TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 368, Number 9, September 2016, Pages 6487–6517
http://dx.doi.org/10.1090/tran/6574
Article electronically published on December 22, 2015

SCALING LIMITS FOR CONDITIONAL DIFFUSION EXIT PROBLEMS AND ASYMPTOTICS FOR NONLINEAR ELLIPTIC EQUATIONS

YURI BAKHTIN AND ANDRZEJ ŚWIĘCH

Abstract. The goal of this paper is to supplement the large deviation principle of the Freidlin–Wentzell theory on exit problems for diffusion processes with results of classical central limit theorem type. Namely, we describe a class of situations where conditioning on exit through unlikely locations leads to a Gaussian scaling limit for the exit distribution. Our results are based on Doob's h-transform and new asymptotic convergence gradient estimates for elliptic nonlinear equations that allow one to reduce the problem to the Levinson case. We devote an appendix to a rigorous and general discussion of h-transform.

1. Introduction

The study of dynamical systems under small noise perturbations has a long history. Although the limiting behavior of the resulting stochastic dynamics as the noise intensity vanishes depends very much on the character of the system, in many cases the key features can be described in terms of the celebrated Freidlin–Wentzell theory (FW); see [20].

Exit problems associated with these diffusions provide important information that can be used to study transitions of the system between various regions in the phase space. FW theory provides the asymptotic description of the exit distributions at the level of large deviations estimates. In several important situations, besides finding the points where the exit distribution concentrates as the noise intensity vanishes, it also produces exact exponential concentration rates. A key notion in FW is that of quasi-potential, a rate of unlikelihood or cost of diffusing from one point to another. One can often say that the exit distribution concentrates near minima of quasi-potential in the small noise limit, although concrete results require more careful statements.

It would be interesting to supplement the large deviation principle results of FW with a result of classical central limit theorem type, i.e., a Gaussian scaling limit.

Although the idea of obtaining such a result is not new — the possibility of such a result in the large deviation framework was hinted at in the original book [20] — no rigorous results are known to us. However, such results may have important

Received by the editors October 22, 2013 and, in revised form, August 25, 2014.

2010 *Mathematics Subject Classification.* Primary 60J60, 35J15, 35F21.

Key words and phrases. Diffusion, exit problems, scaling limit, small noise, Doob's h-transform, Hamilton–Jacobi–Bellman equation, elliptic PDE, viscosity solution, region of strong regularity.

The first author was partially supported by NSF via DMS-1407497 and CAREER DMS-0742424. The second author was partially supported by NSF grant DMS-0856485.

consequences. For example, in [1,3,4], the scaling limits for exit distribution played the crucial role in the analysis of small noise limit for noisy heteroclinic networks. In that context, in the exit problem near a saddle point of the drift vector field, the quasi-potential approach does not produce enough detail. At the same time, the corresponding scaling limits capture the difference between asymptotically symmetric and asymmetric exit distributions responsible for Markov or non-Markov behavior of the limiting process jumping between saddles along heteroclinic connections.

FW theory mainly concerns the situation where for any finite time horizon the exit from a domain O gets extremely unlikely as the noise intensity ε tends to zero. Such situations arise if the unperturbed dynamical system has an attractor or several attractors in O. Then the system spends a long time in a small neighborhood of an attractor and rarely makes excursions away from it as the drift brings it back since the noise is small. After a long time and multiple escape attempts, a larger fluctuation inevitably occurs and makes one of these attempts succeed, so the system either exits the domain or approaches one of the other attractors and the same scenario repeats there. The arising phenomenon of noise induced rare jumps between attractors is often called metastability.

In this setting, the reactive trajectories or transition paths usually travel through a small neighborhood of a saddle critical point of the drift, and the location of the exit from that neighborhood may critically impact the remainder of the transition path by influencing the choice of the next attractor visited by the system. So, obtaining distributional scaling limits for such situations would be highly desirable as it may lead to non-Markovian effects of asymmetrical decisions analogous to [1, 3, 4].

We are not ready to make any mathematical claims concerning this important case. Instead we consider a simpler situation. Although transition or exit without immediate return to the neighborhood of the attractor is a rare event, one can study the system conditioned on this rare event. The goal of this paper is to describe a class of situations where conditioning on exit through unlikely locations leads to a Gaussian scaling limit for the exit distribution.

Let us briefly describe our program.

The first step is an application of Doob's h-transform from which it follows that a diffusion conditioned on an exit event is again a diffusion process with the original diffusion coefficient and a new drift that is obtained from the original drift by a nonlinear correction that depends on ε. The h-transform is a well-known tool (see, e.g., a recent work [27] that uses h-transform to support the theory of transition paths [10, 11, 29]). A self-contained discussion of h-transform is given in Appendix A.

The second step is to establish sufficiently fast convergence of the corrected drift to a limiting vector field as $\varepsilon \to 0$. The idea is that the convergence in question can often be studied with the help of an asymptotic analysis of a stationary Hamilton–Jacobi–Bellman (HJB) equation with positive viscosity ε.

In fact, a large part of this paper is a discussion of a class of such situations where solutions of viscous HJB equations and their gradients converge in domains of regularity, as $\varepsilon \to 0$, to viscosity solutions of the inviscid HJB equations and, moreover, an expansion for the solution with the leading correction term of the order of ε^2 holds true in C^1 norm. Such expansions are known in C^1 for evolutionary HJB equations (see [15]) and in C^0 for stationary ones (see [19]). In both papers the equations have zero boundary conditions and the authors use an analogue of Lemma 5.8. An alternative approach for proving asymptotic series expansion in the C^0 norm without the need to show convergence of gradients (i.e. Lemma 5.8) was developed in [32] and later used in [17] for finite horizon exit problems. This approach is also described in [18], Section VI.7. Since no results on C^1 expansions in the stationary case with our boundary condition are known to us, we prove a new theorem in this direction. We have to impose certain restrictions on the problem, namely, we assume that the noise is additive and the boundary is flat. Our analysis is based on PDE methods and involves elements of stochastic control. It is worth mentioning that it is the infinite time horizon in the corresponding stochastic control problems that makes the analysis of stationary HJB equations harder than that of evolutionary ones, where the time horizon is finite. Let us also mention the connection between the FW quasi-potential, the rate function in the large deviation principle for the conditional exit point, and the viscosity solution of the stationary HJB equation (see [16] for a more PDE and stochastic control viewpoint than [20]).

The third step in our program is to check that the limiting drift satisfies the Levinson condition. If it does, then the CLT for conditioned diffusion follows from the exit CLT for diffusions in the Levinson case obtained in [2].

We will describe the setting, explain our program in more detail, and formulate the main results in Section 2. In Section 3 we show a very simple example where our program can be carried out by explicit computations. In Section 4 we explain the connection with HJB equations and state relevant results on asymptotics of solutions of HJB equations guaranteeing the CLT of Section 2. Proofs of these results are given in Section 5. Appendix A is devoted to Doob's h-transform.

2. Main results

Let us now be more precise. We begin with the deterministic dynamics in $\mathbb{R}^n$, $n \in \mathbb{N}$, defined by a C^2 vector field $b(x) = (b^1(x), \dots, b^n(x)), x \in \mathbb{R}^n$:

$$\dot{X}(t) = b(X(t)).$$

For $\varepsilon > 0$, we consider an elliptic stochastic perturbation of this system given by the following Itô equation:

$$dX_\varepsilon(t) = b(X_\varepsilon(t))dt + \varepsilon\sigma(X_\varepsilon(t))dW(t). \tag{1}$$

Here $W = (W^1, \dots, W^m)$, $m \in \mathbb{N}$, is a standard m-dimensional Wiener process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ satisfying the usual conditions, and $\sigma(x)=(\sigma^i_j(x))$ is a C^2 $n \times m$ matrix-valued diffusion coefficient function such that the C^2 $n \times n$ matrix-valued function $a(x) = \sigma(x)\sigma^*(x)$ is nondegenerate for all x.

Let us assume for simplicity (although one can work without this assumption) that solutions of SDE (1) are nonexplosive, i.e., for any initial condition $X_\varepsilon(0) = x_0$, the solution process $X_\varepsilon(t)$ is defined for all times $t \geq 0$ with probability 1. These solutions can be described as continuous Markov processes with generator A_ε whose action on smooth functions f with compact support is given by

$$A_\varepsilon f(x) = \langle b(x), Df(x)\rangle + \frac{\varepsilon^2}{2}\operatorname{Tr}(a(x)D^2 f(x)) \\ = \sum_{i=1}^n b^i(x)\partial_i f(x) + \frac{\varepsilon^2}{2}\sum_{i,k=1}^n a^{ik}(x)\partial_{ik} f(x), \quad x \in \mathbb{R}^n,$$

where $Df(x) = (\partial_1 f(x), \dots, \partial_n f(x))$, $D^2 f(x) = (\partial_{ik} f(x))_{i,k=1}^n$, and the angular brackets denote the standard inner product in $\mathbb{R}^n$. The distribution of the process X_ε started at time 0 at a point x will be denoted by $\mathsf{P}_{\varepsilon,x}$.

We will observe the process X_ε only while it evolves within a domain $O \subset \mathbb{R}^n$. In this section, the domain O does not have to be bounded, and its boundary ∂O is not required to be smooth. However, ∂O is required to contain a C^2 hypersurface M such that $O \cup M$ is a path-connected set.

If $X_\varepsilon(0) = x \in O$, we define $\tau_\varepsilon = \inf\{t \geq 0 : X_\varepsilon(t) \in \partial O\} \in (0, \infty]$. We are interested in the distribution of the exit point $X_\varepsilon(\tau_\varepsilon)$ conditioned on the event $C_\Gamma = \{\tau_\varepsilon < \infty, X_\varepsilon(\tau_\varepsilon) \in \Gamma\}$, where Γ is a subset of ∂O containing M.

The assumptions that we have made guarantee that for any $\varepsilon > 0$ and any $x \in O$, under $\mathsf{P}_{\varepsilon,x}$ the diffusion X_ε conditioned on C_Γ and stopped at ∂O is also a diffusion process. The generator $A_{\Gamma,\varepsilon}$ of the conditioned diffusion is well defined on every $f \in C_0^2(O)$ (which we understand to be defined on the whole of $\mathbb{R}^n$) and is given by

$$A_{\Gamma,\varepsilon} f(x) = A_\varepsilon f(x) + \frac{\varepsilon^2}{h^\varepsilon(x)} \sum_{i,j=1}^n a^{ij}(x)\partial_j h^\varepsilon(x)\partial_i f(x), \quad x \in O, \tag{2}$$

where h^ε, defined by

$$h^\varepsilon(x) = \mathsf{P}_{\varepsilon,x}(C_\Gamma), \quad x \in \bar{O},$$

is C^2 and strictly positive in O for all $\varepsilon > 0$. In other words, the conditioned diffusion has the diffusion coefficients of the original unconditioned diffusion, but the drift coefficient $\bar{b}_\varepsilon$ of the conditioned diffusion is given by

$$\bar{b}_\varepsilon(x) = b(x) + \varepsilon^2 a(x)\frac{Dh^\varepsilon(x)}{h^\varepsilon(x)}, \quad x \in O. \tag{3}$$

Formulas (2) and (3) can be viewed as specific cases of Doob's h-transform. Although they are well known and valid under very mild assumptions, no rigorous and complete exposition on Doob's h-transform for conditioned diffusions is known to us. Appendix A aims to be such an exposition containing some relevant results that are rigorous and general. Specifically, Lemma A.3 implies that h^ε is differentiable (so (2) and (3) make sense), and Theorem A.6 implies the rest of the claims made in the last paragraph.

Let us now study the effect of some natural assumptions on the limiting behavior of the drift $\bar{b}_\varepsilon$ of the conditioned diffusion introduced in (3).

Let us suppose that there is an open set G, a point $x_0 \in O \cap G$, and a vector field $\bar{b}_0 \in C^2(G)$ satisfying the following properties:

(A1) Let $(S^t)_{t\geq 0}$ denote the flow generated by $\bar{b}_0$. We assume that there are $T > 0$ and $z \in M \cap G$ such that
  (i) $S^t x_0 \in O \cap G$ for all $t \in [0, T)$;
  (ii) $z = S^T x_0$;
  (iii) $\bar{b}_0$ is transversal to M at z, i.e., $\bar{b}_0(z)$ does not belong to the tangent hyperplane $T_z M$.

(A2) We assume that $\bar{b}_\varepsilon$ admits a C^2 continuation onto G and there are positive constants C and α such that for $\varepsilon \in (0,1)$,

$$\left|\bar{b}_\varepsilon(x) - \bar{b}_0(x)\right| \le C\varepsilon^{1+\alpha}, \quad x \in G.$$

The first of these properties is called the Levinson condition. It describes the mutual geometry of the vector field $\bar{b}_0$ and the domain O. The second property also involves $\bar{b}_\varepsilon$ and states that it converges to $\bar{b}_0$ uniformly and sufficiently fast.

Assumptions (A1) and (A2) are imposed so that we can directly apply the main result of [2] that states that if in the Levinson case the deterministic perturbation of the drift converges to zero faster than the stochastic perturbation and the perturbation of the initial condition (the latter is identically zero in our situation, although one can easily consider nonzero perturbations of initial conditions as well), then, under appropriate rescaling, the distribution of the exit location is asymptotically Gaussian. Moreover, an explicit expression for the latter is available. We now introduce the relevant notation and state one of our main results.

Let I be the $n \times n$ identity matrix. Let $\Phi(t)$ be the linearization of the flow S along the orbit of x_0:

$$\dot{\Phi}(t) = D\bar{b}_0(S^t x_0)\Phi(t), \quad \Phi(0) = I.$$

Also, for any vector $v \in \mathbb{R}^n$, we introduce its projections $\pi_b v \in \mathbb{R}$ and $\pi_M v \in T_z M$ by

$$v = \pi_b v \cdot \bar{b}_0(z) + \pi_M v.$$

The following is a direct consequence of Theorem A.6 on h-transform and the main result of [2].

Theorem 2.1. *Suppose* (A1) *and* (A2) *hold for some choice of* G, x_0, *and* $\bar{b}_0$. *Let* $X_\varepsilon(0) = x_0$ *for all* $\varepsilon > 0$. *Then the conditional distribution of* $\varepsilon^{-1}(\tau_\varepsilon - T, X_\varepsilon(\tau_\varepsilon) - z)$ *given* C_Γ *converges weakly* (*as* $\varepsilon \to 0$), *to a nondegenerate* n-*dimensional Gaussian distribution given by* $(-\pi_b\phi, \pi_M\phi)$, *where*

$$\phi = \Phi(T)\int_0^T \Phi^{-1}(t)\sigma(S^t x_0)dW(t).$$

Although this is our central result, we need to explain why assumptions (A1) and (A2) hold true in a large class of cases. Of course, if $h^\varepsilon \equiv 1$, then the conditioning is trivial and if the original vector field satisfies the Levinson condition, then Theorem 2.1 applies. We are interested though in conditioning on unlikely events, and we begin with a simple guiding example.

