The Annals of Probability
1989, Vol. 17, No. 2, 559–572

THE AVERAGING PRINCIPLE FOR DIFFUSIONS WITH A SMALL PARAMETER IN THE CASE OF A NONCHARACTERISTIC BOUNDARY[1]

BY ROSS PINSKY

Technion—Israel Institute of Technology

Let $L_\varepsilon = \varepsilon L_0 + L_1$, where L_0 is a nondegenerate elliptic operator on R^2 and $L_1 = \frac{1}{2}A(r,\theta)\partial^2/\partial\theta^2 + B(r,\theta)\partial/\partial\theta$. We assume that for fixed r, L_1 generates a positive recurrent diffusion on the circle with invariant measure $\mu_r(d\theta)$. Let $X^\varepsilon(t) = (r^\varepsilon(t), \theta^\varepsilon(t))$ denote the diffusion generated by $(1/\varepsilon)L_\varepsilon$ and let u_ε be the solution to the Dirichlet problem $L_\varepsilon u = 0$ on D and $u = f$ on ∂D, where $D = \{x: r_1 < |x| < r_2\}$ and f is continuous. Then $u_\varepsilon(x) = E_x f(X^\varepsilon(\tau_D^\varepsilon))$, where τ_D^ε is the first exit time from D. By the averaging principle, the process $r^\varepsilon(t)$ converges weakly to the process $r^0(t)$ generated by $\overline{L}_0$, the operator obtained from L_0, by restricting to functions depending only on r and averaging the coefficients with respect to $\mu_r(d\theta)$. Furthermore, $P_x(\theta^\varepsilon(\tau_D^\varepsilon \in d\theta|\tau_{r_i}^\varepsilon < \tau_{r_j}^\varepsilon)$ converges weakly as $\varepsilon \to 0$, to a measure which can be calculated in terms of $\mu_{r_i}(\partial\theta)$ and the diffusion matrix of L_0, where τ_r^ε denotes the hitting time of the circle of radius r and $(i, j) = (1,2)$ or $(2,1)$. The above information allows one to evaluate the limiting distribution of $(r^\varepsilon(\tau_D^\varepsilon), \theta^\varepsilon(\tau_D^\varepsilon))$ and thus also the asymptotics of $u_\varepsilon(x)$. Call θ the fast variable and r the slow variable. In this paper we investigate what happens to the averaging principle in the case that the boundary of D is no longer characteristic for the equation slow variable = constant.

1. Let

$$L_\varepsilon = \varepsilon \sum_{i,j=1}^{d} a_{ij}\frac{\partial^2}{\partial x_i\,\partial x_j} + \varepsilon\sum_{i=1}^{d} b_i\frac{\partial}{\partial x_i} + \sum_{i=1}^{d} B_i\frac{\partial}{\partial x_i}$$

in a smooth bounded region $D \subset R^d$ and consider the Dirichlet problem

$$\begin{aligned} L_\varepsilon u &= 0 \quad \text{in } D,\\ u|_{\partial D} &= f, \end{aligned} \tag{1.1}$$

where f is continuous. Then the solution u_ε has the stochastic representation $u_\varepsilon(x) = E_x f(X^\varepsilon(\tau_D^\varepsilon))$, where $X^\varepsilon(t)$ is the diffusion process generated by L_ε and τ_D^ε is the first exit time of $X^\varepsilon(t)$ from D. Thus, studying the asymptotic behavior of u_ε is equivalent to studying the asymptotic behavior of the exit distribution $x^\varepsilon(\tau_D^\varepsilon)$. The methodology employed to study this problem differs radically for each of three "extreme" cases. The general case, which may be very complicated, is, in essence, some combination of the three extreme case with the possible addition of certain degeneracies.

Received June 1987; revised May 1988.

[1]Research supported by the Fund for the Promotion of Research at the Technion.

AMS 1980 *subject classifications*. Primary 60J60; secondary 60H10.

Key words and phrases. Diffusion processes, small parameter diffusion, random perturbation of dynamical systems, the averaging principle.

We describe very briefly two of the cases and then go into a little more detail with respect to the third one, the one that is relevant to the present paper. [See Freidlin ([1], Chapter 4) for a good survey and for references.]

CASE 1. Assume that starting from each $x \in D$, the solution $X_x(t)$ of $X'(t) = B(X(t))$ with $X(0) = x$ exits D in a finite amount of time. Call $T(x)$ the exit time of $X_x(t)$ and let $\widetilde{\partial D} = \{X_x(T(x)), x \in D\}$. $\widetilde{\partial D}$ is called the regular part of the boundary. Also assume that $(n(X_x(T(x))), B(X_x(T(x)))) > 0$, for all $x \in \widetilde{\partial D}$, where $n(z)$ denotes the outward unit normal at $z \in \partial D$. If the above occurs. B is said to fulfill the Levinson condition and $\lim_{\varepsilon \to 0} u_\varepsilon(x) = u_0(x)$, for all $x \in D \cup \widetilde{\partial D}$, where $u_0(x)$ solves

$$B \cdot \nabla u_0 = 0 \quad \text{in } D,$$
$$u_0|_{\widetilde{\partial D}} = f.$$

This is the simplest of the extreme cases.

CASE 2. Assume that there exists a unique stable equilibrium point $x_0 \in D$, such that $X_x(t)$ as given above approaches x_0 as $t \to \infty$ for every $x \in D$. To handle this case, one must turn to the large deviations theory of Wentzell and Freidlin for small parameter diffusion. This case is the most difficult one.

CASE 3 (The averaging principle). Whereas in the above two cases, the drift B either aided or hindered the process with respect to exiting from D, we now assume that the drift is neutral in this regard. For example, let $D = \{x \in R^2 : r_1 < |x| < r_2\}$ and let

$$\begin{aligned} L_\varepsilon = \varepsilon\bigg(&\frac{1}{2}a_{11}(r,\theta)\frac{\partial^2}{\partial r^2} + a_{12}(r,\theta)\frac{\partial^2}{\partial r \partial\theta} + \frac{1}{2}a_{22}(r,\theta)\frac{\partial^2}{\partial\theta^2} \\ &+ b_1(r,\theta)\frac{\partial}{\partial r} + b_2(r,\theta)\frac{\partial}{\partial\theta}\bigg) \\ &+ B(r,\theta)\frac{\partial}{\partial\theta} + \frac{1}{2}A(r,\theta)\frac{\partial^2}{\partial\theta^2}, \end{aligned} \tag{1.2}$$

where $B > 0$ on the set $\{(r,\theta) \in \bar{D} : A(r,\theta) = 0\}$. [Note that the term $\frac{1}{2}A(r,\theta)\partial^2/\partial\theta^2$ actually makes the L_ε here more general than before.] Then for small ε, the process will loop around many times in the θ-direction before it moves much in the r-direction. In this case, an averaging principle occurs. For any $r \in [r_1, r_2]$, consider the process on the circle generated by $\frac{1}{2}A(r,\theta)\, d^2/d\theta^2 + B(r,\theta)\, d/d\theta$ and call its invariant probability density $\mu(r,\theta)$. Let $X^\varepsilon(t) = (r^\varepsilon(t), \theta^\varepsilon(t))$ denote the process generated by L_ε and let $\hat{X}^\varepsilon(t) = (\hat{r}^\varepsilon(t), \hat{\theta}^\varepsilon(t))$ denote the process generated by $1/\varepsilon L_\varepsilon$. Then the process $\hat{r}^\varepsilon(t)$, frozen upon exiting D, will converge weakly to the process

