for $\varepsilon \leq \varepsilon'$ there is a set Ω_ε with $\nu_\varepsilon(\Omega_\varepsilon) \geq 1 - \delta$ such that $d\mu_\varepsilon/d\nu_\varepsilon \in (1 - \delta, 1 + \delta)$ on Ω_ε.*

Proof. To choose the set Ω_ε, note that

$$\nu^\varepsilon(\sigma = k) = \tilde{\mathbf{E}}\big[e^{-\sqrt{\varepsilon}(J(\tilde{Z}_1^{x,\varepsilon})+\cdots+J(\tilde{Z}_{k-1}^{x,\varepsilon}))}\big(1 - e^{-\sqrt{\varepsilon}J(\tilde{Z}_k^{x,\varepsilon})}\big)\big].$$

Using the law of large numbers for the Markov chain $\tilde{Z}^{x,\varepsilon}$, which can be applied uniformly in ε due to the uniform mixing (a consequence of assumption (5)), and the boundedness of J (a consequence of assumption (3)), we conclude that for every $\eta > 0$ there is a k_0 independent of ε such that

$$\tilde{\mathbf{P}}\left(\left|\frac{1}{k}\sum_{j=0}^{k-1} J\big(\tilde{Z}_j^{x,\varepsilon}\big) - J_\varepsilon\right| \geq \eta\right) \leq \eta$$

for $k \geq k_0$, where $J_\varepsilon = \int_X J(u)\,d\lambda^\varepsilon(u)$. Therefore

$$\nu^\varepsilon(\sigma < a/\sqrt{\varepsilon}) \leq \nu^\varepsilon(\sigma < k_0) + \eta + \big(1 - e^{-\sqrt{\varepsilon}\sup_{u\in X} J(u)}\big)\sum_{k=k_0}^{[a/\sqrt{\varepsilon}]} e^{-\sqrt{\varepsilon}(kJ_\varepsilon - k\eta)}.$$

Since $J_\varepsilon \to J_0 > 0$ and since η was arbitrary, we have $\nu^\varepsilon(\sigma < a/\sqrt{\varepsilon}) < \delta/8$ (for all sufficiently small ε) if a is small enough. Similarly one can show that $\nu^\varepsilon(\sigma > b/\sqrt{\varepsilon}) < \delta/8$ if we choose b to be sufficiently large. We set $\Omega_\varepsilon^1 = \{\sqrt{\varepsilon}\sigma \in [a,b]\}$. Note that $\nu_\varepsilon(\Omega_\varepsilon^1) \geq 1 - \delta/4$. Also note that

$$\nu^\varepsilon\big(\sigma = k, h_i(x_k) < \eta; i\big) = \tilde{\mathbf{E}}\left[e^{-\sqrt{\varepsilon}\sum_{j=0}^{k-1} J(\tilde{Z}_j^{x,\varepsilon})}\big(1 - e^{-\sqrt{\varepsilon}J(\tilde{Z}_k^{x,\varepsilon})}\big)\chi_{\{h_i(\tilde{Z}_k^{x,\varepsilon})<\eta\}}\frac{h_i(\tilde{Z}_k^{x,\varepsilon})}{J(\tilde{Z}_k^{x,\varepsilon})}\right].$$

Using the inequality $x^{-1}(1 - e^{-cx}) < c$ for $x, c > 0$, this is less than or equal to $\eta\sqrt{\varepsilon}$. This means that if $\eta > 0$ is choosen small enough, then

$$\nu^\varepsilon\big(\sqrt{\varepsilon}\sigma \in [a,b], h_i(x_\sigma) < \eta; i\big) < \delta/4n \quad \text{for each } i = 1,\ldots,n.$$

We set $\Omega_\varepsilon^2 = \bigcup_{i=1}^n \{\sqrt{\varepsilon}\sigma \in [a,b], h_i(x_\sigma) < \eta; i\}$.

Fix $\gamma > 0$ to be specified later. Let $K_0 \subset X$ be a compact set such that $\lambda^0(X \setminus K_0) < \gamma/3$. This is possible by the σ-compactness of X. Take an open set $U \subseteq X$ such that $K_0 \subseteq U$ and $K = \bar{U}$ is compact, which is possible by local compactness of X. Note that $\lambda^0(X \setminus U) < \gamma/3$. By the weak law of large numbers (which holds uniformly in ε due to the uniform mixing),

$$\sup_{\varepsilon\in(0,\varepsilon_0]} \tilde{\mathbf{P}}\left(\left|\frac{1}{N}\sum_{j=0}^{N-1}\chi_{\{\tilde{Z}_j^{x,\varepsilon}\notin K\}} - \tilde{\lambda}^\varepsilon(X \setminus K)\right| > \gamma/3\right) < \delta/4$$

for large enough N. Elementary properties of weak convergence imply

$$\tilde{\lambda}^\varepsilon(X \setminus K) \leq \tilde{\lambda}^\varepsilon(X \setminus U) \leq \lambda^0(X \setminus U) + \gamma/3 < 2\gamma/3$$

for small enough ε. This means that the set

$$\Omega_\varepsilon^3 = \left\{\sqrt{\varepsilon}\sigma \in [a,b], \sum_{j=0}^{\sigma-1}\chi_{\{x_j\notin K\}} \geq \gamma\frac{2b}{a}\sigma\right\} \subseteq \left\{\sqrt{\varepsilon}\sigma \in [a,b], \sum_{j=0}^{[b/\sqrt{\varepsilon}]}\chi_{\{x_j\notin K\}} \geq \gamma\big([b/\sqrt{\varepsilon}] + 1\big)\right\}$$

has $\nu^\varepsilon(\Omega_\varepsilon^3) < \delta/4$ if ε is sufficiently small.

Similarly, by the ergodic theorem, one can show, by possibly making K larger, that $\Omega_\varepsilon^4 = \{\sqrt{\varepsilon}\sigma \in [a,b], x_\sigma \notin K\}$ has $\nu^\varepsilon(\Omega_\varepsilon^4) < \delta/4$ for sufficiently small ε. Therefore $\Omega_\varepsilon = \Omega_\varepsilon^1 \setminus (\Omega_\varepsilon^2 \cup \Omega_\varepsilon^3 \cup \Omega_\varepsilon^4)$ has $\nu^\varepsilon(\Omega_\varepsilon) > 1 - \delta$.