3. Example

In this short section we consider a simple guiding example where all calculations can be done explicitly. In this example, $O = \{x = (x_1, x_2) \in \mathbb{R}^2 : \; x_1 > 0\}$, $b(x) \equiv (b^1, b^2)$ is a constant vector field with $b^1 > 0$, and $a(x)$ is the identity matrix for all x. In other words, this is Brownian motion in a half-plane, with drift directed away from the boundary of the half-plane. It can be described by the following stochastic equation with additive noise:

$$dX^1_\varepsilon(t) = b^1 dt + \varepsilon dW^1(t),$$
$$dX^2_\varepsilon(t) = b^2 dt + \varepsilon dW^2(t).$$

For any initial condition $x \in O$, the process X_ε reaches $\partial O = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = 0\}$ with positive probability $h^\varepsilon(x_1, x_2)$. Due to the translational invariance of the system along the x_2-axis, this probability depends only on x_1. We have $h^\varepsilon(x_1, x_2) = h^\varepsilon(x_1)$, and this function of one variable is C^2 and satisfies

$$\begin{aligned} &b^1 \partial_1 h^\varepsilon(x_1) + \frac{\varepsilon^2}{2} \partial_{11} h^\varepsilon(x_1) = 0, \quad x_1 > 0, \\ &h^\varepsilon(0) = 1, \\ &h^\varepsilon(x_1) \to 0, \quad x_1 \to \infty. \end{aligned}$$

Solving this linear ODE, we conclude

$$h^\varepsilon(x_1, x_2) = e^{-\frac{2b^1 x_1}{\varepsilon^2}}, \quad x_1 \geq 0.$$

Applying Doob's h-transform (Theorem A.6) we obtain that the process X_ε, conditioned to hit ∂O, is a diffusion with the same diagonal diffusion matrix and with drift given by

$$\begin{aligned} &\bar{b}^1_\varepsilon(x) = b^1 + \varepsilon^2 \frac{-\frac{2b_1}{\varepsilon^2} e^{-\frac{2b^1 x_1}{\varepsilon^2}}}{e^{-\frac{2b^1 x_1}{\varepsilon^2}}} = -b^1, \\ &\bar{b}^2_\varepsilon(x) = b^2, \quad x \in O. \end{aligned}$$

Therefore, $\bar{b}_\varepsilon \equiv (-b^1, b^2)$ for all ε. The conditioned diffusion is thus a Brownian motion with drift $(-b^1, b^2)$. This drift is directed towards the boundary and does not depend on ε, so conditions (A1) and (A2) follow and Theorem 2.1 applies. The Gaussian vector ϕ of Theorem 2.1 can also be easily computed explicitly since $\bar{b}_0 \equiv (-b^1, b^2)$, $D\bar{b}_0 = 0$, $\sigma = I$, and $\Phi(t) = I$ for all t.

In general, when the drift b is not constant and boundaries are curved, we do not expect the conditional drift to be independent of ε. However, in that case one can use PDE methods to check that conditions (A1) and (A2) still hold.

4. PDE approach to checking conditions of Theorem 2.1

From now on, besides the assumptions on O and Γ we made earlier, we assume that the domain O is bounded and Γ is open in the relative topology of ∂O, although some of our arguments can be modified to fit other situations as well.

According to Lemma A.4, the function h^ε satisfies

$$\sum_{i=1}^n b^i(x) \partial_i h^\varepsilon(x) + \frac{\varepsilon^2}{2} \sum_{i,j=1}^n a^{ij}(x) \partial_{ij} h^\varepsilon(x) = 0, \quad x \in O. \tag{4}$$

We also know that on the boundary

$$h^\varepsilon(x) = \begin{cases} 1, & x \in \Gamma, \\ 0, & x \in \partial O \setminus \Gamma, \end{cases}$$

but continuity of h^ε at a boundary point x depends on whether x is *regular*, i.e., whether for all $\varepsilon, t > 0$

$$\lim_{O \ni y \to x} \mathsf{P}_{\varepsilon, x}\{\tau_\varepsilon > t\} = 0.$$

In fact (see, e.g., implication (i)⇒(ii) of Theorem 2.3.3 in [31])

$$\lim_{O\ni y\to x} h^\varepsilon(y) = \begin{cases} 1, & \text{regular } x\in\Gamma, \\ 0, & \text{regular } x\in\partial O\setminus\bar\Gamma. \end{cases} \tag{5}$$

Let us make an assumption that all boundary points are regular. This condition is implied by the following *external cone condition*: for every boundary point x there are a cone K with base x and a neighborhood U of x such that $K\cap U\cap O=\emptyset$; see, e.g., [31, Section 2.3].

Since $0<h^\varepsilon(x)\le 1$ for $x\in O\cup\Gamma$, we can make a Hopf–Cole type logarithmic change of variables:

$$v^\varepsilon(x) = -\varepsilon^2\log h^\varepsilon(x),\quad x\in O\cup\Gamma,$$

so that $0\le v^\varepsilon(x)<\infty$ for $x\in O\cup\Gamma$. The reason for using this transformation is that now $\bar b_\varepsilon$ can be represented as

$$\bar b_\varepsilon(x) = b(x) - a(x)Dv^\varepsilon(x),\quad x\in O, \tag{6}$$

and one may hope to study the behavior of $\bar b_\varepsilon$ via the analysis of Dv^ε. Plugging $h^\varepsilon=\exp\{-v^\varepsilon/\varepsilon^2\}$ into (4) and (5), we obtain

$$-\frac{\varepsilon^2}{2}\sum_{i,j=1}^n a^{ij}\partial_{ij}v^\varepsilon - \sum_{i=1}^n b^i\partial_i v^\varepsilon + \frac12\sum_{i,j=1}^n a^{ij}\partial_i v^\varepsilon\partial_j v^\varepsilon = 0 \quad \text{in } O, \tag{7}$$

and

$$\lim_{O\ni y\to x} v^\varepsilon(y) = \begin{cases} 0, & x\in\Gamma, \\ +\infty, & x\in\partial O\setminus\bar\Gamma. \end{cases} \tag{8}$$

Equation (7) is a stationary HJB equation with positive viscosity and, since the boundary condition (8) does not depend on ε, one should expect that solutions of such equations converge as $\varepsilon\to0$ to a solution v^0 of the inviscid HJB equation

$$-\sum_{i=1}^n b^i\partial_i v^0 + \sum_{i,j=1}^n a^{ij}\partial_i v^0\partial_j v^0 = 0, \tag{9}$$

equipped with the boundary condition

$$v^0(x) = \begin{cases} 0, & x\in\Gamma, \\ +\infty, & x\in\partial O\setminus\bar\Gamma. \end{cases} \tag{10}$$

Therefore, a natural candidate for the vector field $\bar b_0$ of conditions (A1) and (A2) is $b-aDv^0$. However, several difficulties arise. First of all, classical smooth solutions of (9)–(10) often do not exist in the entire domain, and one has to deal with generalized solutions. Viscosity solutions (see, e.g., [8, 18, 26]) form a natural class of solutions, however even then the boundary condition (10) cannot be satisfied on the entire boundary and must be understood in a generalized sense. Still, one can prove convergence $v^\varepsilon\to v^0$, where v^0 is the value function for a variational problem associated with the HJB equation (9)–(10) and is a viscosity solution of this equation (see [16], [14]). For singular perturbation problems, one can establish estimates on convergence rates (see, e.g., [26, Chapter 6]) and even establish uniform asymptotic expansions of the form

$$v^\varepsilon = v^0 + \varepsilon^2 v_1 + \varepsilon^4 v_2 + \ldots + \varepsilon^{2k}v_k + o(\varepsilon^{2k})$$

in regions of strong regularity of the solution v^0, that is, in regions where it is smooth and coincides with the classical solution obtained by the method of characteristics (see [15,19]; for the method of characteristics, we refer the reader to [13, Chapter 3]). However, we need estimates of this kind not just for v^ε, but also for Dv^ε. Such estimates have been obtained in [15] for evolutionary problems with continuous boundary values. Our case is more delicate and no appropriate results seem to exist in the literature. Major difficulties come from the nature of the boundary condition and the fact that on the level of stochastic control our problem corresponds to an infinite horizon one.

In the remainder of this section we will provide sufficient conditions for the estimate

$$Dv^\varepsilon = Dv^0 + \varepsilon^2 Dv_1 + o(\varepsilon^2) \tag{11}$$

to hold uniformly in regions of strong regularity.

From now on we will restrict ourselves to the case of isotropic additive noise and assume that $a(x)$ is the identity matrix, i.e., $a^{ij}(x) \equiv \delta^{ij}$ (although the choice of $\sigma(x)$ guaranteeing this is not unique, the distribution of the resulting diffusion process does not depend on that choice, so for definiteness one may choose $\sigma(x)$ to be the $n \times n$ identity matrix). We will also assume that b is smooth (i.e., $b \in C^\infty(\bar{O})$), and that O is a bounded domain with smooth boundary ∂O. These will be our standing assumptions and they will not be repeated in the rest of the paper. The smoothness assumptions may not be necessary. However, we impose them for simplicity and to be able to cite explicitly results from papers where they were used. We remark that, requiring less regularity for the region of strong regularity, to carry out the program of our paper it would be enough to assume that $b \in C^4(\bar{O})$.

Equation (7) now becomes

$$-\frac{\varepsilon^2}{2}\Delta v^\varepsilon - \langle b, Dv^\varepsilon \rangle + \frac{1}{2}|Dv^\varepsilon|^2 = 0, \tag{12}$$

where Δ is the Laplace operator $\Delta f = \partial_{11} f + \ldots + \partial_{nn} f$, and $|\cdot|$ is the Euclidean norm. Equation (9) becomes

$$-\langle b, Dv \rangle + \frac{1}{2}|Dv|^2 = 0. \tag{13}$$

It is well known that natural candidates for viscosity solutions of HJB equations are value functions of the associated control/variational problems. Let us make this precise for equation (13). To that end we make the following additional assumptions on b:

(B1) There are a constant $\lambda > 0$ and a relatively open set $U \subset \bar{O}$ such that $\bar{\Gamma} \subset U$ and

$$\langle b(x), \nu(x) \rangle \le -\lambda < 0, \quad x \in U \cap \partial O, \tag{14}$$

where for a point $x \in \partial O$, $\nu(x)$ denotes the outward unit normal vector to ∂O at x.

(B2) For any absolutely continuous path $\gamma(\cdot) : [0, +\infty) \to \bar{O}$,

$$\int_0^{+\infty} |\dot{\gamma}(t) - b(\gamma(t))|^2 dt = +\infty.$$

The assumption (B2) was introduced in [16] and was also used in [14].

For a curve $\gamma \in H^1_{\mathrm{loc}}([0,+\infty);\mathbb{R}^n)$ we set

$$\tau_\gamma = \inf\{t \geq 0 : \gamma(t) \in \partial O\},$$

$$\bar{\tau}_\gamma = \inf\{t \geq 0 : \gamma(t) \in \mathbb{R}^n \setminus \bar{O}\},$$

and for $T > 0$ and a curve $\gamma \in H^1([0,T];\bar{O})$ we define the action functional

$$A^{0,T}(\gamma) = \frac{1}{2}\int_0^T |\dot{\gamma}(t) - b(\gamma(t))|^2 dt, \tag{15}$$

and the value function $v^0 : \bar{O} \to \mathbb{R}$ for the variational problem associated with the HJB equation

$$v^0(x) = \inf\Big\{A^{0,\bar{\tau}_\gamma}(\gamma) : \ \gamma \in H^1_{\mathrm{loc}}([0,+\infty);\mathbb{R}^n), \ \gamma(0) = x, \ \gamma(\bar{\tau}_\gamma) \in \Gamma\Big\}. \tag{16}$$

Lemma 4.1. *Under assumptions* (B1) *and* (B2)*, there is a relatively open subset N of $O \cup \Gamma$ satisfying the following conditions:*

(a) $\bar{N} \subset (O \cup \Gamma) \cap U$.

(b) *For every $x \in N$, there is a curve at which the infimum is attained in* (16)*. If γ_x^1, γ_x^2 are two minimizing curves, then $\tau_{\gamma_x^1} = \bar{\tau}_{\gamma_x^1} = \tau_{\gamma_x^2} = \bar{\tau}_{\gamma_x^2}$ and $\gamma_x^1(t) = \gamma_x^2(t), t \in [0, \tau_{\gamma_x^1}]$. In this sense the minimizing curve is unique and we will denote it by γ_x.*

(c) *For every $x \in N$ and all $t \in [0, \tau_{\gamma_x}]$, $\gamma_x(t) \in N$.*

(d) *The function $v^0 \in C^\infty(N)$.*

(e) *For every $x \in N$,*

$$\dot{\gamma}_x(t) = -b_0(\gamma_x(t)), \quad t \in [0, \tau_{\gamma_x}],$$

where

$$b_0(y) = -b(y) + Dv^0(y), \quad y \in N, \tag{17}$$

and $b_0(\gamma_x(\tau_{\gamma_x}))$ is transversal to Γ. The curve $\varphi_x(t) := \gamma_x(\tau_{\gamma_x} - t), t \in [0, \tau_{\gamma_x}]$, is the characteristic curve passing through x and it solves the equation $\dot{\varphi}_x(t) = b_0(\varphi_x(t))$.

(f) *For every characteristic curve $\varphi(t), 0 \leq t \leq s$, in N, there exists an open set $B \subset \mathbb{R}^{n-1}$ and a local smooth parametrization $\psi : B \to \Gamma$ such that $\psi(y_0) = \varphi(0)$ for some $y_0 \in B$ and if $x(t,y)$ denotes the characteristic curve such that $x(0,y) = \psi(y)$ (in particular, $\varphi(t) = x(t,y_0)$), then $x(t,y)$ is a smooth diffeomorphism of $[0,s] \times B$ onto its image.*

Any region N satisfying conditions (a)–(f) of Lemma 4.1 will be called a *region of strong regularity.* Our definition of a region of strong regularity is designed specifically for our case and just lists all the properties such a region should possess. In the literature, the term "region of strong regularity" may refer to a slightly different and more general object; see [15, 19]. In Section 5.1, we prove Lemma 4.1, stating the existence of regions of strong regularity. Although the method employed there proves strong regularity only in a small neighborhood of Γ, in reality regions of strong regularity may be quite large.

Part (e) of Lemma 4.1 implies that condition (A1) holds true for any x_0 in a region of strong regularity. Let us now address condition (A2) and introduce our last assumption.

(C) The set Γ is flat, i.e., it is contained in a hyperplane, and the entire domain O lies entirely on one side of that hyperplane.