$\bar{r}(t)$, frozen upon exiting D, generated by

$$\bar{L} = \frac{1}{2}\bar{a}_{11}(r)\frac{d^2}{dr^2} + \bar{b}_1(r)\frac{d}{dr}, \tag{1.3}$$

where

$$\bar{a}_{11}(r) = \int_0^{2\pi} a_{11}(r,\theta)\mu(r,\theta)\,d\theta$$

and

$$\bar{b}_1(r) = \int_0^{2\pi} b_1(r,\theta)\mu(r,\theta)\,d\theta.$$

Thus, it follows quite readily that the probability that $\hat{r}^\varepsilon(t)$ reaches r_1 before r_2 converges to the probability that $\bar{r}(t)$ reaches r_1 before r_2. Since $\hat{r}^\varepsilon(t)$ and $r^\varepsilon(t)$ are really the same process run at different speeds, in fact we can also conclude that the probability that $r^\varepsilon(t)$ reaches r_1 before r_2 also converges to the probability that $\bar{r}(t)$ reaches r_1 before r_2.

Consequently, in the case that the boundary function f depends only on r, it is easy to obtain $\lim_{\varepsilon\to 0} u_\varepsilon(x) = u_0(|x|)$, for all $x \in D$, where $u_0(r)$ satisfies

$$\bar{L}u_0 = 0, \qquad r_1 < r < r_2,$$
$$u_0(r_i) = f(r_i), \qquad i = 1,2.$$

In the general case that f may depend on θ, the explicit exit distribution on each circle is required. Khasminskii [2] has shown that

$$\lim_{\varepsilon\to 0} P_x\big(\hat{X}^\varepsilon(\tau_D^\varepsilon) \in (r_i, d\theta) \,|\, |\hat{X}^\varepsilon(\tau_D^\varepsilon)| = r_i\big) = \nu_{r_i}(\theta), \qquad i = 1,2, \tag{1.4}$$

where $\nu_r(\theta)$ is given by

$$\nu_r(\theta) = \frac{\mu(r,\theta)a_{11}(r,\theta)}{\int_0^{2\pi}\mu(r,s)a_{11}(r,s)\,ds}. \tag{1.5}$$

Consequently, it follows easily that $\lim_{\varepsilon\to 0} u_\varepsilon(x) = u_0(|x|)$, for all $x \in D$, where u_0 solves

$$\bar{L}_0 = 0, \qquad r_1 < r < r_2,$$
$$u(r_i) = \int_0^{2\pi} f(r_i,\theta)\nu_{r_i}(\theta)\,d\theta.$$

We will refer to r as the "slow" variable and to θ as the "fast" variable.

For the general formulation of the averaging principle and its proof under certain simplifying assumptions, one may consult Freidlin's book. The theorem is due to Khasminskii and its complete proof maybe found in [3].

Now, in general, one can make a small perturbation of the boundary without destroying the property of belonging to Case 1 or Case 2; this is manifestly false for Case 3. Indeed, the formulation depended on the boundary being characteristic for the equation: slow variable = constant. In this paper, we will study a

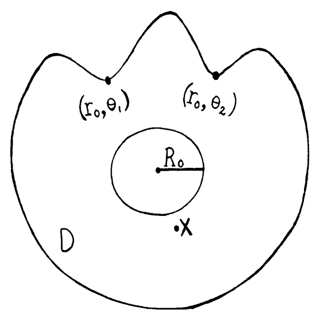

FIG. 1.

situation which is actually a combination of Cases 1 and 3; however, it will be more natural to regard it as a case of the averaging principle when the boundary is noncharacteristic.

We will consider a region D as in Figure 1. D is a smooth annular region which has been perturbed by two depressions which cause the outer boundary to be noncharacteristic with respect to the equation slow variable = constant. To be more precise, we formulate

ASSUMPTION 1. D is a bounded C^1-domain lying outside the ball of radius R_0 and satisfies the following conditions (see Figure 1): (i) $r_0 = \inf_{x \in D^c, |x| > R_0}|x|$; (ii) $(r_0, \theta_i) \in \partial D$, $i = 1, 2$, and if $x \in \partial D$ and $x \neq (r_0, \theta_i))$, then $|x| > r_0$.

We will consider the operator L_ε as in (1.2). For convenience, we will assume that all the coefficients are defined on all of R^2. We will also require

ASSUMPTION 2. (i) $\sqrt{A}$ and all the coefficients of L_ε are uniformly Lipschitz on compacts; (ii) $B > 0$ on $\{(r, \theta): A(r, \theta) = 0\}$ and (iii) $a_{ij}(r, \theta)$ is positive definite for each $r > 0$ and $\theta \in S^1$.

Assumption 2(ii) guarantees that the process on the circle generated by $\frac{1}{2}A(r, \theta)d^2/d\theta^2 + B(r, \theta)d/d\theta$ is positive recurrent for all r.

We will now analyse the asymptotics of the distribution of $X^\varepsilon(\tau_D^\varepsilon)$, starting from $z \in D$. First consider $z = x_1$ or $z = x_2$ as in Figure 2. Our intuition concerning the fast and slow variables should lead us heuristically to the following conclusion:

PROPOSITION. *Starting from x_i, $i = 1, 2$, the exit distribution $X^\varepsilon(\tau_D^\varepsilon)$ converges weakly as $\varepsilon \to 0$ to $p_i\delta_{y_{i1}} + (1 - p_i)\delta_{y_{i2}}$, where p_i is given as follows. Let ϕ project $x = (r, \theta)$ onto the unit circle S^1. Then $p_i = \tilde{P}_{\phi(x_i)}(\tau_{\phi(y_{i1})} < \tau_{\phi(y_{i2})})$, where $\tilde{P}_{\phi(x_i)}$ is the measure on paths associated with the process on the circle starting from $\phi(x_i)$ and generated by $\frac{1}{2}A(|x_i|, \theta)d^2/d\theta^2 + B(|x_i|, \theta)d/d\theta$, and τ_c denotes the first hitting time of c.*

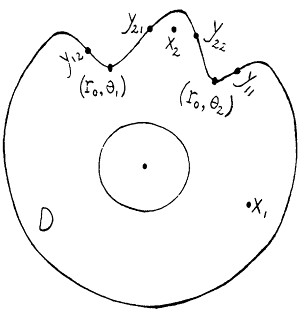

FIG. 2. $|x_1| = |y_{11}| = |y_{12}| > r_0$ *and* $|x_2| = |y_{21}| = |y_{22}| > r_0$.