Observe that

$$\frac{d\mu_\varepsilon}{d\nu_\varepsilon}(x, x_1, \dots, x_k, i) = \frac{p_\varepsilon(x, X) \cdots p_\varepsilon(x_{k-1}, X)}{e^{-\sqrt{\varepsilon}(J(x)+\cdots+J(x_{k-1}))}} \frac{p_\varepsilon(x_k, C_i)}{1 - e^{-\sqrt{\varepsilon}J(x_k)}} \frac{J(x_k)}{h_i(x_k)} \quad \text{on } \Omega_\varepsilon.$$

By the definition of Ω_ε^1, it suffices to consider $k(\varepsilon) \in [a/\sqrt{\varepsilon}, b/\sqrt{\varepsilon}]$. By the definition of h_i and J, the product of the last two fractions converges to 1 uniformly as $\varepsilon \downarrow 0$ (here we use the definition of Ω_ε^2, Ω_ε^4, and assumption (3)). Also note that

$$\left| \prod_{j=0}^{k(\varepsilon)-1} p_\varepsilon(x_j, X) - e^{-\sqrt{\varepsilon}\sum_{j=0}^{k(\varepsilon)-1} J(x_j)} \right| = \left| \prod_{j=0}^{k(\varepsilon)-1} \left(1 - \sqrt{\varepsilon}\sum_{i=1}^{n} \varepsilon^{-1/2} p_\varepsilon(x_j, C_i)\right) - \prod_{j=0}^{k(\varepsilon)-1} e^{-\sqrt{\varepsilon}J(x_j)} \right|.$$

Using the fact that $|\prod a_i - \prod b_i| \le \sum |a_i - b_i|$ when $|a_i|, |b_i| \le 1$ and the boundedness part of assumption (3), this is less than or equal to

$$\sum_{j=0}^{k(\varepsilon)-1} \left| 1 - \sqrt{\varepsilon}\sum_{i=1}^{n} \varepsilon^{-1/2} p_\varepsilon(x_j, C_i) - e^{-\sqrt{\varepsilon}J(x_j)} \right| \le \sqrt{\varepsilon} \sum_{j=0}^{k(\varepsilon)-1} \sum_{i=1}^{n} \left| h_i(x_j) - \varepsilon^{-1/2} p_\varepsilon(x_j, C_i) \right| + o(1),$$

where we used the Taylor expansion of the exponential. Note that by assumption (3), we have for small enough ε that

$$\sqrt{\varepsilon} \sum_{j=0}^{k(\varepsilon)-1} \left| h_i(x_j) - \varepsilon^{-1/2} p_\varepsilon(x_j, C_i) \right| \le \frac{2bc}{k(\varepsilon)} \sum_{j=0}^{k(\varepsilon)-1} \chi_{\{x_j \notin K\}} + b\gamma < \gamma b(4bc/a + 1),$$

where the definition of Ω_ε^3 was used in the last inequality. Since γ was arbitrary and

$$\sqrt{\varepsilon} \sum_{j=0}^{k(\varepsilon)-1} J(x_j) \le nc\sqrt{\varepsilon}k(\varepsilon) \le ncb,$$

we have shown that

$$\left| \prod_{j=0}^{k(\varepsilon)-1} p_\varepsilon(x_j, X) / e^{-\sqrt{\varepsilon}\sum_{j=0}^{k(\varepsilon)-1} J(x_j)} - 1 \right| < \delta$$

for small enough ε provided that $k(\varepsilon) \in [a/\sqrt{\varepsilon}, b/\sqrt{\varepsilon}]$, which implies the desired result. □

Proof of Lemma 2.4. Using Lemma 2.6, we restate Lemma 2.4 in terms of the Markov chain $\tilde{Z}_k^{x,\varepsilon}$. Note first that we can restrict the function g (originally defined on M) to the space X at the expense that the average of g is not zero anymore, but satisfies (2.7) instead. Recall that $\bar{Q}$ is the matrix such that

$$\left(g\left(Z_1^{x,0}\right) + \cdots + g\left(Z_k^{x,0}\right)\right)/\sqrt{k} \to N(0, \bar{Q})$$

in distribution as $k \to \infty$. Let $\bar{Q}(\varepsilon)$ be such that

$$\left(g\left(\tilde{Z}_1^{x,\varepsilon}\right) + \cdots + g\left(\tilde{Z}_k^{x,\varepsilon}\right) - k\int_X g\, d\tilde{\lambda}_\varepsilon\right)/\sqrt{k} \to N\left(0, \bar{Q}(\varepsilon)\right)$$

in distribution as $k \to \infty$. From (2.8) with $k = 1$ and g replaced by an arbitrary bounded continuous function f on X it follows that $\tilde{p}_\varepsilon(x, dy) \overset{\varepsilon \to 0}{\Rightarrow} p_0(x, dy)$ uniformly in $x \in K$ for $K \subseteq X$ compact, since we assumed that the same convergence holds for $p_\varepsilon(x, dy)$. This and the strong Doeblin property for $\tilde{p}_\varepsilon(x, dy)$ easily imply that $\bar{Q}(\varepsilon) \to \bar{Q}$ as $\varepsilon \downarrow 0$ (this was proved in Lemma 2.1(c) of [11] under an additional assumption that $\int_X g\, d\tilde{\lambda}^\varepsilon = 0$, which is now replaced by (2.7)).

We still have the functions h_i defined on X, and we assume that the chain gets killed by entering the state $i \in \{1, \dots, n\}$ with probability $(1 - e^{-\sqrt{\varepsilon} J(x)}) h_i(x)/J(x)$. Let σ be the time when the chain gets killed. Let the random variable $\tilde{e}$ be equal to i if the process gets killed by entering the state i. Since the function g is bounded, omitting one last term in the sum on the left-hand side of (2.2) does not affect the limiting distribution. Now we can recast (2.2) as follows:

$$\left(\varepsilon^{\frac{1}{4}}\left(g\left(\tilde{Z}_1^{x,\varepsilon}\right) + \dots + g\left(\tilde{Z}_\sigma^{x,\varepsilon}\right)\right), \tilde{e}\right) \to (F_1, F_2)$$

in distribution. Fix $t \in \mathbb{R}^2$. For $i \in \{1, \dots, n\}$, we have that

$$\begin{aligned}&\tilde{\mathbf{E}}\left(e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{\sigma} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}; \tilde{e} = i\right) \\ &\quad = \tilde{\mathbf{E}}\left(e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{\sigma} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}; \tilde{e} = i; [a/\sqrt{\varepsilon}] \le \sigma \le [b/\sqrt{\varepsilon}]\right) + \delta(a, b, \varepsilon),\end{aligned}$$

where

$$\left|\delta(a, b, \varepsilon)\right| = \left|\tilde{\mathbf{E}}\left(e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{\sigma} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}; \tilde{e} = i; \sigma < [a/\sqrt{\varepsilon}] \text{ or } \sigma > [b/\sqrt{\varepsilon}]\right)\right| \le \nu^\varepsilon\left(\sigma < [a/\sqrt{\varepsilon}] \text{ or } \sigma > [b/\sqrt{\varepsilon}]\right),$$

which was shown in the proof of Lemma 2.6 to converge to zero as $a \to 0, b \to \infty$ uniformly in ε. Let ξ be an exponential random variable with parameter one on some probability space $(\Omega', \mathcal{F}', P')$ independent of the process. By summing over different possible values of σ,