Theorem 4.2. *Suppose assumptions* (B1), (B2), *and* (C) *hold. Let N be a region of strong regularity. Then there is a function $v_1 \in C^\infty(N)$ such that* (11) *holds uniformly on compact subsets of N.*

Section 5 is devoted to the proof of Theorem 4.2. This is the most technical part of this paper. Our multi-stage hard-analysis proof is based on the approach developed in [15] and [19], and involves PDE and stochastic methods.

The following is a simple corollary of Lemma 4.1 and Theorem 4.2.

Theorem 4.3. *Suppose assumptions* (B1), (B2), *and* (C) *hold. We recall that $\bar{b}_\varepsilon = b - Dv^\varepsilon$; see* (6). *Also, let $\bar{b}_0 = -b_0$, where b_0 is defined in* (17). *Then conditions* (A1) *and* (A2) *hold for any point x_0 belonging to a region of strong regularity and vector fields $\bar{b}_\varepsilon$ and $\bar{b}_0$. In particular, the CLT for conditioned exit distributions stated in Theorem* 2.1 *applies.*

While assumptions (B1) and (B2) define the setting, the technical assumption (C) does not seem to be necessary, and we believe that one can obtain similar results without it. However, our current approach depends on the additive noise character of equation (12) and does not work for curved boundaries since flattening transformations do not preserve its additive noise structure.

5. Asymptotics for the elliptic HJB equation

The goal of this section is to prove Lemma 4.1 and Theorem 4.2.

5.1. **Regions of strong regularity. Proof of Lemma 4.1.**

We recall the following results from [14, 16].

Theorem 5.1. *Let* (B2) *be satisfied. Then:*

(i) *For every domain O' such that $\bar{O}' \subset O \cup \Gamma$, there exists a constant $C(O')$ independent of ε such that*

$$\sup_{x\in O'} (|v^\varepsilon(x)| + |Dv^\varepsilon(x)|) \le C(O'). \tag{18}$$

(ii) *The function v^0 defined by* (16) *is Lipschitz on $\bar{O}$ and*

$$v^0(x) = \inf\Big\{A^{0,\tau_\gamma}(\gamma) : \ \gamma \in H^1_{\mathrm{loc}}([0,+\infty);\mathbb{R}^n),\ \gamma(0) = x,\ \gamma(\tau_\gamma) \in \Gamma\Big\},$$

on $O \cup \Gamma$. It is a viscosity solution of

$$\begin{cases} -\langle b, Dv^0\rangle + \frac{1}{2}|Dv^0|^2 = 0 & \text{in } O, \\ v^0 = 0 & \text{on } \Gamma. \end{cases} \tag{19}$$

(iii) *Uniformly on compact subsets of $O \cup \Gamma$,*

$$\lim_{\varepsilon\to 0} v^\varepsilon = v^0.$$

Part (i) is proved in Lemma 2.2 of [14] and part (ii) in Lemmas 2.3 and 2.4 of [14]. Part (iii) was first proved in [16] and later a proof that used PDE methods more heavily was given in Theorem 2.1 of [14]. Condition (B1) is not needed in Theorem 5.1.

For $U, V \subset \mathbb{R}^n$, $x \in \mathbb{R}^n$, we denote by $\mathrm{dist}(U,V)$ and $\mathrm{dist}(x,V)$, respectively, the Hausdorff distance between sets U, V and between x and V. We denote

$$O_\delta = \{x \in O : \mathrm{dist}(x, \partial O) < \delta\}, \quad d(x) = \mathrm{dist}(x, \partial O).$$

We recall that $d \in C^1(O_\delta)$ for some $\delta > 0$, and $Dd(x) = -\nu(x)$ on ∂O. Thus there exist $c, \alpha, \delta > 0$ such that the function $\tilde{d}(x) = cd(x)$ satisfies

$$-\langle b(x), D\tilde{d}(x)\rangle + \frac{1}{2}|D\tilde{d}(x)|^2 \le -\alpha < 0, \quad x \in U \cap O_\delta. \tag{20}$$

The proof of Lemma 4.1 is based on the method of characteristics. We refer the reader to [13, Chapter 3], [7], [22], [28], and Appendices in [19] and [15] for an overview of the method of characteristics. We sketch the proof below.

Let $\psi : B \to \Gamma$ be a local smooth parametrization of Γ, where $B \subset \mathbb{R}^{n-1}$ is an open set, and let $x_0 = \psi(y_0) \in \psi(B)$. It is easy to see that any solution u of (13) must either satisfy $Du(\psi(y)) = 0$ or $Du(\psi(y)) = 2\langle b(\psi(y)), \nu(\psi(y))\rangle \nu(\psi(y)) \neq 0$. We will look for a solution that satisfies the latter. The characteristic system for (13) is then the following (see [13, Chapter 3]). Functions

$$x(t) := x(t,y), \quad z(t) := z(t,y), \quad p(t) := p(t,y)$$

satisfy

$$\begin{cases} \dot{p}(t) = Db(x(t))^* p(t), \\ \dot{z}(t) = \langle -b(x(t)) + p(t), p(t)\rangle, \\ \dot{x}(t) = -b(x(t)) + p(t), \end{cases} \tag{21}$$

with initial conditions

$$x(0) = \psi(y), \quad z(0) = 0, \quad p(0) = 2\langle b(\psi(y)), \nu(\psi(y))\rangle \nu(\psi(y)).$$

The curve $x(t)$ is called the characteristic starting at $\psi(y)$. Since the right-hand side of (21) is a smooth function of x, z, p, the system has a unique local solution for any initial conditions which is smooth with respect to t and the initial conditions. Since b and ψ are smooth, we thus get that $x(t,y), z(t,y), p(t,y)$ are smooth functions of t, y in some neighborhood of $(0, y_0)$. Since

$$\langle \dot{x}(0,y), \nu(\psi(y))\rangle = \langle b(\psi(y)), \nu(\psi(y))\rangle \neq 0,$$

$\dot{x}(0,y)$ is transversal to Γ at every point $\psi(y)$ (the so-called noncharacteristic condition), and thus $\det(Dx(0,y_0)) \neq 0$ (where D is the space-time differential), which implies that $x(t,y)$ is a local smooth diffeomorphism, and hence the inverse functions $t(x), y(x)$ exist and are smooth. Therefore $u(x) := z(t(x), y(x))$ is a smooth function in some relatively open set containing x_0. One then proves (see [13, Chapter 3]) that $Du(x) = p(t(x), y(x))$ and u satisfies (13) in this set. Thus for every $x_0 \in \Gamma$, u is C^∞ in some relatively open region $C_{\tilde{B},\psi,\tau} = \{x(t,y) : y \in \tilde{B}, 0 \le t < \tau\}$ for some $\tau > 0$ and an open set $\tilde{B} \subset \mathbb{R}^{n-1}$. The union of such sets for $x_0 \in \Gamma$ will give us a relatively open region $\tilde{N}$ of $\bar{O}$ such that $\tilde{N} \cap \partial O = \Gamma$, and $u \in C^\infty(\tilde{N})$ and satisfies (13). We now claim that

$$u(x) = \min_{\mathcal{U}_{x,\tilde{N}}} A^{0,\tau_\gamma}(\gamma),$$

where

$$\mathcal{U}_{x,\tilde{N}} = \{\gamma(\cdot) \in H^1_{\rm loc}([0,+\infty);\mathbb{R}^n) : \\ \gamma(0)=x,\ \gamma(\tau_\gamma)\in\Gamma,\ \tau_\gamma<+\infty,\ \gamma(s)\in\tilde{N},\ 0\le s\le\tau_\gamma\},$$

and the minimizing curve is unique in $\mathcal{U}_{x,\tilde{N}}$. We will use that for $p\in\mathbb{R}^n$,

$$\frac{1}{2}|p|^2 = \sup_{y\in\mathbb{R}^n}\{\langle y,p\rangle - \frac{1}{2}|y|^2\}$$

with equality if and only if $y=p$. Let $\gamma(\cdot)\in\mathcal{U}_{x,\tilde{N}}$. Then

$$u(x) = u(x) - u(\gamma(\tau_\gamma)) = -\int_0^{\tau_\gamma}\langle Du(\gamma(t)),\dot\gamma(t)\rangle dt \\ = \int_0^{\tau_\gamma}\left(-\langle b(\gamma(t)),Du(\gamma(t))\rangle + \langle Du(\gamma(t)), b(\gamma(t))-\dot\gamma(t)\rangle - \frac{1}{2}|b(\gamma(t))-\dot\gamma(t)|^2\right)dt \\ + \int_0^{\tau_\gamma}\frac{1}{2}|b(\gamma(t))-\dot\gamma(t)|^2dt \le \int_0^{\tau_\gamma}\frac{1}{2}|b(\gamma(t))-\dot\gamma(t)|^2dt$$

with equality if and only if $b(\gamma(t))-\dot\gamma(t) = Du(\gamma(t))$. Thus the minimizing curve exists and is unique and thus conditions (c)–(f) are satisfied if v^0 there is replaced by u and $N=\tilde{N}$. Moreover, we can assume that $\tilde{N}\subset U\cap O_\delta$.

We now define

$$\mathcal{U}_x = \{\gamma(\cdot)\in H^1_{\rm loc}([0,+\infty);\mathbb{R}^n) : \gamma(0)=x, \gamma(\tau_\gamma)\in\Gamma, \tau_x<+\infty\}.$$

We will show that $u=v^0$ in a subset of $\tilde{N}$.

Lemma 5.2. *For every $\sigma>0$ there is $\delta(\sigma)>0$ such that $u=v^0$ in $N_\sigma=\{x\in\tilde{N}\cap O_{\delta(\sigma)} : {\rm dist}(x,\partial\tilde{N}\setminus\partial O)>\sigma\}$.*

Proof. Let $x\in N_\sigma$. It is enough to show that if $\gamma(\cdot)\in\mathcal{U}_x\setminus\mathcal{U}_{x,\tilde{N}}$, then

$$\frac{1}{2}\int_0^{\tau_\gamma}|\dot\gamma(s)-b(\gamma(s))|^2ds > u(x). \tag{22}$$

Let $t\in(0,\tau_\gamma)$ be the smallest t such that $\gamma(t)\in O\setminus\tilde{N}$. Then, using (20),

$$\tilde{d}(\gamma(t)) - \tilde{d}(x) = \int_0^t\langle D\tilde{d}(\gamma(s)),\dot\gamma(s)\rangle ds \\ = \int_0^t[\langle D\tilde{d}(\gamma(s)),\dot\gamma(s)-b(\gamma(s))\rangle + \frac{1}{2}|\dot\gamma(s)-b(\gamma(s))|^2 \\ + \langle D\tilde{d}(\gamma(s)),b(\gamma(s))\rangle - \frac{1}{2}|\dot\gamma(s)-b(\gamma(s))|^2]ds \\ \ge \int_0^t[\langle D\tilde{d}(\gamma(s)),b(\gamma(s))\rangle - \frac{1}{2}|D\tilde{d}(\gamma(s))|^2]ds - \frac{1}{2}\int_0^t|\dot\gamma(s)-b(\gamma(s))|^2ds \\ \ge \alpha t - \frac{1}{2}\int_0^t|\dot\gamma(s)-b(\gamma(s))|^2ds.$$

Since $\tilde{d}(x)\ge 0$, we obtain

$$\frac{1}{2}\int_0^t|\dot\gamma(s)-b(\gamma(s))|^2ds \ge \alpha t - \tilde{d}(\gamma(t)). \tag{23}$$

On the other hand,

$$\sigma \leq |\gamma(t) - x| = \left|\int_0^t \dot{\gamma}(s)ds\right| \leq \int_0^t |\dot{\gamma}(s) - b(\gamma(s))|ds + \int_0^t |b(\gamma(s))|ds$$
$$\leq \sqrt{t}\left(\int_0^t |\dot{\gamma}(s) - b(\gamma(s))|^2 ds\right)^{\frac{1}{2}} + Ct,$$

where $C > 0$ is a constant. Thus

$$\frac{1}{2}\int_0^t |\dot{\gamma}(s) - b(\gamma(s))|^2 ds \geq \frac{{(\sigma - Ct)_+}^2}{2t},$$

where for $a \in \mathbb{R}$, $a_+ = \max(a, 0)$. Since there is a constant $C_1 > 0$ such that

$$u(x) \leq C_1\delta, \quad x \in \tilde{N} \cap O_\delta, \tag{24}$$

we conclude that there are $t_0(\sigma) > 0$ and $\delta_0 > 0$ such that $\delta < \delta_0$ and $t < t_0(\sigma)$ imply (independently of x) inequality (22). Hence from now on we can assume that $\delta < \delta_0$ and $t_0(\sigma) \leq t < \tau_\gamma$.

Let us denote $\eta = \tilde{d}(\gamma(t))$. Now let $t_1 \in [t, \tau_\gamma)$ be such that $\gamma(t_1) \in \partial(O_\eta \cap U)\backslash\partial O$, and $\gamma(s) \in O_\eta \cap U$ for $t_1 < s \leq \tau_\gamma$. Arguing as in the derivation of(23), we obtain

$$\frac{1}{2}\int_{t_1}^{\tau_\gamma} |\dot{\gamma}(s) - b(\gamma(s))|^2 ds \geq \tilde{d}(\gamma(t_1)) + \alpha(\tau_\gamma - t_1). \tag{25}$$

Denote $r = \text{dist}(\partial U, \Gamma)$. If $\tilde{d}(\gamma(t_1)) = \eta = \tilde{d}(\gamma(t))$, then, combining (23) and (25), we obtain

$$\frac{1}{2}\int_0^{\tau_\gamma} |\dot{\gamma}(s) - b(\gamma(s))|^2 ds \geq \alpha(t_0(\sigma) + \tau_\gamma - t_1),$$

and, due to (24), there is $\delta_1(\sigma) > 0$ such that (22) holds true for $\delta < \delta_1(\sigma)$. If $\tilde{d}(\gamma(t_1)) \neq \eta$, we must have $\gamma(t_1) \in \partial U$, and then

$$\frac{1}{2}\int_{t_1}^{\tau_\gamma} |\dot{\gamma}(s) - b(\gamma(s))|^2 ds \geq \frac{{(r - C(\tau_\gamma - t_1))_+}^2}{2(\tau_\gamma - t_1)},$$

which together with (24) implies that there is $\delta_2 > 0$ such that if $\delta < \delta_2$, then (22) holds unless $(\tau_\gamma - t_1) \geq \tilde{t}_0 > 0$ for some $\tilde{t}_0 = \tilde{t}_0(r)$. In this case, combining (23), (25), and the fact that $\eta \leq \delta$, we obtain that there is $\delta_3(r) > 0$ such that for $\delta < \delta_3(r)$,

$$\frac{1}{2}\int_0^{\tau_\gamma} |\dot{\gamma}(s) - b(\gamma(s))|^2 ds \geq \alpha(t_0(\sigma) + \tilde{t}_0(r)) - \eta > \alpha t_0(\sigma).$$

Using (24) once again we conclude that there is $\delta_4(\sigma, r) > 0$ such that in this case (22) holds for $\delta < \delta_4(\sigma, r)$. The lemma now follows with $\delta(\sigma) = \min\{\delta_i,\ i = 0, \ldots, 4\}$. □

The region N_σ itself may not be a region of strong regularity since it may not satisfy condition (f). However we can now take N to be the union over all $x_0 \in \Gamma$ of sets of the form $C_{\tilde{B},\psi,\tau} \subset N_\sigma$ described before.