The proof of the proposition follows the same methodology that we employ to prove that starting from (r_0, θ), the support of the exit distribution converges weakly to the two points (r_0, θ_1) and (r_0, θ_2) (see the proof of the theorem); thus we leave it to the reader. With this proposition, one can determine the limit of $u_\varepsilon(x_i)$ as $\varepsilon \to 0$, where u_ε solves (1.1) and L_ε is as in (1.2).

Now consider what happens starting from a point x with $R_0 < |x| \le r_0$. From the averaging principle, it follows that the probability that $X_\varepsilon(t)$ reaches the circle of radius r_0 before reaching the inner circle of radius R_0 converges to the probability, $p(|x|)$, that the radial process, $\bar{r}(t)$, generated by (1.3) and starting from $|x|$ reaches r_0 before reaching R_0. For later use, we note this as

$$p(r) = P_r(\bar{r}(t) \text{ reaches } r_0 \text{ before } R_0). \tag{1.6}$$

It also follows from (1.4) that, starting from $R_0 < |x| < r_0$, the distribution of $\theta^\varepsilon(\tau_D^\varepsilon)$, conditioned on $r^\varepsilon(t)$ reaching r_0 before R_0 (R_0 before r_0), converges to $\nu_{r_0}(\theta)\,d\theta(\nu_{R_0}(\theta)\,d\theta)$. Now, again using the intuition of the fast and slow variables, one is lead heuristically to the conclusion that the limiting support of the exit distribution is the circle of radius R_0 with the two points (r_0, θ_1) and (r_0, θ_2) adjoined and that the limiting probability of exiting D through the inner circle is $1 - p(|x|)$. What is not clear is how the limiting probability $p(|x|)$ is distributed over the two points (r_0, θ_1) and (r_0, θ_2). The main result of this paper is to identify this limiting distribution. This will in turn allow us to identify $\lim_{\varepsilon \to 0} u_\varepsilon(x)$ for $R_0 < |x| \le r_0$.

THEOREM. *Let D satisfy Assumption* 1 *and let L_ε be as in* (1.2) *with coefficients satisfying Assumption* 2.

(i) *Let $x \in D$ satisfy $R_0 < |x| < r_0$ and let $X^\varepsilon(t) = (r^\varepsilon(t), \theta^\varepsilon(t))$ denote the process starting from x and generated by L_ε. Define $\tau_D^\varepsilon = \inf\{t \ge 0: X^\varepsilon(t) \notin D\}$. Then the distribution of $X^\varepsilon(\tau_D^\varepsilon)$ converges weakly as $\varepsilon \to 0$ to*

$$(1 - p(|x|))\delta_{R_0}(dr)\nu_{R_0}(\theta)\,d\theta + p(|x|)\delta_{r_0}(dr)\big(\rho\delta_{\theta_1}(d\theta) + (1 - \rho)\delta_{\theta_2}(d\theta)\big),$$

where $p(r)$ is as in (1.6), *$\nu_r(\theta)$ is as in* (1.5) *and ρ is given as follows. Let $\hat{\theta}(t, \theta)$ denote the diffusion on the circle starting from θ and generated by*

$\frac{1}{2}A(r_0,\theta)\,d^2/d\theta^2 + B(r_0,\theta)\,d/d\theta$ and represent $\hat{\theta}(t,\theta)$ by

$$\hat{\theta}(t,\theta) = \theta + \int_0^t \sqrt{A\big(r_0,\hat{\theta}(s,\theta)\big)}\,d\hat{\omega}(s) + \int_0^t B\big(r_0,\hat{\theta}(s,\theta)\big)\,ds,$$

where $\hat{\omega}(t)$ is a Brownian motion on a probability space $(\Omega,\mathscr{F},P)$. Define the stochastic integral

$$\hat{r}_\theta(t) = \int_0^t \Big(\sigma_{11}^2\big(r_0,\hat{\theta}(s,\theta)\big) + \sigma_{12}^2\big(r_0,\hat{\theta}(s,\theta)\big)\Big)^{1/2}\,d\hat{\hat{\omega}}(s),$$

where $\sigma = \{\sigma_{ij}\}$ is the positive square root of $a = \{a_{ij}\}$ and $\hat{\hat{\omega}}(t)$ is a Brownian motion on $(\Omega,\mathscr{F},P)$ independent of $\hat{\omega}(t)$. Then $\rho = \int_0^{2\pi} h(\theta)\nu_{r_0}(\theta)\,d\theta$, where $h(\theta) = P(\hat{\theta}(\tau,\theta) = \theta_1)$ and $\tau = \inf\{t \geq 0 \colon \hat{\theta}(t,\theta) = \theta_1$ or θ_2 and $\hat{r}_\theta(t) \geq 0\}$.

(ii) *If $x = (r_0,\theta_0)$ for some θ_0, then the distribution of $x^\varepsilon(\tau_D^\varepsilon)$ converges weakly as $\varepsilon \to 0$ to $\delta_{r_0}(dr)(h(\theta_0)\delta_{\theta_1}(d\theta) + (1-h(\theta_0))\delta_{\theta_2}(d\theta))$.*

From the theorem, we immediately obtain

COROLLARY. (i) *For $R_0 < |x| < r_0$, the solution u_ε of (1.1) satisfies*

$$\lim_{\varepsilon\to 0} u_\varepsilon(x) = u_0(|x|),$$

where u_0 satisfies $\bar{L}u_0 = 0$ in $R_0 < r < r_0$ with the boundary conditions $u_0(R_0) = \int f(R_0,\theta)\nu_{R_0}(\theta)\,d\theta$ and $u_0(r_0) = \rho f(r_0,\theta_1) + (1-\rho)f(r_0,\theta_2)$. Here $\bar{L}$ is the averaged operator appearing in (1.3).