$$\begin{aligned}&\tilde{\mathbf{E}}\left(e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{\sigma} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}; \tilde{e} = i\right) \\ &\quad = \delta(a, b, \varepsilon) \\ &\qquad + \sum_{k=[a/\sqrt{\varepsilon}]}^{[b/\sqrt{\varepsilon}]} \tilde{\mathbf{E}}\left(\frac{h_i(\tilde{Z}_k^{x,\varepsilon})}{J(\tilde{Z}_k^{x,\varepsilon})} e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{k} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle} \mathbf{P}'\left(\sqrt{\varepsilon} \sum_{j=0}^{k-1} J\left(\tilde{Z}_j^{x,\varepsilon}\right) < \xi \le \sqrt{\varepsilon} \sum_{j=0}^{k} J\left(\tilde{Z}_j^{x,\varepsilon}\right)\right)\right),\end{aligned} \tag{2.9}$$

where we used the definition of ν_ε and the fact that

$$\mathbf{P}'(c < \xi \le d) = e^{-c}\left(1 - e^{-(d-c)}\right). \tag{2.10}$$

Note that by the law of large numbers, (2.10), and the uniform exponential mixing property of $\tilde{Z}^{x,\varepsilon}$,

$$\tilde{\mathbf{E}} \sum_{k=[a/\sqrt{\varepsilon}]}^{[b/\sqrt{\varepsilon}]} \left|\mathbf{P}'\left(\sum_{j=0}^{k-1} J\left(\tilde{Z}_j^{x,\varepsilon}\right) < \frac{\xi}{\sqrt{\varepsilon}} < \sum_{j=0}^{k} J\left(\tilde{Z}_j^{x,\varepsilon}\right)\right) - \sqrt{\varepsilon} e^{-k\tilde{J}_\varepsilon \sqrt{\varepsilon}} J\left(\tilde{Z}_k^{x,\varepsilon}\right)\right| \to 0 \tag{2.11}$$

as $\varepsilon \to 0$ uniformly in $0 < a < b$, where $\tilde{J}_\varepsilon = \int_X J(u)\, d\tilde{\lambda}^\varepsilon(u) > 0$. Note that the fact that there are $\mathcal{O}(1/\sqrt{\varepsilon})$ terms in the sum is not a problem since the contribution from each term is $\mathcal{O}(\varepsilon)$. Observe that $h_i(x)/J(x) \le 1$ and therefore the factor proceeding P' on the right-hand side of (2.9) is bounded. Therefore, due to (2.11), the main term in (2.9) can be replaced by

$$\sqrt{\varepsilon} \sum_{k=[a/\sqrt{\varepsilon}]}^{[b/\sqrt{\varepsilon}]} \tilde{\mathbf{E}}\left(h_i\left(\tilde{Z}_k^{x,\varepsilon}\right) e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{k} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}\right) e^{-k\tilde{J}_\varepsilon \sqrt{\varepsilon}}. \tag{2.12}$$

Uniform exponential mixing also tells us that there is a constant C such that for every $0 < k_0 < k$ we have

$$\left|\tilde{\mathbf{E}}\left(h_i\left(\tilde{Z}_k^{x,\varepsilon}\right) e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{k-k_0} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}\right) - \tilde{\mathbf{E}}\left(h_i\left(\tilde{Z}_k^{x,\varepsilon}\right)\right) \tilde{\mathbf{E}}\left(e^{i\langle \varepsilon^{1/4} \sum_{j=1}^{k-k_0} g(\tilde{Z}_j^{x,\varepsilon}), t\rangle}\right)\right| < Ce^{-\Lambda k_0}. \tag{2.13}$$

It is easy to see that fixing $k_0 > 0$, i.e., dropping finitely many terms from the sum in the exponent in (2.12) does not change the limit (it only introduces an overall error term of order $\varepsilon^{1/4}$).

Since we have uniform exponential mixing in ε for the transition function $\tilde{p}_\varepsilon(x,dy)$ (i.e. for the process $\tilde{Z}_k^{x,\varepsilon}$), and from the fact that $\tilde{\lambda}^\varepsilon \Rightarrow \lambda^0$, it follows that

$$\sup_{k\in[[a/\sqrt{\varepsilon}],[b/\sqrt{\varepsilon}]]}\left|\tilde{\mathbf{E}}\big(h_i\big(\tilde{Z}_k^{x,\varepsilon}\big)\big)-\int_X h_i(u)\,d\lambda^0(u)\right|\to 0, \tag{2.14}$$

as $\varepsilon\downarrow 0$. Choosing $\alpha<1/4$, it follows from (2.7) that

$$\sup_{k\in[[a/\sqrt{\varepsilon}],[b/\sqrt{\varepsilon}]]}\left|\tilde{\mathbf{E}}\big(e^{i\langle\varepsilon^{1/4}\sum_{j=1}^{k-k_0}g(\tilde{Z}_j^{x,\varepsilon}),t\rangle}\big)-\tilde{\mathbf{E}}\big(e^{i\langle\varepsilon^{1/4}\sum_{j=1}^{k-k_0}(g(\tilde{Z}_j^{x,\varepsilon})-\int_X g\,d\tilde{\lambda}^\varepsilon),t\rangle}\big)\right|\to 0,$$

as $\varepsilon\downarrow 0$. On the other hand, we have the following version of the central limit theorem:

$$\sup_{k\in[[a/\sqrt{\varepsilon}],[b/\sqrt{\varepsilon}]]}\left|\tilde{\mathbf{E}}\big(e^{i\langle\varepsilon^{1/4}\sum_{j=1}^{k-k_0}(g(\tilde{Z}_j^{x,\varepsilon})-\int_X g\,d\tilde{\lambda}^\varepsilon),t\rangle}\big)-\tilde{\mathbf{E}}e^{i\langle\sqrt{k}\varepsilon^{1/4}\cdot N(0,\bar{Q}),t\rangle}\right|\to 0,$$

as $\varepsilon\downarrow 0$, which holds thanks to the uniform strong Doeblin property and the fact that $\bar{Q}(\varepsilon)\to\bar{Q}$ as $\varepsilon\downarrow 0$.

Combining this with (2.9), (2.12), (2.13), and (2.14), and using the fact that $\tilde{J}_\varepsilon\to J_0$, we obtain that

$$\limsup_{\varepsilon\downarrow 0}\left|\tilde{\mathbf{E}}\big(e^{i\langle\varepsilon^{1/4}\sum_{j=1}^{k}g(\tilde{Z}_j^{x,\varepsilon}),t\rangle};\tilde{e}=i\big)-\frac{\int_X h_i\,d\lambda^0}{\int_X J\,d\lambda_0}\int_0^\infty\tilde{\mathbf{E}}e^{i\sqrt{s}\langle N(0,\bar{Q}),t\rangle}J_0e^{-sJ_0}\,ds\right|\le ce^{-\Lambda k_0}.$$

Since t and k_0 were arbitrary, this implies the desired result. □

We close this section by stating a technical lemma that gives us control over how far away the process wanders during an upcrossing. Its proof relies on the same arguments as the proof of Lemma 2.4 considering the maximum of $\sum_{j=1}^k g(\tilde{Z}_j^{x,\varepsilon})$ until σ and using the invariance principle for Markov chains.