5.2. **Convergence of derivatives.** The main goal of this section is to prove Lemma 5.8 stating uniform convergence of Dv^ε to Dv^0 as $\varepsilon \to 0$. Since b and ∂O are smooth it is well known that in this case the function $h^\varepsilon \in C^2(\Omega)$, it is continuous at every point of Γ, $0 < h^\varepsilon < 1$ in O, and, moreover (see, e.g., [23, Section 6.5], h^ε is smooth in every relatively open region O' such that $\bar{O}' \subset O \cup \Gamma$. Therefore v^ε is also smooth in every relatively open region O' such that $\bar{O}' \subset O \cup \Gamma$.

Without further loss of generality we suppose that $\Gamma \subset \{x_n = 0\}$ and $N \subset \{x_n \geq 0\}$. Under this assumption, (14) implies that $-\langle b, e_n\rangle \leq -\lambda < 0$ in a relative neighborhood of Γ in ∂O, where $e_n = (0, \dots, 0, 1)$ is the standard basis vector.

For $r > 0$, we denote

$$N_r = \{x = (x', x_n) \in N : \mathrm{dist}((x', 0), \partial N \setminus \Gamma) > 2r, x_n < r\}.$$

Lemma 5.3. *For every $r > 0$ there are positive numbers L, C, ε_0 such that for $\varepsilon < \varepsilon_0$,*

$$v^\varepsilon(x) \leq (1 + L\varepsilon^2)v^0(x) + Cx_n^2, \quad x \in N_r. \tag{26}$$

Proof. For $y \in N_r$, we denote $W_r(y) = \{x = (x', x_n) : |x' - y'| < r, x_n < r\}$. Let us define

$$\psi_y^\varepsilon(x) = (1 + L\varepsilon^2)v^0(x) + Cx_n^2 + C_1 x_n |x' - y'|^2, \quad x, y \in N_r.$$

We will prove that there is a choice of constants L, C, C_1 such that

$$v^\varepsilon(x) \leq \psi_y^\varepsilon(x), \quad y \in N_r, x \in W_r(y).$$

This will obviously imply (26) since for any $x \in N_r$ one can choose $y = x$. The task will be accomplished by showing that ψ_y^ε is a supersolution of equation (12) in W, for which v^ε is a solution, and by using the comparison principle.

To simplify notation, without loss of generality we will assume that $y' = 0$ and use ψ^ε for ψ_0^ε.

Directly from (13) we know that

$$Dv^0(x) = 2\langle b, e_n\rangle e_n, \quad x \in \Gamma. \tag{27}$$

Therefore, we can choose r such that for some $c_0 > 0$,

$$|Dv^0(x)| \geq c_0, \quad \langle -b(x) + Dv^0(x), e_n\rangle \geq c_0, \quad x \in W_r(y), \tag{28}$$

and from this point on we write W for $W_r(0)$. Due to (18), the inequality

$$v^\varepsilon(x) \leq C_1 r^2 x_n \quad \text{on } \{x \in \partial W : |x'| = r\}$$

holds with $C_1 = C(N)/r^2$. Since v^ε converges uniformly to v^0 in N, the estimate

$$v^\varepsilon(x) \leq v^0(x) + Cr^2 \leq \psi^\varepsilon(x) \quad \text{on } \{x \in \partial W : x_n = r\}$$

holds true if we choose, say, $C \geq 1$ in the definition of ψ^ε and let ε be small enough. We have

$$D\psi^\varepsilon(x) = (1 + L\varepsilon^2)Dv^0(x) + (2Cx_n + C_1|x'|^2)e_n + 2C_1 x_n(x', 0).$$

If $L\varepsilon^2 \leq 1$, then, due to the smoothness of v^0, there is a number $M(C, C_1) > 0$ such that for all ε,

$$|D\psi^\varepsilon(x)|, \|D^2\psi^\varepsilon(x)\| \leq M(C, C_1), \quad x \in W. \tag{29}$$

Here, for an $n \times n$ matrix A, $\|A\| = \max_{i,j=1,\dots,n} |A_{ij}|$.

Using (28) and (29), we obtain

$$
\begin{aligned}
&-\frac{\varepsilon^2}{2}\Delta\psi^\varepsilon(x) - \langle b(x), D\psi^\varepsilon(x)\rangle + \frac{1}{2}|D\psi^\varepsilon(x)|^2 \geq -\frac{\varepsilon^2 nM}{2} \\
&+ (1+L\varepsilon^2)\left(-\langle b(x), Dv^0(x)\rangle + \frac{1}{2}|Dv^0(x)|^2\right) + \frac{L\varepsilon^2(1+L\varepsilon^2)}{2}|Dv^0(x)|^2 \\
&+ \langle -b(x) + Dv^0(x), e_n\rangle(2Cx_n + C_1|x'|^2) + L\varepsilon^2\langle Dv^0(x), e_n\rangle(2Cx_n + C_1|x'|^2) \\
&- 2C_1x_n\langle b(x), (x',0)\rangle + 2C_1(1+L\varepsilon^2)x_n\langle Dv^0(x), (x',0)\rangle \\
&+ \frac{1}{2}\left|(2Cx_n + C_1|x'|^2)e_n + 2C_1x_n(x',0)\right|^2 \\
&\geq -\frac{\varepsilon^2 nM}{2} + c_0(2Cx_n + C_1|x'|^2) + \frac{L\varepsilon^2 c_0^2}{2} - C_2x_n|x'|,
\end{aligned}
$$

where C_2 depends only on C_1 and the bounds on $|b|, |Dv^0|$. It is now clear that the last line of the above inequality can be made nonnegative on W by taking L and C large enough. Therefore ψ^ε is a supersolution of (12) in W and $v^\varepsilon \leq \psi^\varepsilon$ on ∂W by (12) and (13). Therefore, by the comparison principle (see for instance [23, Theorem 17.1]) we obtain $v^\varepsilon \leq \psi^\varepsilon$ in W. ☐

Corollary 5.4. *For every $r > 0$ there exist $L_1, \varepsilon_0 > 0$ such that for $\varepsilon < \varepsilon_0$,*

$$
\partial_n v^\varepsilon(x) \leq (1+L\varepsilon^2)\partial_n v^0(x) \leq \partial_n v^0(x) + L_1\varepsilon^2, \quad x \in N(r)\cap\Gamma. \tag{30}
$$

Corollary 5.5. *Let ε_0 be as in Corollary* 5.4*. Then there is a constant $L_2 > 0$ such that for $\varepsilon < \varepsilon_0$,*

$$
\partial_{nn} v^\varepsilon(x) \leq L_2, \quad x \in N(r)\cap\Gamma. \tag{31}
$$

Proof. Let $x \in N(r)\cap\Gamma$. Then

$$
\partial_{nn}v^\varepsilon(x) = \Delta v^\varepsilon(x) = \frac{2}{\varepsilon^2}\left(-\langle b(x), Dv^\varepsilon(x)\rangle + \frac{1}{2}|Dv^\varepsilon(x)|^2\right).
$$

Since $v_\varepsilon \equiv 0$ on Γ, we have $Dv^\varepsilon(x) = t(x)e_n$ for some $t(x) \geq 0$. Combining this with Corollary 5.4 and (27), we obtain $t(x) \leq 2\langle b(x), e_n\rangle + L_1\varepsilon^2$. Hence

$$
\partial_{nn}v^\varepsilon(x) = \frac{2}{\varepsilon^2}\left(-\langle b(x), e_n\rangle t + \frac{1}{2}t^2\right) \leq \frac{2}{\varepsilon^2}\left(\langle b(x), e_n\rangle L_1\varepsilon^2 + \frac{1}{2}L_1^2\varepsilon^4\right),
$$

which implies our claim. ☐

Remark 5.6. A version of Lemma 5.3 could be obtained without the assumption that Γ is flat. Corollaries 5.4 and 5.5 are also true in this case if $\partial_n v^0$, $\partial_n v^\varepsilon$, and $\partial_{nn}v^\varepsilon$ are replaced by the respective derivatives along the inward normal vector. Lemma 5.7 below is the only place where we have to assume that Γ is flat.

The following lemma and its proof is a standard semiconcavity result from [26] which was also used in [19]. The difference between it and Lemma 3.1 of [19] is that we prove the semiconcavity up to the boundary Γ, although only along coordinate directions. This is made possible by Corollary 5.5. On compact subsets of O a semiconcavity estimate (32) holds along every direction but we do not need it here. We include the proof of the lemma for completeness.

Lemma 5.7. *If N' is a relatively compact subset of $O\cup\Gamma$, then there is a constant $C(N')>0$ such that*

$$\partial_{ii}v^\varepsilon(x)\le C(N'),\quad x\in N', i=1,\dots,n. \tag{32}$$

Proof. Since $\operatorname{dist}(N',\partial O\setminus\Gamma)>0$, there exists a function $\xi\in C_0^2(O\cup\Gamma)$ such that $0\le\xi\le 1$ for all x, $\xi=1$ on N', and

$$\frac{|D\xi|^2}{\xi}\le C_1\quad\text{on the support of }\xi$$

for some C_1. To simplify notation, for a function f with real or vector values, we will write f_i and f_{ii} for $\partial_i f$ and $\partial_{ii}f$. Differentiating (12) with respect to x_i we have

$$-\frac{\varepsilon^2}{2}\Delta v_i^\varepsilon-\langle b_i,Dv^\varepsilon\rangle-\langle b,Dv_i^\varepsilon\rangle+\langle Dv^\varepsilon,Dv_i^\varepsilon\rangle=0,$$

$$-\frac{\varepsilon^2}{2}\Delta v_{ii}^\varepsilon-\langle b_{ii},Dv^\varepsilon\rangle-2\langle b_i,Dv_i^\varepsilon\rangle-\langle b,Dv_{ii}^\varepsilon\rangle+|Dv_i^\varepsilon|^2+\langle Dv^\varepsilon,Dv_{ii}^\varepsilon\rangle=0.$$

Therefore, setting $w=\xi v_{ii}^\varepsilon$, we have

$$\begin{aligned}
&-\frac{\varepsilon^2}{2}\Delta w+\varepsilon^2\operatorname{tr}\left(\frac{D\xi}{\xi}\otimes Dw\right)=-\frac{\varepsilon^2}{2}\xi\Delta v_{ii}^\varepsilon-\frac{\varepsilon^2}{2}v_{ii}^\varepsilon\Delta\xi+\varepsilon^2\frac{|D\xi|^2}{\xi}v_{ii}^\varepsilon\\
&=\xi\langle b_{ii},Dv^\varepsilon\rangle+2\xi\langle b_i,Dv_i^\varepsilon\rangle+\xi\langle b,Dv_{ii}^\varepsilon\rangle-\xi|Dv_i^\varepsilon|^2-\xi\langle Dv^\varepsilon,Dv_{ii}^\varepsilon\rangle\\
&-\frac{\varepsilon^2}{2}\left(\Delta\xi-2\frac{|D\xi|^2}{\xi}\right)v_{ii}^\varepsilon=\xi\langle b_{ii},Dv^\varepsilon\rangle+2\xi\langle b_i,Dv_i^\varepsilon\rangle-\xi|Dv_i^\varepsilon|^2\\
&+\langle b-Dv^\varepsilon,Dw\rangle-\langle b-Dv^\varepsilon,D\xi\rangle v_{ii}^\varepsilon-\frac{\varepsilon^2}{2}\left(\Delta\xi-2\frac{|D\xi|^2}{\xi}\right)v_{ii}^\varepsilon.
\end{aligned} \tag{33}$$

The value of w on ∂O is equal to 0 if $i=1,\dots,n-1$, and it is bounded by Corollary 5.5 if $i=n$. If w has a positive maximum at a point $x\in O$, then $Dw(x)=0$ and $\Delta w(x)\le 0$. Since $|Dv^\varepsilon|$ are bounded on the support of ξ, we thus obtain from (33) that for some numbers C_2,C_3,C_4 depending on ξ but not on ε,

$$(\xi(x))^2(|Dv_i^\varepsilon(x)|^2-C_2|Dv_i^\varepsilon(x)|)\le C_3+C_4w(x).$$

Applying the standard inequality $ab<(a^2/c+cb^2)/2$ to $C_2|Dv_i^\varepsilon(x)|$, we obtain that there are numbers $C_5,C_6>0$ independent of ε such that

$$|Dv_i^\varepsilon(x)|^2-C_2|Dv_i^\varepsilon(x)|\ge C_5|Dv_i^\varepsilon(x)|-C_6.$$

Since ξ^2 is bounded by 1, we obtain

$$(w(x))^2\le(\xi(x))^2|Dv_i^\varepsilon(x)|^2\le(C_6+C_3+C_4w(x))/C_5,$$

and the resulting uniform upper bound on w implies (32). □

The following lemma is now a consequence of Corollary 5.4 and Lemma 5.7 (see the proof of Theorem 3.2 of [19] for a similar argument).

Lemma 5.8. *Dv^ε converges to Dv^0 uniformly on every compact subset of N.*

Proof. Let N' be a compact subset of N. Since Dv^0 is continuous in N, it is enough to show for all $i=1,\dots,n$ that if for all $m\in\mathbb{N}$, $x^{(m)}\in N'$, and $x^{(m)}\to x\in N'$, $\varepsilon_m\to 0$ as $m\to\infty$, then it is possible to extract a subsequence of $(\partial_i v^{\varepsilon_m}(x^{(m)}))$ convergent to $\partial_i v^0(x)$.