(ii) *For $x = (r_0,\theta)$, $\lim_{\varepsilon\to 0} u_\varepsilon(x) = h(\theta)f(r_0,\theta_1) + (1-h(\theta))f(r_0,\theta_2)$.*

REMARK 1. The corollary and the proposition allow us to evaluate $u_0(x) \equiv \lim_{\varepsilon\to 0} u_\varepsilon(x)$, for all $x \in \bar{D}$. In particular, note that in general $u_0(r_0,\theta)$,

$$\lim_{\substack{x\to(r_0,\theta)\\ |x|<r_0}} u_0(x)$$

and

$$\lim_{\substack{x\to(r_0,\theta)\\ |x|>r_0\\ x\in D}} u_0(x)$$

give three different values. At (r_0,θ_i) in fact,

$$u_0(r_0,\theta_i), \qquad \lim_{\substack{x\to(r_0,\theta_i)\\ |x|<r_0}} u_0(x), \qquad \lim_{\substack{x\to(r_0,\theta_i)\\ |x|=r_0\\ x\neq(r_0,\theta_i)}} u_0(x) \quad\text{and}\quad \lim_{\substack{x\to(r_0,\theta_i)\\ |x|>r_0\\ x\in D}} u_0(x)$$

give four distinct values in general. [Note that $\lim_{\theta\to\theta_i} h(\theta)$ is decidedly not 0 or 1; there are discontinuities in $h(\theta)$ at θ_1 and θ_2.] Of course $u_0(x)$ is also discontinuous as a function of the domain D. If we slightly perturb one of the two depressions so that (r_0,θ_1) becomes $(r_0+\delta,\theta_1)$, then depending on the sign of δ, ρ will now be either 0 or 1.

REMARK 2. Note that $\hat{r}_\theta(t)$ is quite simple. For each fixed $\hat{\omega}$, it is a deterministic time change of Brownian motion, i.e., it is Gaussian with independent increments.

REMARK 3. $h(\theta)$ (and thus ρ) is considerably simpler in the case that $A \equiv 0$. Then $\hat{\theta}(t)$ is deterministic and thus the hitting times of θ_1 and θ_2 are deterministic and discrete. Furthermore the integrand appearing in the stochastic integral defining $\hat{r}_\theta(t)$ is now deterministic. [This latter point is actually minor since even in the general case the integrand is independent of $\hat{w}(t)$, as noted in Remark 2.] Thus $h(\theta)$ can be represented as an infinite sum of multiple Gaussian integrals. To be more concrete, for each $\theta \in S^1$, there exist $0 \le l_1(\theta) < l_2(\theta) < \cdots$ such that (depending on the position of θ relative to θ_1 and θ_2) $l_1(\theta), l_3(\theta), \dots,$ are the successive hitting times of θ_1 and $l_2(\theta), l_4(\theta), \dots,$ are the successive hitting times of θ_2 or vice versa. For the sake of concreteness, assume that the former holds. Note that $\hat{r}_\theta(t)$ is simply a time changed Brownian motion. Then $h(\theta) = P(J \text{ is odd})$, where $J = \inf\{j > 0\colon \hat{r}_\theta(l_j(\theta)) \ge 0\}$.

REMARK 4. It should be clear from our proof that the same type of theorem may be proved for more general regions D. For example, we could consider n protruding points $\{r_0, \theta_i\}_{i=1}^n$ on ∂D or allow for intervals of the form $\{(r_0, \theta)\colon \theta \in (\alpha, \beta)\}$ to lie on ∂D. We could also similarly perturb the inner boundary; nor need one restrict to dimension $d = 2$.

2. Proof of the theorem. The theorem will follow from (1.4), (1.6), the strong Markov property and the discussion in the paragraph immediately preceeding the statment of the theorem if we show that

$$P_{r_0,\theta}\big((r(\tau_D^\varepsilon), \theta(\tau_D^\varepsilon))\varepsilon \cdot\big) \Rightarrow h(\theta)\delta_{(r_0,\theta_1)} + \big(1 - h(\theta)\big)\delta_{(r_0,\theta_2)} \tag{2.1}$$

as $\varepsilon \to 0$, for all $\theta \in S^1$. We represent the process $(r^\varepsilon(t), \theta^\varepsilon(t)) = (r^\varepsilon(t, r, \theta), \theta^\varepsilon(t, r, \theta))$ generated by L_ε and starting from $(r, \theta) \in R^+ \times S^1$ as follows. Let $(\Omega, \mathscr{F}, P)$ be a probability space on which live three independent Brownian motions $\omega_1(t)$, $\omega_2(t)$ and $\hat{\omega}(t)$. Then

$$\begin{aligned}
r^\varepsilon(t) &= r^\varepsilon(t, r, \theta)\\
&= r + \sqrt{\varepsilon}\int_0^t \sigma_{11}(r^\varepsilon(s), \theta^\varepsilon(s))\, d\omega_1(s)\\
&\quad + \sqrt{\varepsilon}\int_0^t \sigma_{12}(r^\varepsilon(s), \theta^\varepsilon(s))\, d\omega_2(s) + \varepsilon\int_0^t b_1(r^\varepsilon(s), \theta^\varepsilon(s))\, ds,\\
\theta^\varepsilon(t) &= \theta^\varepsilon(t, r, \theta)\\
&= \theta + \sqrt{\varepsilon}\int_0^t \sigma_{21}(r^\varepsilon(s), \theta^\varepsilon(s))\, d\omega_1(s)\\
&\quad + \sqrt{\varepsilon}\int_0^t \sigma_{22}(r^\varepsilon(s), \theta^\varepsilon(s)), d\omega_2(s) + \varepsilon\int_0^t b_2(r^\varepsilon(s), \theta^\varepsilon(s))\, ds\\
&\quad + \int_0^t B(r^\varepsilon(s), \theta^\varepsilon(s))\, ds + \int_0^t \sqrt{A(r^\varepsilon(s), \theta^\varepsilon(s))}\, d\hat{\omega}(s),
\end{aligned} \tag{2.2}$$

where $\sigma = \{\sigma_{ij}\}$ is the unique positive definite square root of a. Of course $\theta^\varepsilon(t)$ is now defined on R instead of on S^1. However, the coefficients are all periodic and thus, with a slight abuse of notation, we will identify $\theta^\varepsilon(t)$ with $\theta^\varepsilon(t) \bmod 2\pi$. We will also identify $\hat{\theta}(t,\theta)$, which appears in the statement of the theorem, with $\hat{\theta}(t,\theta) \bmod 2\pi$.

In general, we will denote probabilities corresponding to $r^\varepsilon(t,r,\theta)$ and $\theta^\varepsilon(t,r,\theta)$ by P since they are in fact functions of ω_1, ω_2 and $\hat{\omega}$. However, for cases in which an equation holds for a specific value of (r,θ), and this would not otherwise be clear, we will write $P_{r,\theta}$.

By a standard martingale inequality, it follows easily that

$$E \sup_{\theta\in S^1} \sup_{0\le s\le t} |r^\varepsilon(s,r,\theta) - r|^2 \le c(\varepsilon t + \varepsilon^2 t^2), \tag{2.3}$$

for some constant c. Thus

$$\lim_{\varepsilon\to 0} \sup_{\theta\in S^1} \sup_{0\le s\le t} |r^\varepsilon(s,r,\theta) - r| = 0, \quad \text{in probability.} \tag{2.4}$$

Thus, by the Borel–Cantelli lemma,

$$\lim_{n\to\infty} \sup_{\theta\in S^1} \sup_{0\le s\le t} |r^{\varepsilon_n}(s,r,\theta) - r| = 0, \quad \text{a.s.}, \tag{2.5}$$

on all sequences $\{\varepsilon_n\}_{n=1}^\infty$ which decrease to 0 sufficiently rapidly.