Lemma 2.7. *For each $\eta>0$ there is $\delta_0>0$ such that*

$$\lim_{\varepsilon\downarrow 0}\sup_{x\in\mathbb{R}^2}\mathbf{P}\Big(\varepsilon^{1/4}\sup_{0\le t\le\sigma_1^{x,\delta,\varepsilon}}\big|X_t^{x,\varepsilon}-x\big|>\eta\Big)<\eta$$

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

3. Proof of Theorem 1.2

The first step in the proof of Theorem 1.2 is to show tightness of the family of measures induced by $\varepsilon^{1/4}(X_t^{x,\varepsilon}-x)$, $0<\varepsilon\le 1$, $x\in\mathbb{R}^2$. We will then show the convergence of one-dimensional distributions. The convergence of finite-dimensional distributions (and therefore the statement of the theorem) will then follow from the Markov property.

Define $D_t^{y,\delta}$ to be the number of downcrossings from level δ to O by the trajectory of the process Y_t^y up until time t, where we start counting after the first visit to the vertex. Namely, set $\theta_0^\delta=0$, $\tau_0^\delta=\inf\{t\ge 0:Y_t^y=O\}$, and recursively define

$$\theta_n^\delta=\inf\big\{t\ge\tau_{n-1}^\delta:\big|Y_t^y\big|=\delta\big\},\qquad\tau_n^\delta=\inf\big\{t\ge\theta_n^\delta:Y_t^y=O\big\},\quad n\ge 1,$$

where $|Y_t^y|$ is the Euclidean distance of $Y^{t,y}$ from the interior vertex O. Finally, let $D_t^{y,\delta}=\sup\{n\ge 0:\tau_n^\delta\le t\}$.

Lemma 3.1. *We have*

$$\lim_{\delta\downarrow 0}\mathbf{E}\big|\delta D_t^{y,\delta}-L_t^y\big|=0$$

for each $t>0$ and $y\in G$.

The proof of this result is almost identical to [6, Lemma 2.2] , the only difference being the replacement of the condition $a(i,y)\geq c>0$ by the local integrability of $(a(i,y))^{-2}$ (and hence of $(a(i,y))^{-1}$) at the interior vertex. As already noted earlier, this is indeed the case here since our graph process arises from the averaging of a Hamiltonian, see [8, Chapter 8], so that $a^{-2}(i,y)$ only diverges logarithmically as $y\to 0$.

For the proof of tightness, we are going to need the following two simple results.

Lemma 3.2. *Let Z_i be a sequence of independent zero mean variables with a common distribution Z , such that all the moments are finite. Then there exists a universal constant C such that*

$$\mathbf{P}\Big(l^{-1/2}\max_{1\leq m\leq l}|Z_1+\cdots+Z_m|>K\Big)\leq C\frac{\mathbf{E}|Z|^{10}}{K^{10}},$$

for all $K>0$.

Proof. By taking the 10th power and using Chebyshev's inequality,

$$\mathbf{P}\Big(\max_{1\leq m\leq l}|Z_1+\cdots+Z_m|\geq K\sqrt{l}\Big)\leq\frac{1}{K^{10}l^5}\mathbf{E}\max_{1\leq m\leq l}|Z_1+\cdots+Z_m|^{10}. \tag{3.1}$$

Since the Z_i are independent centered random variables,

$$\mathbf{E}(Z_1+\cdots+Z_l)^{10}=\sum_{i_1,\dots,i_{10}=1}^{l}\mathbf{E}Z_{i_1}\cdots Z_{i_{10}}=\sum_{m_1+\cdots+m_5=10,m_i\neq 1}C(l,m_1,\dots,m_5)\mathbf{E}Z^{m_1}\cdot\ldots\cdot\mathbf{E}Z^{m_5},$$

where $|C(l,m_1,\dots,m_5)|\leq Cl^5$ for some constant $C>0$. By Holder's inequality, the sum is bounded by $Cl^5\mathbf{E}|Z|^{10}$ with a possibly different constant $C>0$. The partial sums of the Z_i's form a martingale so that, by Doob's maximal inequality,

$$\sup_{l\geq 1}\Big(l^{-5}\mathbf{E}\max_{1\leq m\leq l}|Z_1+\cdots+Z_m|^{10}\Big)\leq\left(\frac{10}{9}\right)^{10}\sup_{l\geq 1}\mathbf{E}\left|\frac{Z_1+\cdots+Z_l}{\sqrt{l}}\right|^{10}\leq C\mathbf{E}|Z|^{10}.$$

The claim now follows at once. □

Lemma 3.3. *We have* $\limsup_{t\to 0}\mathbf{E}(L_t^0/t^{1/2})^n<\infty$ *for every* $n\in\mathbb{N}$.

Proof. By Lemma 2.3 in [6] with $F(y)=|y-O|$ being the distance of $y\in G$ from the interior vertex, we get that

$$|Y_t^0|=\int_0^t a\big(i(s),Y_s^0\big)\,dW_s+\int_0^t b\big(i(s),Y_s^0\big)\,ds+L_t^0.$$

By the uniqueness of the Skorokhod-reflection, see e.g. [12, Section 3.6.C], we have the representation

$$L_t^0=\max_{0\leq s\leq t}\left(-\int_0^s a\big(i(s),Y_s^0\big)\,dW_s-\int_0^s b\big(i(s),Y_s^0\big)\,ds\right). \tag{3.2}$$

This implies that there is a standard Brownian motion B such that

$$\left(\frac{L_t^0}{t^{1/2}}\right)^n\leq C\left(\max_{0\leq s\leq t}|B_{\frac{1}{t}\int_0^s(a(i(s),Y_s^0))^2\,ds}|+t^{-1/2}\int_0^t\big|b\big(i(s),Y_s^0\big)\big|\,ds\right)^n,$$

and thus the proof is finished by noting that a and b are bounded on the graph. □

Lemma 3.4. *The family of measures induced by the processes $\{\varepsilon^{1/4}(X_t^{x,\varepsilon}-x)\}_{0<\varepsilon\leq 1,x\in\mathbb{R}^2}$ is tight.*

Proof. By the Markov property, it is sufficient to prove that for each $\eta > 0$ there are $r \in (0,1)$ and $\varepsilon_0 > 0$ such that

$$\mathbf{P}\Big(\sup_{0\le t\le r}\big|\varepsilon^{1/4}\big(X_t^{x,\varepsilon}-x\big)\big| > \eta\Big) \le r\eta, \tag{3.3}$$

for all $\varepsilon \le \varepsilon_0$ and $x \in \mathbb{R}^2$.