For $h \in \mathbb{R}$ such that $x^{(m)} + he_i \in N'$, (31) implies

$$v^{\varepsilon_m}(x^{(m)} + he_i) \leq v^{\varepsilon_m}(x^{(m)}) + \partial_i v^{\varepsilon_m}(x^{(m)})h + \frac{L_2}{2}h^2. \tag{34}$$

Since $\partial_i v^{\varepsilon_m}(x^{(m)})$ is bounded, taking a subsequence if necessary, we have $\partial_i v^{\varepsilon_m}(x^{(m)}) \to p$ for some p. Since $v^{\varepsilon_m} \to v^0$ uniformly on N, (34) implies

$$v^0(x + he_i) \leq v^0(x) + ph + \frac{L_2}{2}h^2 \tag{35}$$

for sufficiently small $|h|$ if $x \notin \Gamma$ and for sufficiently small $h > 0$ if $x \in \Gamma$. If $i = 1, \ldots, n-1$, (35) clearly implies $p = \partial_i v^0(x)$. If $i = n$ and $x \notin \Gamma$, we also have $p = \partial_n v^0(x)$. If $x = (x', 0) \in \Gamma$, let $x^{(m)} = (x^{(m)'}, a_m)$, i.e., $a_m = x_n^{(m)}$. Then $a_m \geq 0$, $a_m \to 0$ as $m \to \infty$. Using (31), we get

$$\begin{aligned}\partial_n v^{\varepsilon_m}(x^{(m)'}, a_m) - \partial_n v^{\varepsilon_m}(x^{(m)'}, 0) &= \int_0^1 \frac{d}{dt}\left[\partial_n v^{\varepsilon_m}(x^{(m)'}, ta_m)\right] dt \\ &= a_m \int_0^1 \partial_{nn} v^{\varepsilon_m}(x^{(m)'}, ta_m) dt \leq a_m L_2.\end{aligned}$$

Therefore, by (30),

$$\partial_n v^{\varepsilon_m}(x^{(m)}) \leq \partial_n v^{\varepsilon_m}(x^{(m)'}, 0) + a_m L_2 \leq \partial_n v^0(x^{(m)'}, 0) + \varepsilon_m^2 L_1 + a_m L_2.$$

This implies $p \leq \partial_n v^0(x)$. Combining this with (35), we obtain $p = \partial_n v^0(x)$. □

5.3. **Asymptotics of v^ε.** We follow the method of [19]. We recall that any characteristic γ of (13) satisfies $\dot{\gamma} = b_0(\gamma)$, where $b_0(x) = -b(x) + Dv^0(x)$.

By the construction in [19], page 439, for every $\bar{x} \in N$ there exists a relatively open in N subregion of regularity N_γ which is a neighborhood of the characteristic curve γ connecting $\bar{x}$ with ∂O (and consisting of characteristic curves), and smooth functions F, G on N_γ such that

$$F > 0 \text{ in } N_\gamma, \quad F = 0 \text{ on } \partial N_\gamma \setminus \partial O,$$

$$G > 0 \text{ in } N_\gamma \setminus \partial O, \quad G = 0 \text{ on } N_\gamma \cap \partial O,$$

$$-\langle b_0, DF \rangle = 1 \text{ in } N_\gamma,$$

$$\langle b_0, DG \rangle = 1 \text{ in } N_\gamma.$$

The following lemma was proved in [19].

Lemma 5.9 ([19, Lemma 4.1]). *Let $\beta_\varepsilon \to b_0$ uniformly on compact subsets of N as $\varepsilon \to 0$. Suppose, for each $\varepsilon > 0$, functions w^ε, A^ε and numbers C^ε, a^ε satisfy*

$$\begin{cases} -\dfrac{\varepsilon^2}{2}\Delta w^\varepsilon(x) + \langle \beta_\varepsilon(x), Dw^\varepsilon(x) \rangle = A^\varepsilon(x), & x \in N_\gamma, \\ w^\varepsilon(x) = 0, \quad x \in N_\gamma \cap \partial O, \\ |w^\varepsilon(x)| \leq C^\varepsilon, \quad x \in \partial N_\gamma \setminus \partial O, \end{cases}$$

and $|A^\varepsilon(x)| \leq a^\varepsilon$ for all $x \in N_\gamma$. Then there exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$,

$$|w^\varepsilon(x)| \leq C^\varepsilon e^{-\frac{2F(x)}{\varepsilon^2}} + 4a^\varepsilon G(x), \quad x \in N_\gamma.$$

Theorem 5.10. *As $\varepsilon \to 0$,*

$$v^\varepsilon = v^0 + \varepsilon^2 v_1 + o(\varepsilon^2)$$

uniformly on compact subsets of N. Here, the function $v_1 \in C^\infty(N)$ is a unique solution of

$$\begin{cases} \langle b_0(x), Dv_1(x)\rangle = \frac{1}{2}\Delta v^0(x), & x \in N, \\ v_1(x) = 0, & x \in N \cap \partial O. \end{cases} \tag{36}$$

Proof. First we recall that since the characteristic curves for (36) are the same as for (13) in N, v_1 is obtained by the method of characteristics in N and is in $C^\infty(N)$.

Let N' be a compact subset of N. It can be covered by a finite number of sets $N_{\gamma_i}^{\delta_i} = \{x \in N_{\gamma_i} : \mathrm{dist}(x, \partial N_{\gamma_i} \setminus \partial O) > \delta_i\}, i = 1, \dots, m$. Let us define

$$v_1^\varepsilon = \frac{v^\varepsilon - v^0}{\varepsilon^2}.$$

We will show that $v_1^\varepsilon \to v_1$ uniformly on every $N_{\gamma_i}^{\delta_i}$. On N we have

$$-\frac{\varepsilon^2}{2}\Delta v^\varepsilon - \langle b, Dv^\varepsilon\rangle + \frac{1}{2}|Dv^\varepsilon|^2 = 0$$

and

$$-\frac{\varepsilon^2}{2}\Delta v^0 - \langle b, Dv^0\rangle + \frac{1}{2}|Dv^0|^2 = -\frac{\varepsilon^2}{2}\Delta v^0.$$

Therefore,

$$-\frac{\varepsilon^2}{2}\Delta(v^\varepsilon - v^0) - \langle b, D(v^\varepsilon - v^0)\rangle + \frac{1}{2}\langle Dv^\varepsilon + Dv^0, D(v^\varepsilon - v^0)\rangle = \frac{\varepsilon^2}{2}\Delta v^0,$$

which gives

$$-\frac{\varepsilon^2}{2}\Delta v_1^\varepsilon + \langle \beta_\varepsilon, Dv_1^\varepsilon\rangle = \frac{1}{2}\Delta v^0,$$

where $\beta_\varepsilon(x) = -b(x) + (Dv^\varepsilon(x) + Dv^0(x))/2$. Lemma 5.8 implies that $\beta_\varepsilon \to b_0$ uniformly on compact subsets of N. It follows from (36) that

$$-\frac{\varepsilon^2}{2}\Delta v_1 + \langle \beta_\varepsilon, Dv_1\rangle = \frac{1}{2}\Delta v^0 - \frac{\varepsilon^2}{2}\Delta v_1 + \langle \beta_\varepsilon - b_0, Dv_1\rangle.$$

Therefore, the function $w^\varepsilon = v_1^\varepsilon - v_1$ satisfies

$$-\frac{\varepsilon^2}{2}\Delta w^\varepsilon + \langle \beta_\varepsilon, Dw^\varepsilon\rangle = \frac{\varepsilon^2}{2}\Delta v_1 - \langle \beta_\varepsilon - b_0, Dv_1\rangle \quad \text{in } N_{\gamma_i}$$

and $w^\varepsilon = 0$ on $N_{\gamma_i} \cap \partial O$. Moreover, there is a constant C, such that

$$|w^\varepsilon| \le \frac{C}{\varepsilon^2} \quad \text{on } \partial N_{\gamma_i} \setminus \partial O,\ i = 1, \dots, m,$$

and if $A^\varepsilon(x) = \frac{\varepsilon^2}{2}\Delta v_1(x) - \langle \beta_\varepsilon(x) - b_0(x), Dv_1(x)\rangle$, then

$$|A^\varepsilon(x)| \le a^\varepsilon \to 0 \quad \text{on } \bigcup_{i=1}^m N_{\gamma_i}.$$

By Lemma 5.9 we thus obtain that for any $i = 1, \dots, m$,

$$|w^\varepsilon(x)| \le \frac{C}{\varepsilon^2} e^{-\frac{2F_i(x)}{\varepsilon^2}} + 4a^\varepsilon G_i(x), \quad x \in N_{\gamma_i}.$$

This implies that $w^\varepsilon \to 0$ as $\varepsilon \to 0$ uniformly on $N_{\gamma_i}^{\delta_i}$ for any $i = 1, \dots, m$. □

5.4. **Asymptotics of Dv^ε.** We use the strategy from [15] where the asymptotics of derivatives was proved for parabolic problems with zero boundary value on the parabolic boundary. However, we use simpler PDE arguments whenever possible.

Lemma 5.11. *$Dv_1^\varepsilon \to Dv_1$ uniformly on compact subsets of $N \cap \partial O$.*

Proof. Let A be a compact subset of $N \cap \partial O$. Since $v_1^\varepsilon = v_1 = 0$ on $N \cap \partial O$, we only need to show $\partial_n v_1^\varepsilon \to \partial_n v_1$ on A. Let $x_0 = (x_0', 0) \in A$. We have

$$\begin{aligned} \partial_n v_1^\varepsilon(x_0) - \partial_n v_1(x_0) &= \lim_{x_n \to 0} \frac{v_1^\varepsilon(x_0', x_n) - v_1(x_0', x_n)}{x_n} \\ &= \partial_n v^0(x_0) \lim_{x_n \to 0} \frac{v_1^\varepsilon(x_0', x_n) - v_1(x_0', x_n)}{v^0(x_0', x_n)}. \end{aligned} \tag{37}$$

Let $0 < \sqrt{2}r < \mathrm{dist}(A, \partial N \cap O)$ and $W_r(x_0) = B_r(x_0') \times [0, r)$, where $B_r(x_0')$ stands for the Euclidean ball of radius r centered at x_0'. We also require that r is small enough so that there is a constant $c_1 > 0$ such that

$$|Dv^0(x)| \geq c_1, \quad x \in W_r(x_0), \tag{38}$$

for every $x_0 \in A$. The functions $w^\varepsilon = \pm(v_1^\varepsilon - v_1)$ satisfy

$$-\frac{\varepsilon^2}{2}\Delta w^\varepsilon(x) + \langle \beta_\varepsilon(x), Dw^\varepsilon(x)\rangle = \pm A^\varepsilon(x), \quad x \in W_r(x_0),$$

where $A^\varepsilon(x) = \frac{\varepsilon^2}{2}\Delta v_1(x) - \langle \beta_\varepsilon(x) - b^0(x), Dv_1(x)\rangle$. Also, $w^\varepsilon(x) = 0$ for $x \in W \cap \partial O$, and there are numbers $a^\varepsilon, c^\varepsilon \to 0$ independent of $x_0 \in A$ such that $|A(x)| \leq a^\varepsilon$ for $x \in W_r(x_0)$ and $|w^\varepsilon(x)| \leq c^\varepsilon$ for $x \in \partial W_r(x_0) \setminus \partial O$. We also recall that $\beta_\varepsilon \to b_0$ uniformly on compact subsets of N.

Since

$$\langle b_0(x), Dv^0(x)\rangle = \frac{1}{2}|Dv^0(x)|^2 \geq \frac{c_1^2}{2}, \quad x \in W_r(x_0),$$

it follows that for sufficiently small ε,

$$\langle \beta_\varepsilon(x), Dv^0(x)\rangle \geq \frac{c_1^2}{4}, \quad x \in W_r(x_0).$$

Let $\eta > 0$. We set

$$\psi(x) = \eta v^0(x) + \frac{c^\varepsilon}{r^2}|x' - x_0'|^2.$$

Then for ε small enough (but independent of x_0), $\psi \geq c^\varepsilon$ on $\partial W_r(x_0) \setminus \partial O$, and thus $\psi \geq |w^\varepsilon|$ on $\partial W_r(x_0)$. Moreover, for some constants c_2, c_3

$$-\frac{\varepsilon^2}{2}\Delta\psi(x) + \langle \beta_\varepsilon(x), D\psi(x)\rangle \geq -\varepsilon c_2 + \frac{\eta c_1^2}{4} - c^\varepsilon c_3 \geq \frac{\eta c_1^2}{8} \geq a^\varepsilon$$

if ε is small enough. Therefore, by comparison we obtain $|w^\varepsilon| = \max(\pm w^\varepsilon) \leq \psi$ in $W_r(x_0)$. Hence

$$\left|\frac{v_1^\varepsilon(x_0', x_n) - v_1(x_0', x_n)}{v^0(x_0', x_n)}\right| \leq \eta$$

if $0 < x_n < r$, and the claim follows since η is arbitrary. □

Proof of Theorem 4.2. Our arguments follow those of the proof of Theorem 6.4 of [15].

It is sufficient to show that the convergence is uniform on every relatively open subregion of strong regularity $N_1 \subset N$ such that $\bar{N}_1 \subset N$. Let N_2 be a relatively open subset of N such that $\bar{N}_1 \subset \bar{N}_2 \subset N$. Let us introduce $b_\varepsilon(x) = -b(x) + Dv^\varepsilon(x)$

so that $\bar{b}_\varepsilon(x) = -b_\varepsilon(x)$; see (3). We extend b_0 and b_ε outside N_2 to be Lipschitz functions on $\mathbb{R}^n$ such that

$$\sup_{\mathbb{R}^n} |b_\varepsilon - b_0| = c^\varepsilon \to 0 \quad \text{as } \varepsilon \to 0. \tag{39}$$

We will denote by L the Lipschitz constant of b_0. For $x \in N_1$, let X_0 be the solution of

$$\begin{cases} \dot{X}_0(t) = -b_0(X_0(t)), \\ X_0(0) = x, \end{cases} \tag{40}$$

and let X_ε be the strong solution of the Itô equation

$$\begin{cases} dX_\varepsilon(t) = -b_\varepsilon(X_\varepsilon(t)) + \varepsilon dW(t), \\ X_\varepsilon(0) = x, \end{cases} \tag{41}$$

where W is a standard n-dimensional Wiener process defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. We denote $\tau_x^0, \tau_x^\varepsilon$ to be respectively the first exit times of X_0 and X_ε from N_2. We recall that $X_0(t)$ is the characteristic of (13) and (36) passing through x with its time parametrization reversed. Let T_1 be such that $\tau_x^0 \le T_1$ for all $x \in N_1$ and let $T = T_1 + 1$. Since $\langle b_0(x), e_n\rangle > c_0 > 0$ for some c_0 and all $x \in N_1$ sufficiently close to ∂O, there are $c_1, s_0 > 0$ such that

$$\operatorname{dist}(X_0(\tau_x^0 + s), \partial O) \ge c_1|s| \quad \text{for all } x \in N_1 \text{ and } |s| \le s_0. \tag{42}$$

Let us introduce $A_\varepsilon := \{\omega \in \Omega : \sqrt{\varepsilon}\sup_{0\le t\le T} |W(t)(\omega)| \le 1\}$. Notice that the sets form a monotone family: $A_{\varepsilon_1} \supset A_{\varepsilon_2}$ if $\varepsilon_1 \le \varepsilon_2$. Also, $\mathsf{P}(\bigcup_{\varepsilon>0} A_\varepsilon) = 1$. A standard maximal inequality for W implies that for some $C, \gamma > 0$,

$$\mathsf{P}(A_\varepsilon^c) \le Ce^{-\frac{\gamma}{\varepsilon}}. \tag{43}$$

It follows from (39), (40), and (41) that

$$|X_\varepsilon(t) - X_0(t)| \le L\int_0^t |X_\varepsilon(s) - X_0(s)|ds + c^\varepsilon t + \varepsilon|W(t)|.$$

Therefore, for all $x \in N_1$, $\omega \in A_\varepsilon$,

$$|X_\varepsilon(t) - X_0(t)| \le L\int_0^t |X_\varepsilon(s) - X_0(s)|ds + c^\varepsilon t + \sqrt{\varepsilon},$$

and, by Gronwall's inequality,

$$\sup_{0\le t\le T} |X_\varepsilon(t) - X_0(t)| \le k^\varepsilon, \tag{44}$$

where $k^\varepsilon = (c^\varepsilon T + \sqrt{\varepsilon})e^{TL} \to 0$, as $\varepsilon \to 0$. Since $X_0(t) \in N_1$ for $0 \le t \le \tau_x^0$, this implies $X_\varepsilon(t) \in N_2$ for $0 \le t \le \tau_x^0 \wedge \tau_x^\varepsilon$ for $\omega \in A_\varepsilon$ if $\varepsilon < \varepsilon_0$ for some ε_0 independent of x and only depending on $\operatorname{dist}(N_1, O \setminus N_2)$. Therefore, (42) and (44) imply that if $\varepsilon < \varepsilon_0$, $x \in N_1$, and $\omega \in A_\varepsilon$, then

$$|\tau_x^\varepsilon - \tau_x^0| \le \frac{k^\varepsilon}{c_1} \tag{45}$$

and $X_\varepsilon(\tau_x^\varepsilon) \in \Gamma$.