Using the same martingale inequality again and the Lipschitz continuity of A and B, we have that

$$\begin{aligned} E \sup_{\theta\in S^1} \sup_{0\le s\le t} &|\theta^\varepsilon(s,r_0,\theta) - \hat{\theta}(s,\theta)|^2 \\ &\le c(\varepsilon t + \varepsilon^2 t^2) + (ct + ct^2) E \sup_{\theta\in S^1} \sup_{0\le s\le t} |r^\varepsilon(s,r_0,\theta) - r_0|^2 \\ &\quad + (c + ct)\int_0^t E|\theta^\varepsilon(s,r_0,\theta) - \hat{\theta}(s,\theta)|^2\, ds, \end{aligned}$$

where $\hat{\theta}(s,\theta)$ is as in the statement of the theorem.

Thus, by Gronwall's inequality and (2.3),

$$E \sup_{\theta\in S^1} \sup_{0\le s\le t} |\theta^\varepsilon(s,r_0,\theta) - \hat{\theta}(s,\theta)|^2 \le c_t \varepsilon,$$

for some constant c_t depending on t. Consequently,

$$\lim_{\varepsilon\to 0} \sup_{\theta\in S^1} \sup_{0\le s\le t} |\theta^\varepsilon(s,r_0,\theta) - \hat{\theta}(s,\theta)|^2 = 0, \quad \text{in probability,} \tag{2.6}$$

and, by the Borel–Cantelli lemma,

$$\lim_{n\to\infty} \sup_{\theta\in S^1} \sup_{0\le s\le t} |\theta^{\varepsilon_n}(s,r_0,\theta) - \hat{\theta}(s,\theta)| = 0, \quad \text{a.s.}, \tag{2.7}$$

on all sequences $\{\varepsilon_n\}_{n=1}^\infty$ which decrease to 0 sufficiently rapidly.

We now want to show that the distribution of $r^\varepsilon(\tau_D^\varepsilon, r_0, \theta)$ converges weakly to δ_{r_0} as $\varepsilon \to 0$, for all $\theta \in S^1$, that is, that

$$\lim_{\varepsilon\to 0} P\big(|r^\varepsilon(\tau_D^\varepsilon, r_0,\theta) - r_0| > \delta\big) = 0, \quad \text{for all } \delta > 0. \tag{2.8}$$

To show (2.8), it is enough to show that

$$\lim_{\varepsilon \to 0} P\big(|r^\varepsilon(\tau_D^\varepsilon, r_0, \theta)| = R_0\big) = 0 \tag{2.9}$$

and that

$$\lim_{\varepsilon \to 0} P\big(|r^\varepsilon(\tau_D^\varepsilon, r_0, \theta)| > r_0 + \delta\big) = 0, \quad \text{for all } \delta > 0. \tag{2.10}$$

Let $A_\delta^\varepsilon = \{r^\varepsilon(\cdot, r_0, \theta) \text{ hits } r_0 + \delta \text{ before hitting } R_0\}$. Then $\lim_{n \to \infty} P(A_{1/n}^\varepsilon) = 1$ and in fact

$$\lim_{n \to \infty} \inf_{\varepsilon > 0} P\big(A_{1/n}^\varepsilon\big) = 1. \tag{2.11}$$

(2.11) follows from the fact that one can pick an appropriate Liapunov function independent of ε which in turn follows from the form of the generator L_ε—namely that all the terms involving r differentiation are multiplied by ε which can thus be factored out. By (2.11) and the strong Markov property, to prove (2.9) it is enough to show that

$$\lim_{\varepsilon \to 0} P\big(|r^\varepsilon(\tau_D^\varepsilon, r_0 + \delta, \theta)| = R_0\big) = 0, \tag{2.12}$$

for all sufficiently small $\delta > 0$ and all θ.

Using the notation $\tau_r^\varepsilon = \inf\{t \geq 0 \colon r^\varepsilon(t) = r\}$, we have by the strong Markov property that

$$\begin{aligned} P\big(r^\varepsilon(\tau_D^\varepsilon, r_0, \theta) > r_0 + \delta\big) &= P\big(\tau_{r_0+\delta/2}^\varepsilon < \tau_D^\varepsilon, r^\varepsilon(\tau_D^\varepsilon, r_0, \theta) > r_0 + \delta\big) \\ &\leq \sup_{\theta \in S^1} P\big(r^\varepsilon(\tau_D^\varepsilon, r_0 + \delta/2, \theta) > r_0 + \delta\big). \end{aligned} \tag{2.13}$$

Now (2.9) and (2.10) will follow from (2.12) and (2.13) if we show that

$$\lim_{\varepsilon \to 0} \sup_{\theta \in S^1} P\big(r_0 \leq r^\varepsilon(\tau_D^\varepsilon, r_0 + \delta/2, \theta) \leq r_0 + \delta\big) = 1, \tag{2.14}$$

for all sufficiently small $\delta > 0$.

Fix $\delta > 0$ sufficiently small that $(r_0 + \eta, \theta_i) \in D^c$ for all $0 \leq \eta \leq \delta$ and $i = 1, 2$. This is possible by the smoothness of D. Now, from (2.4) and (2.6) and the positive recurrence of $\hat{\theta}(s, \theta)$, it follows that for any $\gamma > 0$, there exists a T_0 and an $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$,

$$\sup_{\theta \in S^1} P\big(\inf\{t \geq 0 \colon \theta^\varepsilon(t, r_0, \theta) = \theta_1 \text{ or } \theta_2\} \leq T_0\big) > 1 - \frac{\gamma}{2} \tag{2.15}$$

and

$$\sup_{\theta \in S^1} P\left(\sup_{0 \leq s \leq T_0} \left| r^\varepsilon\left(s, r_0 + \frac{\delta}{2}, \theta\right) \right| - \left(r_0 + \frac{\delta}{2}\right) \right| < \frac{\delta}{2}\right) > 1 - \frac{\gamma}{2}. \tag{2.16}$$

(2.14) now follows from (2.15), (2.16) and the constraint placed on δ. This concludes the proof of (2.8).

From (2.8) it follows that the support of the exit distribution for the process starting from (r_0, θ) converges weakly to the limiting support consisting of the two points (r_0, θ_1) and (r_0, θ_2). To complete the proof, we must show that the limiting distribution on these two points is as in (2.1).