Take $Z = \sqrt{\xi}N(0,Q)$ and let Z_1^δ, Z_2^δ, etc. be independent identically distributed. Assume that their distribution coincides with the distribution of $\sqrt{\delta}(1+a(\delta))Z$, where $a(\delta)$ is the same as in the right-hand side of (2.1).

Applying Lemma 3.2 with $K = \eta k^{-1/2}/4$, we see that for a given $\eta > 0$, there are $k_0 \in (0,1)$ and $\delta_1 > 0$ such that

$$\mathbf{P}\Big(\max_{1\le m\le k/\delta}\big|Z_1^\delta+\cdots+Z_m^\delta\big| > \eta/4\Big) \le k^4\eta/4, \tag{3.4}$$

whenever $k \in (0,k_0)$ and $\delta \in (0,\delta_1)$. From (3.4) and Lemma 2.3, it follows that there is $\varepsilon_1(k,\delta) > 0$ such that

$$\mathbf{P}\Big(\max_{1\le m\le k/\delta}\varepsilon^{1/4}\big|S_1^{x,\delta,\varepsilon}+\cdots+S_m^{x,\delta,\varepsilon}\big| > \eta/3\Big) \le k^4\eta/3, \tag{3.5}$$

provided that $\varepsilon \le \varepsilon_1(k,\delta)$. It is not difficult to see that this estimate and those below are uniform in x. Combining (3.5) and Lemma 2.7, it now follows that there is $\varepsilon_2(k,\delta) > 0$ such that

$$\mathbf{P}\Big(\sup_{0\le t\le \sigma^{x,\delta,\varepsilon}_{[k/\delta]}}\varepsilon^{1/4}\big|X_t^{x,\varepsilon}-x\big| > \eta/2\Big) \le k^4\eta/2 \tag{3.6}$$

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

Note that by Lemma 3.1 for a given $\eta > 0$, we can find $r > 0$ and $\delta_2 = \delta_2(r) > 0$ such that

$$\sup_{y\in G}\mathbf{P}\big(D_r^{y,\delta} \ge r^{1/4}/\delta\big) < \sup_{y\in G}\mathbf{P}\big(L_r^y \ge r^{1/4}\big) + \eta r/4 \le r^2\mathbf{E}\big(L_r^0/r^{1/2}\big)^8 + \eta r/4 \le \eta r/3 \tag{3.7}$$

if $\delta \le \delta_2$, where the second inequality follows from the Chebyshev inequality and the strong Markov property, while the last inequality follows from Lemma 3.3. As a consequence of the averaging principle, we see that there is $\varepsilon_3(r,\delta)$ such that

$$\mathbf{P}\big(\sigma^{x,\delta,\varepsilon}_{[r^{1/4}/\delta]} < r\big) \le \mathbf{P}\big(D_r^{\Gamma(x),\delta} \ge r^{1/4}/\delta\big) + \eta r/6 \tag{3.8}$$

if $\varepsilon \le \varepsilon_3(r,\delta)$.

Clearly,

$$\mathbf{P}\Big(\sup_{0\le t\le r}\big|\varepsilon^{1/4}\big(X_t^{x,\varepsilon}-x\big)\big| > \eta\Big) \le \mathbf{P}\big(\sigma^{x,\delta,\varepsilon}_{[r^{1/4}/\delta]} < r\big) + \mathbf{P}\Big(\sup_{0\le t\le \sigma^{x,\delta,\varepsilon}_{[r^{1/4}/\delta]}}\varepsilon^{1/4}\big|X_t^{x,\varepsilon}-x\big| > \eta\Big)$$

so that, choosing $r > 0$ sufficiently small, combining (3.6) with $k = r^{1/4}$, (3.7), and (3.8) with $\delta < \min(\delta_1,\delta_2)$ and $\varepsilon < \min(\varepsilon_1(k,\delta),\varepsilon_2(k,\delta),\varepsilon_3(r,\delta))$, we obtain (3.3), which implies tightness. □

For the proof of convergence of one-dimensional distributions, we are going to need a lemma that is a straightforward consequence of tightness.

Lemma 3.5. *For $\eta > 0$ and $f \in \mathcal{C}_b(\mathbb{R}^2)$ uniformly continuous, we can find an $r > 0$ such that*

$$\sup_{\varepsilon\in(0,1]}\big|\mathbf{E}f\big(\varepsilon^{1/4}\big(X_{\tau''}^{x,\varepsilon}-x\big)\big) - \mathbf{E}f\big(\varepsilon^{1/4}\big(X_{\tau'}^{x,\varepsilon}-x\big)\big)\big| < \eta, \tag{3.9}$$

$$\big|\mathbf{E}f\big(\tilde{W}^Q_{\tau''}\big) - \mathbf{E}f\big(\tilde{W}^Q_{\tau'}\big)\big| < \eta \tag{3.10}$$

for each pair of stopping times $\tau' \le \tau''$ that satisfy $\mathbf{P}(\tau'' > \tau' + r) \le r$.

Proof. By the tightness result above, for each $\alpha > 0$ we can find $r > 0$ such that

$$\sup_{x\in\mathbb{R}^2} \mathbf{P}\Big(\varepsilon^{1/4} \sup_{0\le t\le r} \big|X_t^{x,\varepsilon} - x\big| > \alpha\Big) < \alpha.$$

Using that f is uniformly continuous, we can choose $\alpha(\eta)$ small enough so that we can write

$$\mathbf{E}\big|f\big(\varepsilon^{1/4}\big(X_{\tau''}^{x,\varepsilon} - x\big)\big) - f\big(\varepsilon^{1/4}\big(X_{\tau'}^{x,\varepsilon} - x\big)\big)\big| < \frac{\eta}{3} + \mathbf{P}\big(\varepsilon^{1/4}\big|X_{\tau''}^{x,\varepsilon} - X_{\tau'}^{x,\varepsilon}\big| > \alpha\big).$$

After conditioning on $X_{\tau'}^{x,\varepsilon}$ and using the strong Markov property, the second term is seen to be bounded from above by

$$\sup_{x\in\mathbb{R}^2} \mathbf{P}\Big(\varepsilon^{1/4} \sup_{0\le t\le r} \big|X_t^{x,\varepsilon} - x\big| > \alpha\Big) + \mathbf{P}\big(\tau'' - \tau' > r\big) \le \alpha + r,$$

which finishes the proof of (3.9) once α and r are chosen to be small enough. The proof of (3.10) is similar. $\square$

Let us fix $t > 0$, $f \in \mathcal{C}_b(\mathbb{R}^2)$ uniformly continuous, and $\eta > 0$. To show the convergence of one-dimensional distributions, it suffices to prove that

$$\big|\mathbf{E}f\big(\varepsilon^{1/4}\big(X_t^{x,\varepsilon} - x\big)\big) - \mathbf{E}f\big(\tilde{W}^Q_{L_t^{\Gamma(x)}}\big)\big| < \eta \tag{3.11}$$

for all sufficiently small ε. As we discussed in the introduction, the main contribution to $X_t^{x,\varepsilon}$ (found in the first term on the left-hand side of (3.11)) comes from the excursions between $\mathcal{L}$ and ∂V^δ, i.e., the upcrossings of V^δ. Also, the local time in the second term on the left-hand side of (3.11) can be related to the number of excursions (i.e., upcrossings) between the interior vertex and the set $\Gamma(\{x : |H(x)| = \delta\})$ on the graph G that happen before time t. These two observations will lead us to the proof of (3.11).