Differentiating equations (12) and (13) with respect to x_i, $i = 1, \dots, n$, we obtain

$$-\frac{\varepsilon^2}{2}\Delta\partial_i v^\varepsilon + \langle Dv^\varepsilon - b, D\partial_i v^\varepsilon\rangle - \langle \partial_i b, Dv^\varepsilon\rangle = 0$$

and

$$-\frac{\varepsilon^2}{2}\Delta\partial_i v^0 - \langle b, D\partial_i v^0\rangle - \langle \partial_i b, Dv^0\rangle + \langle Dv^0, D\partial_i v^0\rangle = -\frac{\varepsilon^2}{2}\Delta\partial_i v^0.$$

Subtracting the above equations and dividing by ε^2 yields

$$\begin{aligned}-&\frac{\varepsilon^2}{2}\Delta\partial_i v_1^\varepsilon + \langle Dv^\varepsilon - b, D\partial_i v_1^\varepsilon\rangle + \frac{1}{\varepsilon^2}\langle Dv^\varepsilon, D\partial_i v^0\rangle \\ &\qquad\qquad - \langle \partial_i b, Dv_1^\varepsilon\rangle - \frac{1}{\varepsilon^2}\langle Dv^0, D\partial_i v^0\rangle = \frac{1}{2}\Delta\partial_i v^0.\end{aligned}$$

Using

$$\frac{1}{\varepsilon^2}\langle Dv^\varepsilon, D\partial_i v^0\rangle - \frac{1}{\varepsilon^2}\langle Dv^0(x), D\partial_i v^0\rangle = \langle D\partial_i v^0, Dv_1^\varepsilon\rangle$$

and $\partial_i b_0 = -\partial_i b + D\partial_i v^0$, we thus obtain

$$-\frac{\varepsilon^2}{2}\Delta\partial_i v_1^\varepsilon + \langle b_\varepsilon, D\partial_i v_1^\varepsilon\rangle + \langle \partial_i b_0, Dv_1^\varepsilon\rangle = \frac{1}{2}\Delta\partial_i v^0, \quad i = 1, \dots, n.$$

We can combine these n identities into one:

$$-\frac{\varepsilon^2}{2}\Delta(Dv_1^\varepsilon) + D(Dv_1^\varepsilon)b_\varepsilon + (Db_0)^* Dv_1^\varepsilon = \frac{1}{2}\Delta(Dv^0). \tag{46}$$

For $\varepsilon \geq 0$, we define the fundamental matrices Y^ε to be the solutions of

$$\begin{cases}\dot{Y}^\varepsilon(t) = -Y^\varepsilon(t)(Db_0)^*(X_\varepsilon(t)),\\ Y^\varepsilon(0) = I.\end{cases}$$

Let us denote $\tilde{\tau}_x^\varepsilon = \tau_x^\varepsilon \wedge T$. Using Itô's formula and (46) we obtain

$$\begin{aligned}&\mathsf{E}[Y^\varepsilon(\tilde{\tau}_x^\varepsilon)Dv_1^\varepsilon(X_\varepsilon(\tilde{\tau}_x^\varepsilon))]\\ &= Dv_1^\varepsilon(x) + \mathsf{E}\biggl[\int_0^{\tilde{\tau}_x^\varepsilon}\biggl(- Y^\varepsilon(t)(Db_0)^*(X_\varepsilon(t))Dv_1^\varepsilon(X_\varepsilon(t))\\ &\qquad - Y^\varepsilon(t)D(Dv_1^\varepsilon)(X_\varepsilon(t))b_\varepsilon(X_\varepsilon(t)) + \frac{\varepsilon^2}{2}Y^\varepsilon(t)\Delta(Dv_1^\varepsilon)(X_\varepsilon(t))\biggr)dt\biggr]\\ &= Dv_1^\varepsilon(x) - \frac{1}{2}\mathsf{E}\biggl[\int_0^{\tilde{\tau}_x^\varepsilon} Y^\varepsilon(t)\Delta(Dv^0)(X_\varepsilon(t))dt\biggr],\end{aligned}$$

which yields

$$Dv_1^\varepsilon(x) = \frac{1}{2}\mathsf{E}\left[\int_0^{\tilde{\tau}_x^\varepsilon} Y^\varepsilon(t)D(\Delta v^0)(X_\varepsilon(t))dt\right] + \mathsf{E}[Y^\varepsilon(\tilde{\tau}_x^\varepsilon)Dv_1^\varepsilon(X_\varepsilon(\tilde{\tau}_x^\varepsilon))]. \tag{47}$$

We will show that the right-hand side of (47) converges to

$$V(x) := \frac{1}{2}\int_0^{\tau_x^0} Y^0(t)D(\Delta v^0)(X_0(t))dt + Y^0(\tau_x^0)Dv_1(X_0(\tau_x^0)) \tag{48}$$

uniformly on N_1. To that end, we need to estimate the difference between the corresponding terms of (47) and (48).

We begin with the nonintegral terms. We notice that there is a number $C(T) > 0$ such that for every ω, ε, and $x \in N_1$

$$\sup_{0\leq t\leq \tilde{\tau}_x^\varepsilon} \|Y^\varepsilon(t)\| \leq C(T). \tag{49}$$

Next, there is $\varepsilon_0 > 0$ such that if $\varepsilon \in [0, \varepsilon_0)$, $\omega \in A_\varepsilon$, and $x \in N_1$, then (i) $\tilde{\tau}_x^\varepsilon = \tau_x^\varepsilon$ and (ii) $X_\varepsilon(\tilde{\tau}_x^\varepsilon) \in \Gamma$. Property (i) along with (44) and standard ODE theory implies that there are positive numbers $(k_1^\varepsilon)_{\varepsilon\in(0,\varepsilon_0)}$ such that $k_1^\varepsilon \to 0$ as $\varepsilon \to 0$, and

$$\sup_{0\le t\le \tilde{\tau}_x^\varepsilon} \|Y^\varepsilon(t) - Y^0(t)\| \le k_1^\varepsilon, \quad \omega \in A_\varepsilon, x \in N_1. \tag{50}$$

Property (ii) allows us to apply Lemma 5.11. So, along with (44), (45), (49), and (50), it implies that there are positive numbers $(k_2^\varepsilon)_{\varepsilon\in(0,\varepsilon_0)}$ such that $k_2^\varepsilon \to 0$ as $\varepsilon \to 0$, and

$$|Y^\varepsilon(\tilde{\tau}_x^\varepsilon) Dv_1^\varepsilon(X_\varepsilon(\tilde{\tau}_x^\varepsilon)) - Y^0(\tau_x^0) Dv_1(X_0(\tau_x^0))| \le k_2^\varepsilon, \quad \omega \in A_\varepsilon, x \in N_1. \tag{51}$$

Finally, we observe that there is a constant C_1, such that for all ε,

$$|Dv_1^\varepsilon(x)| \le \frac{C_1}{\varepsilon^2}, \quad x \in N_2. \tag{52}$$

The difference between the second terms of (47) and (48) can be estimated, due to (49), (51), (52) as

$$\left|\mathsf{E}[Y^\varepsilon(\tilde{\tau}_x^\varepsilon) Dv_1^\varepsilon(X_\varepsilon(\tilde{\tau}_x^\varepsilon))] - Y^0(\tau_x^0) Dv_1(X_0(\tau_x^0))\right| \le k_2^\varepsilon + C_2(1+\varepsilon^{-2})\mathsf{P}(A_\varepsilon^c)$$

for some constant C_2, and by (43) converges to 0 as $\varepsilon \to 0$.

To estimate the difference between the integral terms of (47) and (48), we write

$$\begin{aligned}
&\left|\mathsf{E}\int_0^{\tilde{\tau}_x^\varepsilon} Y^\varepsilon(t) D(\Delta v^0)(X_\varepsilon(t))dt - \int_0^{\tau_x^0} Y^0(t) D(\Delta v^0)(X_0(t))dt\right| \\
&\le \mathsf{E}\left[\mathbf{1}_{A_\varepsilon}\int_0^{\tau_x^\varepsilon\wedge\tau_x^0} \left|Y^\varepsilon(t) D(\Delta v^0)(X_\varepsilon(t)) - Y^0(t) D(\Delta v^0)(X_0(t))\right| dt\right] \\
&\quad + \mathsf{E}\left[\mathbf{1}_{A_\varepsilon}\left(\int_{\tau_x^\varepsilon\wedge\tau_x^0}^{\tau_x^\varepsilon} \left|Y^\varepsilon(t) D(\Delta v^0)(X_\varepsilon(t))\right| dt + \int_{\tau_x^\varepsilon\wedge\tau_x^0}^{\tau_x^0} \left|Y^0(t) D(\Delta v^0)(X_0(t))\right| dt\right)\right] \\
&\quad + \mathsf{E}\left[\mathbf{1}_{A_\varepsilon^c}\left(\int_0^{\tau_x^\varepsilon} \left|Y^\varepsilon(t) D(\Delta v^0)(X_\varepsilon(t))\right| dt + \int_0^{\tau_x^0} \left|Y^0(t) D(\Delta v^0)(X_0(t))\right| dt\right)\right].
\end{aligned}$$

Each of the terms on the r.h.s. uniformly converges to 0 in N_1. For the first term, this follows from (44), (49), and (50), for the second term — from (45) and (49), and for the third term — from (43) and (49).

Thus our claim of uniform convergence of Dv_1^ε to V follows. It remains to notice, differentiating (36), that Dv_1 satisfies in N the system

$$\langle b_0, D\partial_i v_1\rangle + \langle \partial_i b_0, Dv_1(x)\rangle = \frac{1}{2}\Delta\partial_i v^0, \quad i = 1, \dots, n,$$

for which the method of characteristics implies that $Dv_1(x) = V(x)$ in N_1. □

Appendix A. Doob's h-transform for conditioned diffusions

Here we provide a general and rigorous introduction to Doob's h-transform, computing the conditional distribution for diffusions conditioned on exit events. The material of this appendix is not highly original: connections of h-transform to conditioning, to the potential theory of elliptic PDEs, and to Martin boundaries are well known; see, e.g., [9, 31]. However, no rigorous and complete exposition of the main result of this section, Theorem A.6, is known to us, and we decided to include this appendix, also hoping that it will serve as a useful reference point. Our exposition draws from [6] and [34, Chapter 13]. We use minimal information from PDE theory. Other useful sources on Markov processes and diffusions are [33], [12], [24], [25].

Let us first introduce an abstract generalization of a diffusion process in a domain with absorption at the boundary of the domain. We will always work with homogeneous Markov processes, i.e., processes with transition mechanisms that do not depend on initial time.

Let O be a domain in $\mathbb{R}^n$. Let us equip the space $C = C([0,\infty),\bar{O})$ of continuous paths $(X(t))_{t\ge 0}$ with $\mathcal{B} = \mathcal{B}(C)$, the Borel σ-algebra with respect to locally uniform topology.

Suppose that $\mathsf{P}^t(x,dy)$ is a Markov transition kernel on $\bar{O}$. This means that (i) for all $t \ge 0$ and all $x \in \bar{O}$, $\mathsf{P}^t(x,\cdot)$ is a Borel probability measure on $\bar{O}$; (ii) for any $t \ge 0$ and any Borel set A, $\mathsf{P}^t(\cdot, A)$ is a Borel measurable function; (iii) $\mathsf{P}^0(x,dy) = \delta_x(dy)$; and (iv) the Chapman–Kolmogorov equations hold, i.e., for any $s,t \ge 0$ and any Borel set A,

$$\mathsf{P}^{t+s}(x,A) = \int_{\bar{O}} \mathsf{P}^t(x,dy)\mathsf{P}^s(y,A).$$

Let us make the following additional assumptions:

(1) There is a family of measures $(\mathsf{P}_x)_{x\in\bar{O}}$ on paths $(C,\mathcal{B})$ such that for each $x \in \bar{O}$, $\mathsf{P}_x\{X(0) = x\} = 1$ and under P_x the process X is a Markov process with a homogeneous transition probability $\mathsf{P}^t(x,dy)$.

(2) For all $x \in \partial O$ and all $t \ge 0$, $\mathsf{P}^t(x,dy) = \delta_x(dy)$.

Processes associated with such transition kernels or semigroups will be called continuous Markov processes on O with absorption at ∂O. We denote by E_x the expectation with respect to P_x. The semigroup (P^t) is defined by

$$\mathsf{P}^t f(x) = \int_{\bar{O}} \mathsf{P}^t(x,dy)f(y), \quad t \ge 0,\ x \in \bar{O},\ f \in \mathbb{B}(\bar{O}),$$

where $f \in \mathbb{B}(\bar{O})$ is the space of bounded measurable functions on $\bar{O}$.

For any $X \in C$, we denote by $\tau(X)$ the first time of exit on the boundary: $\tau = \inf\{t \ge 0 :\ X(t) \in \partial O\}$.