In the sequel, we will use the notation

$$\tau_A^\varepsilon = \inf\{t \ge 0: (r^\varepsilon(t,r,\theta), \theta^\varepsilon(t,r,\theta)) \notin A\}, \quad \text{for } A \subset R^2.$$

Also recall that by (2.2) and the fact that the coefficients are defined on all of R^2, it follows that the process $(r^\varepsilon(t,r,\theta), \theta^\varepsilon(t,r,\theta))$ is defined for all $t \ge 0$, that is, it may enter D^c.

We note that by (2.8), we immediately obtain

LEMMA. *$\lim_{\varepsilon\to 0} \sup P_{r_0,\theta}(\tau^\varepsilon_{(D^c\cap U_1)^c} < \tau^\varepsilon_{(D^c\cap U_2)^c})$ is independent of $U_i \subset R^d$ as long as $(D^c \cap U_1) \cap (D^c \cap U_2) = \varnothing$ and $D^c \cap U_i$ includes a relatively open neighborhood of (r_0, θ_i) in D^c, $i = 1,2$.*

In fact, it is then clear from (2.8) that (2.1) will be proved once we show that

$$\lim_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{(D^c\cap U_1)^c} < \tau^\varepsilon_{(D^c\cap U_2)^c}\right) = h(\theta), \tag{2.17}$$

for some pair U_1, U_2 as above. Define $U_{i,\delta} = \{(r,\theta): r \ge r_0 \text{ and } \theta \in [\theta_i - \delta, \theta_i + \delta]\}$, for $\delta \ge 0$. Then for $0 < \delta < \frac{1}{2}|\theta_2 - \theta_1|$, the $U_{i,\delta}$ may serve as the U_i, $i = 1,2$, described above.

For $0 < \delta < \frac{1}{2}|\theta_2 - \theta_1|$, we have the inequality

$$P_{r_0}\left(\tau^\varepsilon_{(D^c\cap U_{1,\delta})^c} < \tau^\varepsilon_{(D^c\cap U_{2,\delta})^c}\right) \le P_{r_0,\theta}\left(\tau^\varepsilon_{U^c_{1,\delta}} < \tau^\varepsilon_{(D^c\cap U_{2,\delta})^c}\right).$$

Although it is not necessarily true that $U_{2,0} \subset D^c$ (refer to Assumption 1, not to Figure 1), from (2.8) we can conclude that similar to Lemma 1,

$$\limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{U^c_{1,\delta}} < \tau^\varepsilon_{(D^c\cap U_{2,\delta})^c}\right) \le \limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{U^c_{1,\delta}} < \tau^\varepsilon_{U^c_{2,0}}\right),$$

since, by the smoothness of D, $[r_0, r_0 + \nu] \times \{\theta_2\} \subset D^c$ for sufficiently small $\nu > 0$. The above inequalities give

$$\begin{aligned}&\limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{(D^c\cap U_{1,\delta})^c} < \tau^\varepsilon_{(D^c\cap U_{2,\delta})^c}\right)\\ &\qquad \le \limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{U^c_{1,\delta}} < \tau^\varepsilon_{U^c_{2,0}}\right).\end{aligned} \tag{2.18}$$

We will show that

$$\lim_{\delta\to 0}\lim_{n\to\infty} P_{r_0,\theta}\left(\tau^{\varepsilon_n}_{U^c_{1,\delta}} < \tau^{\varepsilon_n}_{U^c_{2,0}}\right) = h(\theta) \tag{2.19}$$

on all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly. By switching the roles of θ_1 and θ_2, we obtain analogous to (2.18),

$$\begin{aligned}&\limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{(D^c\cap U_{2,\delta})^c} < \tau^\varepsilon_{(D^c\cap U_{1,\delta})^c}\right)\\ &\qquad \le \limsup_{\varepsilon\to 0} P_{r_0,\theta}\left(\tau^\varepsilon_{U^c_{2,\delta}} < \tau^\varepsilon_{U^c_{1,0}}\right).\end{aligned} \tag{2.20}$$

By interchanging θ_1 and θ_2 in the proof of (2.19), we obtain analogous to (2.19),

$$\lim_{\delta\to 0}\lim_{n\to\infty} P_{r_0,\theta}\left(\tau^{\varepsilon_n}_{U^c_{2,\delta}} < \tau^{\varepsilon_n}_{U^c_{1,0}}\right) = 1 - h(\theta), \tag{2.21}$$

on all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly.

Now we can complete the proof as follows. From the lemma, it follows that the left-hand sides of (2.18) and (2.20) do not depend on δ as long as $0 < \delta < \frac{1}{2}|\theta_2 - \theta_1|$. Thus, from (2.18)–(2.21), it follows that

$$\limsup_{n\to\infty} P_{r_0,\theta}\big(\tau^{\varepsilon_n}_{(D^c\cap U_{1,\delta})^c} < \tau^{\varepsilon_n}_{(D^c\cap U_{2,\delta})^c}\big) \le h(\theta) \tag{2.22}$$

and

$$\limsup_{n\to\infty} P_{r_0,\theta}\big(\tau^{\varepsilon_n}_{(D^c\cap U_{2,\delta})^c} < \tau^{\varepsilon_n}_{(D^c\cap U_{1,\delta})^c}\big) \le 1 - h(\theta), \tag{2.23}$$

for all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly.

Since the sum of the left-hand sides of (2.22) and (2.23) is greater than or equal to unity, it follows that in fact (2.22) and (2.23) are equalities. It is also clear that if

$$\liminf_{n\to\infty} P_{r_0,\theta}\big(\tau^{\varepsilon_n}_{(D^c\cap U_{1,\delta})^c} < \tau^{\varepsilon_n}_{(D^c\cap U_{2,\delta})^c}\big)$$

were strictly smaller than $h(\theta)$, then

$$\limsup_{n\to\infty} P_{r_0,\theta}\big(\tau^{\varepsilon}_{(D^c\cap U_{2,\delta})^c} < \tau^{\varepsilon}_{(D^c\cap U_{2,\delta})^c}\big)$$

would be strictly greater than $1 - h(\theta)$, contradicting (2.23). Thus, in fact,

$$\lim_{n\to\infty} P_{r_0,\theta}\big(\tau^{\varepsilon_n}_{(D^c\cap U_{1,\delta})^c} < \tau^{\varepsilon_n}_{(D^c\cap U_{2,\delta})^c}\big) = h(\theta). \tag{2.24}$$

Since (2.24) holds for all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly, it follows that

$$\lim_{\varepsilon\to 0} P_{r_0,\theta}\big(\tau^{\varepsilon}_{(D^c\cap U_{1,\delta})^c} < \tau^{\varepsilon}_{(D^c\cap U_{2,\delta})^c}\big) = h(\theta). \tag{2.25}$$

This gives (2.17) and proves the theorem. Thus, it remains to show (2.19).