In order to choose an appropriate value for δ, we need the following lemma (a simple generalization of the CLT).

Lemma 3.6. *Suppose that N_δ are $\mathbb{N}$-valued random variables independent of the family $\{Z_i^\delta\}$ that satisfy $\mathbf{E}N_\delta \le C/\delta$ for some $C > 0$. Let $f \in \mathcal{C}_b(\mathbb{R}^2)$ and let $\tilde{W}_t^Q$ be a Brownian motion with covariance Q, independent of $\{N_\delta\}$. Then*

$$\mathbf{E}f\big(Z_1^\delta + \cdots + Z_{N_\delta}^\delta\big) - \mathbf{E}f\big(\tilde{W}^Q_{\delta N_\delta}\big) \to 0 \quad \text{as } \delta \downarrow 0.$$

Let $e^\delta(t)$ be the (random) time that elapses before the time spent by the process $Y_\cdot^y$, aside from the upcrossings, equals t, i.e.,

$$e^\delta(t) = t + \sum_{n=1}^{\infty}\big(\theta_n^\delta \wedge e^\delta(t) - \tau_{n-1}^\delta \wedge e^\delta(t)\big).$$

In other words, we stop a 'special' clock every time the process hits the vertex O, and re-start it once the process reaches the level set $\{|y| = \delta\}$. Then $e^\delta(t)$ is the actual time that elapses when the special clock reaches time t. Let $N_\delta = N_t^{y,\delta}$ be the number of upcrossings of the interval $[0, \delta]$ by the process $Y_\cdot^y$ prior to time $e^\delta(t)$.

Similarly, let $e^{\delta,\varepsilon}(t)$ be the time that elapses before the time spent by the process $X_t^{x,\varepsilon}$, aside from the upcrossings, equals t. Let $N_t^{x,\delta,\varepsilon}$ be the number of upcrossings by the process $X_t^{x,\varepsilon}$ prior to time $e^{\delta,\varepsilon}(t)$.

Lemma 3.7. *We have $e^\delta(t) \to t$ and $\delta(N_t^{y,\delta} - D_t^{y,\delta}) \to 0$ in L^1 as $\delta \downarrow 0$ for each $y \in G$.*

Proof. The first statement basically means that most of the time is spent on downcrossings rather than upcrossings. Its proof is contained in the proof of Lemma 2.2 in [6]. The second statement follows from the first one together with the strong Markov property of the process and Lemmas 3.1 and 3.3. $\square$

From Lemmas 3.7 and 3.1 it follows that the conditions of Lemma 3.6 are satisfied with our choice of N_δ. We can therefore choose $\delta_0 > 0$ such that

$$\sup_{x\in G}\left|\mathbf{E}f\left(Z_1^\delta+\cdots+Z^\delta_{N_t^{\Gamma(x),\delta}}\right)-\mathbf{E}f\left(\tilde{W}^Q_{\delta N_t^{\Gamma(x),\delta}}\right)\right|\le \eta/10 \tag{3.12}$$

whenever $\delta\le\delta_0$.

Choose r such that (3.9) and (3.10) in Lemma 3.5 hold with $\eta/10$ instead of η. Also, use Lemma 3.1 and Lemma 3.7 to choose $\delta<\delta_0$ sufficiently small so that

$$\left|\mathbf{E}f\left(\tilde{W}^Q_{\delta D_t^{\Gamma(x),\delta}}\right)-\mathbf{E}f\left(\tilde{W}^Q_{L_t^{\Gamma(x)}}\right)\right|<\eta/10 \tag{3.13}$$

and

$$\mathbf{P}\left(\delta N_t^{\Gamma(x),\delta}>\delta D_t^{\Gamma(x),\delta}+r\right)\le r,\qquad \mathbf{P}\left(e^\delta(t)>t+r\right)\le r/2.$$

From the weak convergence of the processes, the latter implies that there is $\varepsilon_0>0$ such that

$$\mathbf{P}\left(e^{\delta,\varepsilon}(t)>t+r\right)\le r$$

for $\varepsilon<\varepsilon_0$. By Lemma 3.5, these inequalities imply that

$$\left|\mathbf{E}f\left(\varepsilon^{1/4}\left(X^{x,\varepsilon}_{e^{\delta,\varepsilon}(t)}-x\right)\right)-\mathbf{E}f\left(\varepsilon^{1/4}\left(X_t^{x,\varepsilon}-x\right)\right)\right|<\eta/10, \tag{3.14}$$

and

$$\left|\mathbf{E}f\left(\tilde{W}^Q_{\delta N_t^{\Gamma(x),\delta}}\right)-\mathbf{E}f\left(\tilde{W}^Q_{\delta D_t^{\Gamma(x),\delta}}\right)\right|<\eta/10. \tag{3.15}$$

In what follows δ is fixed at this value.

Choose N large enough so that

$$\left|\mathbf{E}f\left(Z_1^\delta+\cdots+Z^\delta_{N_t^{\Gamma(x),\delta}}\right)-\mathbf{E}f\left(Z_1^\delta+\cdots+Z^\delta_{N_t^{\Gamma(x),\delta}\wedge N}\right)\right|<\eta/10 \tag{3.16}$$

and by possibly increasing N, let $\varepsilon_1>0$ be such that

$$\left|\mathbf{E}f\left(\varepsilon^{1/4}\left(X^{x,\varepsilon}_{e^{\delta,\varepsilon}(t)}-x\right)\right)-\mathbf{E}f\left(\varepsilon^{1/4}\left(X^{x,\varepsilon}_{e^{\delta,\varepsilon}(t)\wedge\sigma_N^{x,\delta,\varepsilon}}-x\right)\right)\right|<\eta/10 \tag{3.17}$$

for all $\varepsilon\le\varepsilon_1$. This latter can be done by noting that by Lemma 2.3, for every α one can select an N such that

$$\mathbf{P}\left(\sigma_N^{x,\delta,\varepsilon}\le e^{\delta,\varepsilon}(t)\right)<\alpha \tag{3.18}$$

for every small enough ε. Indeed,

$$\mathbf{P}\left(\sigma_N^{x,\delta,\varepsilon}\le e^{\delta,\varepsilon}(t)\right)=\mathbf{P}\left(T_0^{x,\delta,\varepsilon}+\cdots+T_N^{x,\delta,\varepsilon}\le t\right).$$

For fixed N and δ, the random variable $T_0^{x,\delta,\varepsilon}+\cdots+T_N^{x,\delta,\varepsilon}$ converges in distribution to some random variable $\tilde{\tau}_N^\delta$ as $\varepsilon\downarrow 0$. Choose N large enough so that

$$\mathbf{P}\left(\tilde{\tau}_N^\delta\le t\right)<\alpha/2,$$

which implies (3.18). Now we have both N and δ fixed.