Let Γ be a measurable subset of ∂O. We introduce a trajectory set

$$C_\Gamma = \{X \in C :\ \tau(X) < \infty,\ X(\tau(X)) \in \Gamma\}$$

and a measurable bounded function

$$h(x) = h_\Gamma(x) = \mathsf{P}_x(C_\Gamma) = \lim_{n\to\infty} P^n(x,\Gamma), \quad x \in \bar{O}. \tag{53}$$

Let us assume that

$$h(x) > 0, \quad x \in O. \tag{54}$$

Our goal is then to describe the conditional measures $\mathsf{P}_{\Gamma,x}$ defined by

$$\mathsf{P}_{\Gamma,x}(A) = \mathsf{P}_x(A|C_\Gamma), \quad x \in O \cup \Gamma, A \in \mathcal{B}.$$

We will denote the expectation with respect to $\mathsf{P}_{\Gamma,x}$ by $\mathsf{E}_{\Gamma,x}$.

Denoting by $\mathcal{F}_t$ the natural filtration of the process X, we obtain for any $\mathcal{F}_t$-measurable random variable ξ:

$$\mathsf{E}_{\Gamma,x}\xi = \mathsf{E}_x\left[\xi\frac{\mathbf{1}_{C_\Gamma}}{h(x)}\right] = \mathsf{E}_x\left[\xi\mathsf{E}_x\left[\frac{\mathbf{1}_{C_\Gamma}}{h(x)}\Big|\mathcal{F}_t\right]\right] = \mathsf{E}_x\left[\xi\frac{h(X_t)}{h(x)}\right]. \tag{55}$$

Lemma A.1. *If $x \in O \cup \Gamma$, then $\mathsf{P}_{\Gamma,x}$ defines a continuous Markov process on $O \cup \Gamma$ with transition probability*

$$\mathsf{P}^t_\Gamma(x,dy) = \frac{h(y)}{h(x)}\mathsf{P}^t(x,dy). \tag{56}$$

Proof. The continuity is inherited from the original process, so it is sufficient to show that for any bounded measurable function f on $O \cup \Gamma$ and any $s,t \geq 0$, $x \in O \cup \Gamma$,

$$\mathsf{E}_{\Gamma,x}[f(X(s+t))|\mathcal{F}_s] = \mathsf{E}_x\left[f(X(s+t))\frac{h(X(s+t))}{h(X(s))}\Big|\ X(s)\right].$$

The right-hand side is $\mathcal{F}_s$-measurable, so we need to check that integrals with respect to $\mathsf{P}_{\Gamma,x}$ of both sides over any event $A \in \mathcal{F}_s$ coincide. By (55), the integral identity to check is

$$\mathsf{E}_x\left[\mathsf{E}_x\left[f(X(s+t))\frac{h(X(s+t))}{h(X(s))}\Big|\ X(s)\right]\mathbf{1}_A\frac{h(X(s))}{h(x)}\right] = \mathsf{E}_x\left[f(X(s+t))\mathbf{1}_A\frac{h(X(s+t))}{h(x)}\right].$$

To prove this identity we cancel the two instances of $h(X(s))$ on the left-hand side and, using the Markov property of X under P_x, replace conditioning with respect to $X(s)$ by conditioning with respect to $\mathcal{F}_s$. □

If $x \in \partial O \setminus \Gamma$, then the above construction does not make sense, and we simply set $\mathsf{P}^t_\Gamma(x,dy) = \delta_x(dy)$ for all $t \geq 0$. Combining this with (56) we see that the thus defined process is a continuous Markov process in O absorbed at ∂O, and the action of the semigroup associated with transition kernels P^t_Γ can be written as

$$\begin{aligned}\mathsf{P}^t_\Gamma f(x) &= \frac{\mathsf{E}_x f(X(t))h(X(t))}{h(x)}\mathbf{1}_{x\in O} + f(x)\mathbf{1}_{x\in\partial O}\\ &= \frac{\mathsf{P}^t(hf)(x)}{h(x)}\mathbf{1}_{x\in O} + f(x)\mathbf{1}_{x\in\partial O}, \quad t \geq 0,\ x \in \bar{O},\ f \in \mathbb{B}(\bar{O}).\end{aligned} \tag{57}$$

Let us say that a semigroup (P^t) defines a diffusion process in O absorbed at ∂O if in addition to properties (1) and (2) the following holds: for each point $x \in \bar{O}$ there is a positive semidefinite symmetric $n \times n$ matrix $a(x)$ and an n-dimensional vector $b(x)$ such that a and b are Borel functions on $\bar{O}$, bounded on every compact subset of O, and for every function $f \in C^2_0(O) = C^2(O) \cap C_c(O)$, the generator

$$Af = \lim_{t\to 0}\frac{\mathsf{P}^t f - f}{t}$$

is well defined in the space $\mathbb{B}(\bar{O})$ equipped with the sup-norm (i.e., the convergence on the right-hand side is uniform) and

$$Af(x) = Lf(x), \quad x \in O, \tag{58}$$

where we denote

$$Lf(x) = \sum_{i=1}^n b^i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^n a^{ij}(x)\partial_{ij} f(x). \tag{59}$$

Notice that we require (58) to hold for $x \in O$, although it is often convenient to have coefficients a and b defined on $\bar{O}$.

The identification of diffusion processes with solutions of stochastic equations and martingale problems is well known; see, e.g., [33], [12], [25]. The following theorem states that diffusion processes with absorption can also be viewed as solutions of Itô stochastic equations stopped upon reaching the boundary.

Theorem A.2. *Suppose the semigroup* (P^t) *defines a diffusion process on a domain* O *with absorption at* ∂O *and coefficients* a, b. *Let* σ *be a Borel measurable* $n \times n$ *matrix-valued function on* O *such that* $\sigma\sigma^* \equiv a$ *and such that* σ *is bounded on every compact subset of* O. *Then for any* $x \in O$ *there is an extension* $(\tilde{C}, \tilde{\mathcal{B}}, \tilde{\mathsf{P}}_x)$ *of the original probability space* $(C, \mathcal{B}, \mathsf{P}_x)$, *a filtration* $(\tilde{\mathcal{F}}_t)_{t\ge 0}$ *on* $\tilde{\mathcal{B}}$ *satisfying the usual conditions, and an* n*-dimensional Wiener process* W *w.r.t.* $(\tilde{\mathcal{F}}_t)$, *such that the coordinate process* X *is* $(\tilde{\mathcal{F}}_t)$*-adapted and, with probability one,* $X(0) = x$ *and*

$$dX(t) = b(X(t))\mathbf{1}_{\{X(t)\in O\}}dt + \sigma(X(t))\mathbf{1}_{\{X(t)\in O\}}dW(t), \quad t \ge 0.$$

If $a(x)$ *is nondegenerate for all* $x \in O$, *then* $(\tilde{C}, \tilde{\mathcal{B}}, \tilde{\mathsf{P}}_x)$ *may be taken to coincide with* $(C, \mathcal{B}, \mathsf{P}_x)$.

Sketch of the proof. We only indicate the changes one needs in adapting the proof of the same statement for the usual diffusion processes; see Proposition 5.4.6 of [25]. Let us consider a sequence $(U_m)_{m\in\mathbb{N}}$ of open sets such that $K_m = \bar{U}_m$ is a compact subset of O for every $m \in \mathbb{N}$, and $\bigcup_{m\in\mathbb{N}} U_m = O$. For every $m \in \mathbb{N}$, we can find a nonnegative bounded function $g_m \in C_0^2(O)$ such that $g_m \equiv 1$ on K_m. Then functions $f_{m,i}(x) = x_i g_m(x)$ and $f_{m,i,j}(x) = x_i x_j g_m(x)$ belong to the domain of A, and (58) holds for $f = f_{m,i}$ and $f = f_{m,i,j}$. Therefore, if X is the coordinate process under P_x, then, for $f = f_{m,i}$ and $f = f_{m,i,j}$, the process $f(X(t)) - \int_0^t Lf(X(s))ds$ is a bounded martingale under P_x (see, e.g., [12, Proposition 4.1.7]). As in the proof of Proposition 5.4.6 in [25], we can use this to derive that if $\tau_m = \inf\{t \ge 0 : X(t) \notin U_m\}$, then

$$M_m^i(t) = X^i(t \wedge \tau_m) - x_0^i - \int_0^{t\wedge\tau_m} b^i(X(s))ds$$

is a martingale with

$$\langle M_m^i, M_m^j \rangle = \int_0^{t\wedge\tau_m} a^{ij}(X(s))ds.$$

We can then follow the proof of Proposition 5.4.6 in [25] and use the multidimensional version of Doob's representation for continuous martingales (see [25, Theorem 3.4.2 and Remark 3.4.3]) to represent $M_m = (M_m^i)_{i=1}^n$ as

$$M_m(t) = \int_0^{t\wedge\tau_m} \sigma(X(s))dW_m(s)$$

for a Wiener process W_m on a filtered extension $(\tilde{C}_m, \tilde{\mathcal{B}}_m, (\tilde{\mathcal{F}}_{m,t})_{t\ge 0}, \tilde{\mathsf{P}}_{x,m})$ of the original probability space. In fact, one can choose the extended probability space and the Wiener process $W = W_m$ to be independent of m, thus obtaining

$$X(t \wedge \tau) = x_0 + \int_0^{t\wedge\tau} b(X(s))ds + \int_0^{t\wedge\tau} \sigma(X(s))dW(s), \quad s \ge 0,$$

where $\tau = \lim_{m\to\infty} \tau_m \in (0, \infty]$ and we used the continuity of trajectories of X. □

Theorem A.2 uses the existence of a Markov process with given coefficients of drift and diffusion as an assumption. In general, verifying this assumption may be a nontrivial issue. Moreover, for unbounded domains, without imposing some restrictions on the coefficients, such as global Lipschitzness of a and b or existence of an appropriately understood Lyapunov function, one cannot exclude finite time explosion of solutions. However, if the coefficients a and b are defined globally and define a diffusion process in $\mathbb{R}^n$, then stopping that process upon reaching ∂O naturally gives a Markov processes with all the required properties.

To derive the conditioned diffusion generator on $C_0^2(O)$, we will need to check that formula (58) holds true for $f = h$. The Markov property implies that h is harmonic for (P^t), i.e.,

$$\mathsf{P}^t h(x) = h(x), \quad x \in \bar{O},\ t > 0.$$

Therefore, Ah is well defined and identically equal to 0 on $\bar{O}$. However, $h \notin C_0^2(O)$, and *a priori* it is not even clear if $h \in C^2(O)$. To guarantee the latter one needs to make certain assumptions.

Lemma A.3. *Let the coefficients a, b of a diffusion process in O stopped at ∂O satisfy $a, b \in C^1(\bar{O})$ and $\det a(x) \neq 0$ for all $x \in \bar{O}$. Then $h \in C^2(O)$.*

We postpone the proof of this lemma to the end of this section.

Lemma A.4. *Let the conditions of Lemma* A.3 *hold. Then*

$$Ah(x) = Lh(x) = 0, \quad x \in O. \tag{60}$$

Proof. We cannot apply (58) directly since $h \notin C_0^2(O)$. Let us take $r > 0$ such that the ball $B_r(x)$ is contained in O. Then a and b are uniformly bounded in $B_r(x)$. Using Theorem A.2 to represent the process as a solution of an SDE and noticing that the behavior of the process until the first exit from $B_r(x)$ is entirely determined by the behavior of the coefficients in $B_r(x)$, one can derive from standard maximal inequalities for martingales that

$$\mathsf{P}_x \left\{ \sup_{s\in[0,t]} |X(s) - x| \geq r \right\} = o(t), \quad t \to 0.$$

In particular, $\mathsf{P}^t(x, B_r(x)^c) = o(t)$. Let us now find $f \in C_0^2(O)$ such that $0 \leq f(y) \leq 1$ for all $y \in \bar{O}$ and $f(y) = h(y)$ for all $y \in B_r(x)$. Then

$$\begin{aligned}\left| \frac{\mathsf{P}^t h(x) - h(x)}{t} - \frac{\mathsf{P}^t f(x) - f(x)}{t} \right| &= \left| \frac{\int_{\bar{O}} (h(y) - f(y)) \mathsf{P}^t(x, dy)}{t} \right| \\ &\leq \frac{2\mathsf{P}^t(x, B_r(x)^c)}{t} \to 0, \quad t \to 0.\end{aligned}$$

Therefore, $0 = Ah(x) = Af(x)$ and since the partial derivatives of h and f coincide at x, formula (60) is implied by (58) ☐

Lemma A.5. *Let the conditions of Lemma* A.3 *hold and assume that there is a point $x_0 \in O$ such that $h(x_0) > 0$. Then* (54) *holds.*

Proof. The lemma is a direct consequence of Lemma A.4 and the strong maximum principle (see [13, Theorem 3, page 349]) or Harnack's inequality (see [23, Corollary 9.25]). One can also give a more probabilistic argument: Lemma A.3 implies that h is positive in some neighborhood U of x_0. For any starting point $x \in O$ there is a continuous path $\gamma : [0, 1] \to O$ connecting x to x_0, which implies that $\mathsf{P}^1(x, U) > 0$. Now $h(x) > 0$ follows from the Markov property. ☐

Theorem A.6. *Let* P^t *define a diffusion process* X *in a domain* O*, with coefficients* $a, b \in C^1(\bar{O})$ *such that* $\det a(x) \neq 0$ *for all* $x \in \bar{O}$*. Let a measurable set* $\Gamma \subset \partial O$ *be such that* $h(x_0) > 0$ *for some* $x_0 \in O$*, where* h *is defined by* (53)*. Then the process* X *conditioned on exit from* O *through* Γ *is a diffusion process in* O *with coefficients* a *and* b_Γ*, where*

$$b_\Gamma(x) = b(x) + a(x)\frac{Dh(x)}{h(x)}, \quad x \in O.$$

Remark A.7. Under the conditions on a and b imposed by Theorem A.6, the condition $h(x_0) > 0$ is effectively a restriction on the "size" of Γ. The domain O itself is not required to be bounded, and ∂O can be arbitrarily irregular. One natural situation where this condition holds true is where Γ contains a smooth hypersurface Γ' such that $\Gamma' \cup O$ is path-connected. Also, one can often make sense of h-transform for semigroups and their generators even for conditioning on events of zero probability.

Proof of Theorem A.6. Lemma A.3 and condition (54) imply that if $f \in C_0^2(O)$, then $hf \in C_0^2(O)$, and $h(x)$ is bounded away from zero on the support of f. So, by (57), the generator A_Γ of the semigroup (P_Γ^t) is well defined on $C_0^2(O)$ and given by

$$A_\Gamma f = \lim_{t\to 0} \frac{h^{-1}\mathsf{P}^t(hf) - f}{t} = \lim_{t\to 0} \frac{\mathsf{P}^t(hf) - hf}{th} = \frac{A(fh)}{h}, \quad f \in C_0^2(O),$$

and a straightforward computation using Lemma A.4 produces

$$\begin{aligned} A_\Gamma f(x) &= \sum_{i=1}^n \left(b^i(x) + \frac{1}{h(x)} \sum_{j=1}^n a^{ij}(x)\partial_j h(x) \right) \partial_i f(x) + \frac{1}{2} \sum_{i,j=1}^n a^{ij}(x)\partial_{ij} f(x) \\ &= Lf(x) + \frac{1}{h(x)} \sum_{i,j=1}^n a^{ij}(x)\partial_j h(x)\partial_i f(x), \quad x \in O, \end{aligned}$$

which completes the proof. ◻

Proof of Lemma A.3. Let us first summarize the necessary information from the theory of Green's functions of elliptic PDE's in bounded smooth domains (see, e.g., [30, Sections 10, 16, 21, 36]).