Write the condition $r^\varepsilon(t) \ge r_0$ as $(r^\varepsilon(t) - r_0)/\sqrt{\varepsilon} \ge 0$ and note that for $0 < \delta < \frac{1}{2}|\theta_2 - \theta_1|$ and $\varepsilon > 0$,

$$\tau^{\varepsilon}_{U^c_{1,\delta}} \wedge \tau^{\varepsilon}_{U^c_{2,0}} = \inf\left\{t \ge 0: \theta^\varepsilon(t) \in [\theta_1 - \delta, \theta_1 + \delta] \cup \{\theta_2\} \text{ and } \frac{r^\varepsilon(t) - r_0}{\sqrt{\varepsilon}} \ge 0\right\}.$$

From (2.2), we have

$$\frac{r^\varepsilon(t) - r_0}{\sqrt{\varepsilon}} = \int_0^t \sigma_{11}(r^\varepsilon(s), \theta^\varepsilon(s))\, d\omega_1(s) + \int_0^t \sigma_{12}(r^\varepsilon(s), \theta^\varepsilon(s))\, d\omega_2(s) + \sqrt{\varepsilon}\int_0^t b_1(r^\varepsilon(s), \theta^\varepsilon(s))\, ds.$$

Define $\tilde{r}(t) = \tilde{r}(t, \theta)$ by

$$\tilde{r}(t) = \int_0^t \sigma_{11}\big(r_0, \hat{\theta}(s)\big)\, d\omega_1(s) + \int_0^t \sigma_{12}\big(r_0, \hat{\theta}(s)\big)\, d\omega_2(s). \tag{2.26}$$

[Note that $\tilde{r}(t)$ depends on θ through $\hat{\theta}(s) = \hat{\theta}(s, \theta)$.] It follows readily from the assumed Lipschitz continuity that

$$E \sup_{\theta\in S^1} \sup_{0\le s\le t} \left|\tilde{r}(s) - \frac{r^\varepsilon(s) - r_0}{\sqrt{\varepsilon}}\right| \le ctE \sup_{\theta\in S^1} \sup_{0\le s\le t} |r^\varepsilon(s) - r_0|^2 + c\varepsilon t^2,$$

for some $c > 0$. By (2.3), we obtain

$$\lim_{\varepsilon \to 0} \sup_{\theta \in S^1} \sup_{0 \le s \le t} \left| \tilde{r}(s) - \frac{r^{\varepsilon}(s) - r_0}{\sqrt{\varepsilon}} \right| = 0, \quad \text{in probability,}$$

and, by the Borel–Cantelli lemma,

$$\lim_{n \to \infty} \sup_{\theta \in S^1} \sup_{0 \le s \le t} \left| \tilde{r}(s) - \frac{r^{\varepsilon_n}(s) - r_0}{\sqrt{\varepsilon_n}} \right| = 0, \quad \text{a.s.,} \tag{2.27}$$

for all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly.

Now define, analogous to $\tau^{\varepsilon}_{U^c_{1,\delta}}$ and $\tau^{\varepsilon}_{U^c_{2,0}}$,

$$\tau^0_{1,\delta} = \inf\{t \ge 0 : \hat{\theta}(t) \in [\theta_1 - \delta, \theta_1 + \delta] \text{ and } \tilde{r}(t) \ge 0\}, \quad \text{for } \delta \ge 0,$$

and

$$\tau^0_{2,0} = \inf\{t \ge 0 : \hat{\theta}(t) = \theta_2 \text{ and } \tilde{r}(t) \ge 0\}.$$

For $\varepsilon > 0$ and $\delta > 0$, define $A^{\varepsilon,\delta} = \{\tau^{\varepsilon}_{U^c_{1,\delta}} < \tau^{\varepsilon}_{U^c_{2,0}}\}$ and, for $\delta \ge 0$, define $A^{0,\delta} = \{\tau^0_{1,\delta} < \tau^0_{2,0}\}$. We claim that

$$\lim_{n \to \infty} P_{r_0,\theta}(A^{\varepsilon_n,\delta}) = P_{r_0,\theta}(A^{0,\delta}), \quad \text{for } \delta > 0,$$

and all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly. In the interest of brevity, we will only prove that

$$\limsup_{n \to \infty} P_{r_0,\theta}(A^{\varepsilon_n,\delta}) \le P_{r_0,\theta}(A^{0,\delta}), \quad \text{for } \delta > 0, \tag{2.28}$$

and all sequences $\{\varepsilon_n\}$ decreasing to 0 sufficiently rapidly, since this is in fact enough to prove the theorem. Using (2.28), we will be able to show instead of (2.19),

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{r_0,\theta}\left(\tau^{\varepsilon_n}_{U^c_{1,\delta}} < \tau^{\varepsilon_n}_{U^c_{2,0}}\right) \le h(\theta). \tag{2.19$'$}$$

As before, by interchanging the roles of θ_1 and θ_2 we obtain analogous to (2.19′),

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{r_0,\theta}\left(\tau^{\varepsilon_n}_{U^c_{2,\delta}} < \tau^{\varepsilon_n}_{U^c_{1,0}}\right) \le 1 - h(\theta). \tag{2.21$'$}$$

However (2.22) and (2.23) follow just as well from (2.18), (2.19′), (2.20) and (2.21′).

Denote the generic point in Ω by ω. To show (2.28), we will show that for arbitrary $T > 0$, if $\omega \in (A^{0,\delta})^c$ and $\tau^0_{2,0} \le T$, then $\omega \in (A^{\varepsilon_n,\delta})^c$ almost surely for sufficiently large n. Thus assume $\omega \in (A^{0,\delta})^c$ and $\tau^0_{2,0} \le T$. Define $\gamma = \inf\{s > 0 : (\tilde{r}(\tau^0_{2,0} + s), \hat{\theta}(\tau^0_{2,0} + s)) \in [0, \infty) \times [\theta_1 - \delta, \theta_1 + \delta]\}$. Since $(\tilde{r}(t), \hat{\theta}(t))$ almost surely never hits $(0, \theta_2)$, we have $\tilde{r}(\tau^0_{2,0}) > 0$ almost surely, and thus one can pick an $\eta = \eta(\omega) > 0$ satisfying $\tau^0_{2,0} + \eta < \min(\tau^0_{2,0} + \gamma/2, T + 1)$ and such that

$$\tilde{r}(t) > 0, \quad \text{for all } t \in \left(\tau^0_{2,0} - \eta, \tau^0_{2,0} + \eta\right), \quad \text{a.s.} \tag{2.29}$$

Furthermore, there exists a $\nu = \nu(\omega) > 0$ such that

$$\inf_{t \in (\tau^0_{2,0} - \eta, \tau^0_{2,0} + \eta)} \hat{\theta}(t) \le \theta_2 - \nu, \quad \text{a.s.} \tag{2.30}$$

and

$$\sup_{t\in(\tau^0_{2,0}-\eta,\,\tau^0_{2,0}+\eta)} \hat{\theta}(t) \geq \theta_2 + \nu, \quad \text{a.s.} \tag{2.31}$$

In the case $A(r_0, \theta_2) > 0$, (2.30) and (2.31) follow from the law of the iterated logarithm. In the case $A(r_0, \theta_2) = 0$, by assumption $B(r_0, \theta_2) > 0$. Using this, one can show that $\hat{\theta}(t) > \theta_2$ almost surely for $t > \tau^0_{2,0}$ and $\hat{\theta}(t) < \theta_2$ almost surely for $t < \tau^0_{2,0}$. [A comment is in order concerning the above. Recall that $\hat{\theta}(t)$ is actually defined on R, but we have agreed to identify $\hat{\theta}(t)$ with $\hat{\theta}(t) \bmod 2\pi$. Up until (2.30), we have been considering $\hat{\theta}(t)$ on the circle. For (2.30) and (2.31) and the statement following them to be true, one should consider $\hat{\theta}(t)$ on R.]