By Lemma 2.3, there is $\varepsilon_2(\delta)>0$ such that

$$\left|\mathbf{E}f\left(\varepsilon^{1/4}\left(S_1^{x,\delta,\varepsilon}+\cdots+S^{x,\delta,\varepsilon}_{N_t^{\Gamma(x),\delta,\varepsilon}\wedge N}\right)\right)-\mathbf{E}f\left(Z_1^\delta+\cdots+Z^\delta_{N_t^{\Gamma(x),\delta}\wedge N}\right)\right|<\eta/10 \tag{3.19}$$

if $\varepsilon \le \varepsilon_2$. It is here where we used the fact that the displacements during upcrossings become independent, in the limit of $\varepsilon \downarrow 0$, from the times spent on downcrossings. We also have that there is an $\varepsilon_3 > 0$ such that

$$\big|\mathbf{E} f\big(\varepsilon^{1/4}\big(S_1^{x,\delta,\varepsilon} + \cdots + S^{x,\delta,\varepsilon}_{N_t^{\Gamma(x),\delta,\varepsilon}\wedge N}\big)\big) - \mathbf{E} f\big(\varepsilon^{1/4}\big(X^{x,\varepsilon}_{e^{\delta,\varepsilon}(t)\wedge\sigma_N^{x,\delta,\varepsilon}} - x\big)\big)\big| < \eta/10 \tag{3.20}$$

for all $\varepsilon < \varepsilon_3$.

Collecting (3.14), (3.17), (3.20), (3.19), (3.16), (3.12), (3.15) and (3.13), we obtain (3.11) for $\varepsilon \le \min\{\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3\}$, which completes the proof of Theorem 1.2.

Remark 3.8. *It is not difficult to show (and it indeed follows from the proof) that convergence in Theorem 1.2 is uniform in $x \in K$ for every compact K.*

4. Proofs of the PDE results

Proof of Theorem 1.4. *Part* 1. By the representation formula,

$$u^{\varepsilon,R}(x) = \mathbf{E}\int_0^{\tau_{\partial D_R}(X_\cdot^{x,\varepsilon})} f\big(X_s^{x,\varepsilon}/R\big)\,ds,$$

which can be decomposed as

$$\mathbf{E}\int_0^{\tau_{\mathcal{L}}(X_\cdot^{x,\varepsilon})} f\big(X_s^{x,\varepsilon}/R\big)\,ds + \mathbf{E}\int_{\tau_{\mathcal{L}}(X_\cdot^{x,\varepsilon})}^{\tau_{\partial D_R}} f\big(X_s^{x,\varepsilon}/R\big)\,ds,$$

where $\tau_{\mathcal{L}}$ is the first time the process hits the separatrix. The first term can easily be seen to converge by the averaging theorem to $f(0)\mathbf{E}\bar{\tau}_0(Y_\cdot^{\Gamma(x)})$, and thus it remains to show that the second term converges to zero. It suffices to show that $\mathbf{E}(\tau_{\partial D_R}(X_\cdot^{x,\varepsilon}) - \tau_{\mathcal{L}}(X_\cdot^{x,\varepsilon})) \to 0$ as $\varepsilon \to 0$.

Let $\mathcal{T}$ be the periodicity cell that contains the origin. Recall that $\mathcal{L}_{\mathcal{T}}$ is the projection of $\mathcal{L}$ on the torus. Equivalently, we can view it as a set on the plane that is the intersection of $\mathcal{L}$ and $\mathcal{T}$. Thus it is sufficient to show that

$$\sup_{x\in\mathcal{L}_{\mathcal{T}}} \mathbf{E}\tau_{\partial D_R}\big(X_\cdot^{x,\varepsilon}\big) \to 0 \quad \text{as } \varepsilon \downarrow 0,\ R = R(\varepsilon). \tag{4.1}$$

We claim that

$$\sup_{x\in\mathcal{L}_{\mathcal{T}}} \mathbf{P}\big(\tau_{\partial D_R}\big(X_\cdot^{x,\varepsilon}\big) > K\big) \to 0 \quad \text{as } \varepsilon \downarrow 0,\ R = R(\varepsilon) \tag{4.2}$$

for each $K > 0$, and that there is $\varepsilon_0 > 0$ such that

$$\sup_{\varepsilon\in(0,\varepsilon_0]} \sup_{x\in\mathbb{R}^2} \mathbf{P}\big(\tau_{\mathcal{L}}\big(X_\cdot^{x,\varepsilon}\big) > 1\big) < 1. \tag{4.3}$$

The latter easily follows from the averaging principle (see [8], Chapter 8), while the former will be justified below.

Note that

$$\sup_{x\in\mathcal{L}_{\mathcal{T}}} \mathbf{E}\tau_{\partial D_R}\big(X_\cdot^{x,\varepsilon}\big) \le \int_0^\infty \sup_{x\in\mathcal{L}_{\mathcal{T}}} \mathbf{P}\big(\tau_{\partial D_R}\big(X_\cdot^{x,\varepsilon}\big) > K\big)\,dK.$$

By (4.2), the integrand tends to zero for each K. Also note that the integrand decays exponentially in K uniformly in ε, as follows from (4.2), (4.3), and the strong Markov property of the process. This justifies (4.1).

We still need to prove (4.2). For a given value of $\delta > 0$ and all sufficiently small ε, we have

$$\tau_{\partial D_R}\big(X_\cdot^{x,\varepsilon}\big) \le \tau_{B(0,\delta)}\big(\varepsilon^{1/4} X_\cdot^{x,\varepsilon}\big),$$

where $\tau_{B(0,\delta)}$ is the time to reach the boundary of the ball of radius δ centered at the origin. By Theorem 1.2,

$$\mathbf{P}\big(\tau_{B(0,\delta)}\big(\varepsilon^{1/4}X^{x,\varepsilon}_{\cdot}\big) > K\big) \to \mathbf{P}\big(\tau_{B(0,\delta)}\big(\tilde{W}^{Q}_{L^0_{\cdot}}\big) > K\big) \quad \text{as } \varepsilon \downarrow 0,$$

since the boundary of the event on the right-hand side has probability zero. It remains to note that we can make the right-hand side arbitrarily small by choosing a sufficiently small δ. This is possible since $\mathbf{P}(L^0_t > 0) = 1$ for each $t > 0$ (as follows from (3.2) and the elementary properties of the Brownian motion).