Theorem A.8. *Let* B *and* B' *be two balls in* $\mathbb{R}^n$ *and let* $\bar{B} \subset B'$*. Let* $a, b \in C^1(B')$ *and let* $a(x)$ *be nondegenerate for all* $x \in B'$*. Then there is a function* $K_B \in C(B \times \partial B)$ *such that for any* $\phi \in C(\partial B)$ *there is a unique solution* $v \in C(\bar{B}) \cap C^2(B)$ *of*

$$\begin{cases} Lv(x) = 0, & x \in B, \\ v(x) = \phi(x), & x \in \partial B \end{cases}$$

(*where* L *is defined in* (59))*, and the solution can be represented as*

$$v(x) = \int_{\partial B} K_B(x,y)\phi(y)\mu_B(dy), \quad x \in B, \tag{61}$$

where $\mu_B(dy)$ *denotes the surface area on the sphere* ∂B*. If* $\phi \in C^3(\partial B)$*, then* $v \in C^2(\bar{B})$ *and it can be extended to a function* $v \in C_0^2(O)$*.*

Let us now recall the connection with diffusions.

Consider transition probabilities (P^t) and the associated Markov family (P^t_x) of diffusion processes with generator A satisfying the conditions of Theorem A.6. Let B be a ball such that $\bar{B} \subset O$ and let $x \in B$. Under P_x, we define $\tau_B = \inf\{t \geq 0 : \ X(t) \in \partial B\}$. Under the assumptions of Theorem A.6, $\mathsf{P}_x\{\tau_B < \infty\} = 1$; see, e.g., [5, Proposition 8.2].

For any $\phi \in C^3(\partial B)$, the associated function v given by Theorem A.8 and extended to a function in $C^2_0(O)$ belongs to the domain of the generator A, and so $v(X(t)) - \int_0^t Av(X(s))ds$ is a bounded martingale under P_x (see, e.g., [12, Proposition 4.1.7]). Since $v \in C^2_0(O)$, we have $Av(X(t))\mathbf{1}_{t\leq\tau_B} = Lv(X(t))\mathbf{1}_{t\leq\tau_B} = 0$, and Doob's optional sampling theorem [12, Theorem 2.2.13] implies

$$v(x) = \mathsf{E}_x\phi(X(\tau_B)), \quad x \in B.$$

Comparing this to (61), we conclude that for any ball $B \subset O$ and any starting point $x \in O$, $K_B(x, \cdot)$ is the density of the distribution of $X(\tau_B)$ with respect to the surface area on ∂B, so that for any bounded measurable function $f : \partial B \to \mathbb{R}$,

$$\mathsf{E}_x f(X(\tau_B)) = \int_{\partial B} f(y)K_B(x, y)\mu_B(dy). \tag{62}$$

Under the conditions of Theorem A.6 the Feller property holds true (see, e.g., [12, Theorem 8.1.4]) and hence due to the continuity of trajectories, so does the strong Markov property (see, e.g., [12, Theorem 4.2.7]). The latter implies $h(x) = \mathsf{E}_x h(X(\tau_B))$ for any ball $B \subset O$ and any $x \in B$. This, along with (62), implies

$$h(x) = \int_{\partial B} h(y)K_B(x, y)\mu_B(dy), \quad x \in B.$$

In the right-hand side, h is bounded and, for any open set U compactly contained in B, the function K_B is uniformly continuous on $U \times \partial B$. Therefore, $h \in C(B)$. Let us take another ball B' such that $\partial B' \subset B$. Then $h \in C(\partial B')$ and, for $\tau' = \inf\{t \geq 0 : \ X(t) \in \partial B'\}$,

$$h(x) = \mathsf{E}_x h(X(\tau')) = \int_{\partial B'} h(y)K_{B'}(x, y)\mu_{B'}(dy), \quad x \in B'.$$

Theorem A.8 implies $h \in C^2(B')$. Since one can choose balls B and B' to contain any given point in O, we conclude that $h \in C^2(O)$. □

Alternative proof of Lemma A.3. The proof of Lemma A.3 was based on the existence of a Green's function. We want to present an alternative approach based on the strong Feller property of diffusions (which heuristically means that transition probabilities have nice densities). Let us recall (see, e.g., [33, Theorem 7.2.4]) that if coefficients a and b are bounded on $\mathbb{R}^n$, $a \in C(\mathbb{R}^n)$, and $\det a > c_0$ for some $c_0 > 0$, then the corresponding diffusion process is strong Feller, i.e., for any bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and any time $t > 0$, the function $\mathsf{P}^t f$ defined by $\mathsf{P}^t f(x) = \int_{\mathbb{R}^n} \mathsf{P}^t(x, dy)f(y)$ is continuous. (Another proof of the strong Feller property when the coefficients are Lipschitz continuous can be deduced from the results of [21, Chapter 5, Sections 4 and 5].) So, for any open balls B_1 and B_2 such that $\bar{B}_1 \subset \bar{B}_2$ and $\bar{B}_2 \subset O$, we can find bounded coefficients $\tilde{a}, \tilde{b} \in C^1(\mathbb{R}^n)$ such that

$$a(x) = \tilde{a}(x), \quad b(x) = \tilde{b}(x), \quad x \in \bar{B}_2, \tag{63}$$

and the diffusion associated with $\tilde{a}, \tilde{b}$ is strong Feller on $\mathbb{R}^n$. Let us also extend the function h to a bounded measurable function defined on $\mathbb{R}^n$. For any $x \in \bar{B}_1$, we can use Theorem A.2 to realize the diffusion corresponding to coefficients a, b with absorption at ∂O as a solution to an SDE driven by a Wiener process. We can now construct a strong solution of the SDE driven by the same Wiener process, with coefficients $\tilde{a}, \tilde{b}$ and starting point x on the same probability space. Let us keep P_x and E_x as the notation for the respective probability and expectation on this probability space and denote the diffusion processes by $X(t)$ and $\tilde{X}(t)$. These processes coincide at least up to a random time $\nu = \inf\{t \geq 0 : X(t) \in \partial B_2\}$.

We know that the measurable function h satisfies

$$h(x) = \mathsf{E}_x h(X(\tilde{\tau}))$$

for every stopping time $\tilde{\tau} < \tau$. We need to show that h is continuous in O. Denote $\nu_t := t \wedge \nu$. Due to (63), for any $t \geq 0$,

$$h(x) = \mathsf{E}_x h(X(\nu_t)) = \alpha_t(x) + \beta_t(x), \quad x \in B_1, \tag{64}$$

where $\alpha_t(x) = \mathsf{E}_x h(\tilde{X}(t))$ and $\beta_t(x) = \mathsf{E}_x h(X(\nu_t))\mathbf{1}_{\{\nu<t\}} - \mathsf{E}_x h(\tilde{X}(t))\mathbf{1}_{\{\nu<t\}}$. The strong Feller property for $\tilde{X}$ implies that $\alpha_t(\cdot)$ is continuous on B_1. For the second term we have

$$|\beta_t(x)| \leq 2\mathsf{P}_x\{\nu < t\},$$

and the standard maximal inequalities imply that, as $t \to 0$, $\beta_t(\cdot)$ converges to 0 uniformly in B_1. Therefore, due to (64), h is continuous on B_1 being a uniform limit of continuous functions. Since the choice of B_1 was arbitrary, h is continuous on O.

Once we know that h is continuous in O, for every open ball B such that $\bar{B} \subset O$, the problem

$$\begin{cases} Lv(x) = 0, & x \in B, \\ v(x) = h(x), & x \in \partial B, \end{cases}$$

has a unique solution $v \in C^2(B) \cap C(\bar{B})$ (see, e.g., [23, Theorem 6.13]). For any $x \in B$, we can use the Itô formula along with the martingale property to see

$$h(x) = \mathsf{E}_x h(X(\tau_B)) = \mathsf{E}_x v(X(\tau_B)) = v(x) + \mathsf{E}_x \left[\int_0^{\tau_B} Lv(X(t))dt \right] = v(x),$$

where $\tau_B = \inf\{t \geq 0 : \ X(t) \in \partial B\}$. So, h coincides with v in B. Therefore, $h \in C^2(B)$. Since the choice of B is arbitrary, the lemma follows. □

Acknowledgments

The authors would like to thank W. H. Fleming for his comments about the first version of the manuscript, and the anonymous referee for his/her careful reading of the manuscript and remarks that improved the quality of the paper.

References

[1] S. A. Almada Monter and Y. Bakhtin, *Normal forms approach to diffusion near hyperbolic equilibria*, Nonlinearity **24** (2011), no. 6, 1883–1907, DOI 10.1088/0951-7715/24/6/011. MR2802310 (2012j:60143)

[2] S. A. A. Monter and Y. Bakhtin, *Scaling limit for the diffusion exit problem in the Levinson case*, Stochastic Process. Appl. **121** (2011), no. 1, 24–37, DOI 10.1016/j.spa.2010.09.002. MR2739004 (2011m:37083)

[3] Y. Bakhtin, *Small noise limit for diffusions near heteroclinic networks*, Dyn. Syst. **25** (2010), no. 3, 413–431, DOI 10.1080/14689367.2010.482520. MR2731621 (2012e:37104)

[4] Y. Bakhtin, *Noisy heteroclinic networks*, Probab. Theory Related Fields **150** (2011), no. 1-2, 1–42, DOI 10.1007/s00440-010-0264-0. MR2800902 (2012f:60274)

[5] R. F. Bass, *Diffusions and elliptic operators*, Probability and its Applications (New York), Springer-Verlag, New York, 1998. MR1483890 (99h:60136)

[6] A. Bloemendal, *Doob's h-transform: theory and examples*, http://www.math.harvard.edu/~alexb/rm/Doob.pdf (2010).

[7] C. Carathéodory, *Calculus of variations and partial differential equations of the first order. Part I: Partial differential equations of the first order*, Translated by Robert B. Dean and Julius J. Brandstatter, Holden-Day, Inc., San Francisco-London-Amsterdam, 1965. MR0192372 (33 #597)

[8] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) **27** (1992), no. 1, 1–67, DOI 10.1090/S0273-0979-1992-00266-5. MR1118699 (92j:35050)

[9] J. L. Doob, *Classical potential theory and its probabilistic counterpart*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 262, Springer-Verlag, New York, 1984. MR731258 (85k:31001)

[10] W. E and E. Vanden-Eijnden, *Towards a theory of transition paths*, J. Stat. Phys. **123** (2006), no. 3, 503–523, DOI 10.1007/s10955-005-9003-9. MR2252154 (2008g:60223)

[11] Weinan E and Eric Vanden-Eijnden, *Transition-path theory and path-finding algorithms for the study of rare events*, Annual Review of Physical Chemistry **61** (2010), no. 1, 391–420.

[12] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986. Characterization and convergence. MR838085 (88a:60130)

[13] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, RI, 2010. MR2597943 (2011c:35002)

[14] L. C. Evans and H. Ishii, *A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities* (English, with French summary), Ann. Inst. H. Poincaré Anal. Non Linéaire **2** (1985), no. 1, 1–20. MR781589 (86f:35183)

[15] W. H. Fleming, *Stochastic control for small noise intensities*, SIAM J. Control **9** (1971), 473–517. MR0304045 (46 #3180)

[16] W. H. Fleming, *Exit probabilities and optimal stochastic control*, Appl. Math. Optim. **4** (1977/78), no. 4, 329–346, DOI 10.1007/BF01442148. MR512217 (80h:60100)

[17] W. H. Fleming and M. R. James, *Asymptotic series and exit time probabilities*, Ann. Probab. **20** (1992), no. 3, 1369–1384. MR1175266 (93k:60069)

[18] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*, Applications of Mathematics (New York), vol. 25, Springer-Verlag, New York, 1993. MR1199811 (94e:93004)

[19] W. H. Fleming and P. E. Souganidis, *Asymptotic series and the method of vanishing viscosity*, Indiana Univ. Math. J. **35** (1986), no. 2, 425–447, DOI 10.1512/iumj.1986.35.35026. MR833404 (87i:35080a)

[20] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 260, Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szücs. MR722136 (85a:60064)

[21] A. Friedman, *Stochastic differential equations and applications. Vol. 1*, Probability and Mathematical Statistics, vol. 28, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975. MR0494490 (58 #13350a)

[22] P. R. Garabedian, *Partial differential equations*, AMS Chelsea Publishing, Providence, RI, 1998. Reprint of the 1964 original. MR1657375 (99f:35001)

[23] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition. MR1814364 (2001k:35004)

[24] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, 2nd ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989. MR1011252 (90m:60069)

[25] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR1121940 (92h:60127)

[26] P.-L. Lions, *Generalized solutions of Hamilton-Jacobi equations*, Research Notes in Mathematics, vol. 69, Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982. MR667669 (84a:49038)

[27] J. Lu and J. Nolen, *Reactive trajectories and the transition path process*, ArXiv e-prints (2013).

[28] K. Maurin, *Analysis. Part I*, D. Reidel Publishing Co., Dordrecht-Boston, Mass.; PWN—Polish Scientific Publishers; Warsaw, 1976. Elements; Translated from the Polish by Eugene Lepa. MR0507436 (58 #22444)

[29] P. Metzner, C. Schutte, and E. Vanden-Eijnden, *Illustration of transition path theory on a collection of simple examples*, The Journal of Chemical Physics **125** (2006), no. 8, 084110.

[30] C. Miranda, *Partial differential equations of elliptic type*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2, Springer-Verlag, New York-Berlin, 1970. Second revised edition. Translated from the Italian by Zane C. Motteler. MR0284700 (44 #1924)

[31] R. G. Pinsky, *Positive harmonic functions and diffusion*, Cambridge Studies in Advanced Mathematics, vol. 45, Cambridge University Press, Cambridge, 1995. MR1326606 (96m:60179)

[32] S. J. Sheu, *Asymptotic behavior of the invariant density of a diffusion Markov process with small diffusion*, SIAM J. Math. Anal. **17** (1986), no. 2, 451–460, DOI 10.1137/0517034. MR826705 (87f:60124)

[33] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 233, Springer-Verlag, Berlin-New York, 1979. MR532498 (81f:60108)

[34] A. D. Wentzell, *A course in the theory of stochastic processes*, McGraw-Hill International Book Co., New York, 1981. Translated from the Russian by S. Chomet; With a foreword by K. L. Chung. MR781738 (86g:60004)

Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012

School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, Georgia 30332-0160