Now it follows from (2.7), (2.27), (2.29)–(2.31) and the fact that $\tau^0_{2,0} + \eta \leq \min(\tau^0_{2,0} + \gamma/2, T + 1)$ that

$$\tau^{\varepsilon_n}_{U^c_{2,0}} < \min\left(\tau^0_{2,0} + \gamma/2, T + 1\right), \quad \text{a.s.,} \tag{2.32}$$

for sufficiently large n.

To complete the proof that $\omega \in (A^{\varepsilon_n,\delta})^c$ for sufficiently large n, suppose to the contrary that $\omega \in A^{\varepsilon_n,\delta}$ for infinitely many n. Without loss of generality, we may suppose that $\omega \in A^{\varepsilon_n,\delta}$ for all n. By (2.32), $\tau^{\varepsilon_n}_{U^c_{1,\delta}} < \min(\tau^0_{2,0} + \gamma/2, T + 1)$, for sufficiently large n. Let $\{\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}\}_{k=1}^{\infty}$ denote a convergent subsequence with limit $t_0 \leq \min(\tau^0_{2,0} + \gamma/2, T + 1)$. Then by (2.7) and (2.27) and the fact that $t_0 \leq T + 1$, one has

$$\lim_{k\to\infty} \left(\frac{r^{\varepsilon_{n_k}}\left(\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}\right) - r_0}{\sqrt{\varepsilon_{n_k}}}, \theta^{\varepsilon_{n_k}}\left(\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}\right) \right) = \left(\tilde{r}(t_0), \tilde{\theta}(t_0)\right).$$

Now,

$$\left(\frac{r^{\varepsilon_{n_k}}\left(\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}\right) - r_0}{\sqrt{\varepsilon_{n_k}}}, \theta^{\varepsilon_{n_k}}\left(\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}\right) \right) \in [0, \infty) \times [\theta_1 - \delta, \theta_0 + \delta]$$

by the definition of $\tau^{\varepsilon_{n_k}}_{U^c_{1,\delta}}$. Thus $(\tilde{r}(t_0), \hat{\theta}(t_0)) \in [0, \infty) \times [\theta_1 - \delta, \theta_1 + \delta]$. This means that $\tau^0_{1,\delta} \leq t_0 \leq \tau^0_{2,0} + \gamma/2$. But by the definition of γ, this implies that in fact $\tau^0_{1,\delta} < \tau^0_{2,0}$ and consequently that $\omega \in A^{0,\delta}$, contradicting the assumption that $\omega \in (A^{0,\delta})^c$. Thus in fact $\omega \in (A^{\varepsilon_n,\delta})^c$ for all sufficiently large n. This completes the proof of (2.28).

We will now show that

$$\lim_{\delta\to 0} P_{r_0,\theta}(A^{0,\delta}) = h(\theta). \tag{2.33}$$

Then (2.33) and (2.28) give (2.19′) to complete the proof of the theorem. Since $(\tilde{r}(\cdot), \hat{\theta}(\cdot), \tau^0_{2,0})$ [or more explicitly $(\tilde{r}(\cdot, \theta), \hat{\theta}(\cdot, \theta), \tau^0_{2,0})$] and $(\hat{r}_\theta(\cdot), \hat{\theta}(\cdot, \theta), \tau)$ are distributed identically [see the statement of the theorem for the definitions of $\hat{r}_\theta(\cdot)$ and τ], (2.33) will be proved once we show that

$$\lim_{\delta\to 0} P_{r_0,\theta}(A^{0,\delta}) = P_{r_0,\theta}(A^{0,0}). \tag{2.34}$$

It is clear from the definitions of $\tau^0_{1,\delta}$ and $\tau^0_{1,0}$ that $A^{0,0} \subset A^{0,\delta}$ for $\delta > 0$. Thus to prove (2.34), it suffices to show that if $\omega \in (A_{0,0})^c$ and $\delta_n \downarrow 0$ as $n \to \infty$, then $\omega \in (A^{0,\delta_n})^c$ almost surely for sufficiently large n. Assume $\omega \in (A^{0,0})^c$, that is, $\tau_{2,0}^0 < \tau_{1,0}^0$. Then there exist a $\nu = \nu(\omega) > 0$ and a $\eta = \eta(\omega) > 0$ such that

$$\sup_{\substack{t:\, \hat{\theta}(t) \in [\theta_1 - \nu, \theta_1 + \nu] \\ t < \tau_{2,0}^0}} \tilde{r}(t) < -\eta, \quad \text{a.s.} \tag{2.35}$$

To prove (2.35), assume to the contrary that there exists a sequence $\{t_n\}_{n=1}^{\infty}$ with $t_n < \tau_{2,0}^0$, $\tilde{r}(t_n) > -1/n$ and $\lim_{n\to\infty}\hat{\theta}(t_n) = \theta_1$. Then there exists a subsequence t_{n_k} and a t_0 such that $\lim_{k\to\infty} t_{n_k} = t_0$ and thus $\hat{\theta}(t_0) = \theta_1$, $\tilde{r}(t_0) \geq 0$ and $t_0 < \tau_{2,0}^0$. This contradicts the assumption $\omega \in (A_{0,0})^c$. From (2.35), it follows that $\omega \in (A^{0,\delta_n})^c$ almost surely for sufficiently large n. This proves (2.34) and completes the proof of the theorem.

REFERENCES

[1] Freidlin, M. I. (1985). *Functional Integration and Partial Differential Equations*. Princeton Univ. Press, Princeton, N.J.

[2] Khasminskii, R. Z. (1963). On diffusion processes with small parameter. *Izv. Akad. Nauk SSSR Ser. Mat.* **27** 1281–1300. (In Russian.)

[3] Khasminskii, R. Z. (1968). On the averaging principle for stochastic differential equations. *Kybernetika (Prague)* 4 260–279. (In Russian.)

Department of Mathematics
Technion—Israel Institute of Technology
Haifa 32000
Israel