Part 2. Let's first assume that $f \geq 0$. Observe that for each $t > 0$ we have

$$\mathbf{E}\int_0^{\tau_{\partial D_R}(X^{x,\varepsilon}_{\cdot})\wedge t} f\big(X^{x,\varepsilon}_s/R\big)\,ds = \mathbf{E}\int_0^{\tau_{\partial D}(R^{-1}X^{x,\varepsilon}_{\cdot})\wedge t} f\big(R^{-1}X^{x,\varepsilon}_s\big)\,ds =: \mathbf{E}I^t_f\big(R^{-1}X^{x,\varepsilon}_{\cdot}\big).$$

By Theorem 1.2, the processes $R^{-1}X^{x,\varepsilon}_{\cdot}$ converge weakly to $C^{-1}W^{Q}_{L^{\Gamma(x)}_{\cdot}}$. Since I^t_f is bounded and is continuous almost surely with respect to the measure induced by $C^{-1}W^{Q}_{L^{\Gamma(x)}_{\cdot}}$, we have

$$\mathbf{E}\int_0^{\tau_{\partial D_R}(X^{x,\varepsilon}_{\cdot})\wedge t} f\big(X^{x,\varepsilon}_s/R\big)\,ds \to \mathbf{E}\int_0^{\tau_{\partial D}(C^{-1}W^{Q}_{L^{\Gamma(x)}_{\cdot}})\wedge t} f\big(C^{-1}W^{Q}_{L^{\Gamma(x)}_s}\big)\,ds \quad \text{as } \varepsilon \downarrow 0. \tag{4.4}$$

As in the proof of Part 1, we have that $\mathbf{P}(\tau_{\partial D_R}(X^{x,\varepsilon}_{\cdot}) > K)$ decays exponentially in K uniformly in ε, which justifies the fact that we can take $t = \infty$ in (4.4). The general case follows by taking $f = f_+ - f_-$.

Part 3. The PDE result easily follows from the weak convergence of the corresponding processes. More precisely, let $\bar{X}^{x,\varepsilon}_t = R^{-1}(\varepsilon)X^{x,\varepsilon}_{\varepsilon^{1/2}R(\varepsilon)^2 t}$. We need to show that

$$\bar{X}^{x,\varepsilon}_{\cdot} \Rightarrow \tilde{W}^{cQ}_{\cdot} \quad \text{as } \varepsilon \downarrow 0. \tag{4.5}$$

It follows from [11] that

$$\frac{\varepsilon^{1/4}X^{x,\varepsilon}_{k\cdot}}{\sqrt{k}} \quad \Rightarrow \quad \tilde{W}^{D(\varepsilon)}_{\cdot} \quad \text{as } k \to \infty, \tag{4.6}$$

where $D(\varepsilon) = D_0 + o(1)$ and D_0 is a constant multiple of Q. (Strictly speaking, the result in [11] concerns the finite dimensional distributions, but the generalization to the functional CLT is standard in this situation.) Moreover, it is not difficult to show (by following the proof in [11] and using arguments similar to those in the proof of Lemma 2.4) that the convergence is uniform in ε. Therefore, (4.6) implies (4.5) with $cQ = D_0$. □

Acknowledgements

The authors are grateful to D. Dolgopyat, G. Iyer, and A. Novikov for various helpful suggestions. We would also like to thank the referees for reading the paper very carefully and suggesting many important corrections. While working on the paper, L. Koralov was partially supported by the Simons Fellowship in Mathematical Sciences as well as the NSF grant number 1309084. Z. Pajor-Gyulai was partially supported by the NSF grant number 1309084. M. Hairer was partially supported by the Royal Society and by the Leverhulme Trust.

References

[1] G. Ben Arous and J. Černý. Scaling limit for trap models on $\mathbb{Z}^d$. *Ann. Probab.* **35** (6) (2007) 2356–2384. MR2353391

[2] Y. Bakhtin. Noisy heteroclinic networks. *Probab. Theory Related Fields* **150** (2011) 1–42. MR2800902

[3] E. Bolthausen and I. Goldsheid. Lingering random walks in random environment on a strip. *Comm. Math. Phys.* **278** (1) (2008) 253–288. MR2367205

[4] A. Bensoussan, J. Lions and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures*. AMS Chelsea Publishing, Providence, RI, 2011. MR2839402 (Corrected reprint of the 1978 original *Asymptotic Analysis for Periodic Structures. Studies in Mathematics and Its Applications* **5**. North-Holland Publishing Co., Amsterdam. MR0503330)

[5] D. Dolgopyat and L. Koralov. Averaging of Hamiltonian flows with an ergodic component. *Ann. Probab.* **36** (6) (2008) 1999–2049. MR2478675

[6] M. Freidlin and S.-J. Sheu. Diffusion processes on graphs: Stochastic differential equations, large deviation principle. *Probab. Theory Related Fields* **116** (2) (2000) 181–220. MR1743769

[7] M. I. Freidlin and A. D. Wentzell. Diffusion processes on graphs and the averaging principle. *Ann. Probab.* **21** (4) (1993) 2215–2245. MR1245308

[8] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*, 3rd edition. *Grundlehren der Mathematischen Wissenschaften* **260**. Springer, Heidelberg, 2012. MR2953753

[9] G. Iyer, T. Komorowski, A. Novikov and L. Ryzhik. From homogenization to averaging in cellular flows. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **5** (2014) 957–983. MR3258362

[10] G. Iyer and K. C. Zygalakis. Numerical studies of homogenization under a fast cellular flow. *Multiscale Model. Simul.* **10** (3) (2012) 1046–1058. MR3022031

[11] L. Koralov. Random perturbations of 2-dimensional Hamiltonian flows. *Probab. Theory Related Fields* **129** (1) (2004) 37–62. MR2052862

[12] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, 2nd edition. *Graduate Texts in Mathematics* **53**. Springer, New York, 1991. MR1121940

[13] G. Pavliotis and A. Stuart. *Multiscale Methods: Averaging and Homogenization*. *Texts in Applied Mathematics*. Springer, New York, 2008. MR2382139

[14] G. Pavliotis, A. Stuart and K. Zygalakis. Calculating effective diffusivities in the limit of vanishing molecular diffusion. *J. Comput. Phys.* **228** (4) (2009) 1030–1055. MR2489942

[15] V. Zhikov, S. Kozlov and O. Oleĭnik. *Homogenization of Differential Operators and Integral Functionals*. Springer, Berlin, 1994. MR1329546