The Annals of Probability
1999, Vol. 27, No. 1, 284–296

ON LARGE DEVIATIONS IN THE AVERAGING PRINCIPLE FOR SDEs WITH A "FULL DEPENDENCE"[1]

BY A. YU. VERETENNIKOV

Institute for Information Transmission Problems

We establish the large deviation principle for stochastic differential equations with averaging in the case when all coefficients of the fast component depend on the slow one, including diffusion.

1. Introduction. We consider the SDE system

$$
\begin{aligned}
&dX_t = f(X_t, Y_t)\,dt, \qquad X_0 = x_0,\\
&dY_t = \varepsilon^{-2}B(X_t, Y_t)\,dt + \varepsilon^{-1}C(X_t, Y_t)\,dW_t, \qquad Y_0 = y_0.
\end{aligned}
\tag{1}
$$

Here $X_t \in E^d$, $Y_t \in M$, M is a compact manifold of dimension l (e.g., torus T^l), f is a function with values in d-dimensional Euclidean space E^d, B is a function with values in TM, C is a function with values in $(TM)^l$ (i.e., in local coordinates an $l \times l$ matrix), W_t is an l-dimensional Wiener process on some probability space (Ω, F, P), $\varepsilon > 0$ is a small parameter. Concerning SDE's on manifolds we refer to Watanabe and Ikeda (1989).

The large deviation principle (LDP) for such systems with a "full dependence", that is, $C(X_t, Y_t)$, was not treated earlier. Only the case $C(Y_t)$ was considered in papers by Freidlin (1976), Freidlin (1978), Freidlin and Wentzell (1984) for a compact state space and by Veretennikov (1994) for a noncompact one. There are, as well, recent papers on more general systems with small additive diffusions by Liptser and by the author which also only concern the case $C(Y_t)$.

The LDP for systems like (1) is important in averaging and homogenization, in the KPP equation theory, for stochastic approximation algorithms with averaging and so forth. The problem of an LDP for the case $C(X_t, Y_t)$ has arisen since Freidlin (1976) and Freidlin (1978). Intuitively, the scheme used for $C(Y_t)$ should work; at least, almost all main steps go well. Indeed, there was only one lacuna; the use of Girsanov's transformation did not allow freezing of X_t if C depended on the slow motion while it worked well and very naturally for the drift $B(X_t, Y_t)$. Yet the problem remained unresolved for years and the answer was not clear at all.

Notice that this difficulty does not appear in analogous discrete-time systems [see Gulinsky and Veretennikov (1993), Chapter 11].

Received June 1998.
[1]Supported by grants RFBR 98-01-00062, INTAS 93-0378, INTAS 93-1585-ext.
AMS 1991 *subject classifications*. 60F10, 60J60.
Key words and phrases. Large deviations, averaging, stochastic differential equation.

It turned out that the use of Girsanov's transformation in some sense prevented resolving the problem. Our approach in this paper is based on a new technical lemma, Lemma 5 below. The main idea is to use two different scales of partitions of the interval $[0, T]$, a "first-order partition" by points Δ, $2\Delta, \dots,$ which do not depend on the small parameter ε and "second-order partitions" which depend on ε in a special way, by points $\varepsilon^2 t(\varepsilon), 2\varepsilon^2 t(\varepsilon), \dots$. Then the exponential estimates needed for the proof of the result can be established by two steps. First, the estimates for a "small" partition interval are derived using the uniform bound of Lemma 3 (see below) and the estimates for stochastic integrals. It is important that, in the "second" scale, the fast motion is still close enough to its frozen version [the bound (11) below]. Second, the bounds for "small" partitions and induction give one the estimate for a "large" partition interval.

The main result is stated in Section 2. In Section 3 we expose auxiliary lemmas, among them the main technical Lemma 5 with proof and a version of an important lemma from Freidlin and Wentzell (1984) (see Lemma 6), which requires certain comments. Those comments are given in the Appendix. The proof of the main theorem is presented in Section 4.

2. Main result.

We make the following assumptions.

(A_f) The function f is bounded and satisfies the Lipschitz condition.
(A_C) The function CC^* is bounded, uniformly nondegenerate, C satisfies the Lipschitz condition.
(A_B) The function B is bounded and satisfies the Lipschitz condition.

Some conditions may be relaxed; for example, B can be locally bounded, C locally (w.r.t. x) nondegenerate and so on.

The family of processes X^ε satisfies a large deviation principle in the space $C([0, T]; R^d)$ with a normalizing coefficient ε^{-2} and a rate function $S(\varphi)$ if three conditions are satisfied:

$$\limsup_{\delta \to 0} \limsup_{\varepsilon \to 0} \varepsilon^2 \log P_x(X^\varepsilon \in F) \le -\inf_F S(\varphi) \quad \forall\, F \text{ closed}, \tag{2}$$

$$\liminf_{\delta \to 0} \liminf_{\varepsilon \to 0} \varepsilon^2 \log P_x(X^\varepsilon \in G) \ge -\inf_G S(\varphi), \quad \forall\, G \text{ open} \tag{3}$$

and S is a "good" rate function; that is, for any $s \ge 0$, the set

$$\Phi(s) := \big(\varphi \in C([0, T]; R^d) \colon S(\varphi) \le s,\ \varphi(0) = x\big)$$

is compact in $C([0, T]; R^d)$.

Let $\tilde{W}_t = \varepsilon^{-1} W_{t\varepsilon^2}$, $y_t = Y_{t\varepsilon^2}$, $x_t = X_{t\varepsilon^2}$ and let y_t^x denote a solution of SDE,

$$dy_t^x = B(x, y_t^x)\, dt + C(x, y_t^x)\, d\tilde{W}_t, \qquad y_0^x = y_0. \tag{4}$$

THEOREM 1. *Let* (A_f), (A_B), (A_C) *be satisfied. Then the family* ($X_t^\varepsilon = X_t$, $0 \le t \le T$) *satisfies the LDP as* $\varepsilon \to 0$ *in the space* $C([0, T]; R^d)$ *with a rate function*

$$S(\varphi) = \int_0^T L(\varphi_t, \dot{\varphi}_t)\, dt,$$

where

$$L(x, \alpha) = \sup_{\beta}(\alpha\beta - H(x, \beta)),$$

$$H(x, \beta) = \lim_{t \to \infty} t^{-1} \log E \exp\left(\int_0^t f(x, y_s^x)\, ds \right).$$

The limit H exists and is finite for any β, *the functions H and L are convex in their last arguments* β *and* α *correspondingly*, $L \geq 0$ *and H is continuously differentiable in* β.

The differentiability of H at any β is provided by the compactness of the state space of the fast component.

3. Auxiliary lemmas. Let us consider the semigroup of operators T_t^β, $t \geq 0$ on $C(M)$ defined by the formula

$$T_t^{x', x, \beta} g(y) = T_t^\beta g(y) = E_y g(y_t^x) \exp\left(\int_0^t \beta f(x', y_s^x)\, ds \right),$$

where $\beta \in E^d$, βf is a scalar product.

LEMMA 1. *Let assumptions* (A_f), (A_B), (A_C) *be satisfied. Then for any* β, *the operator* T_1^β *is compact in the space* $C(M)$.

LEMMA 2. *Let assumptions* (A_f), (A_B), (A_C) *be satisfied. Then the spectral radius* $r(T_1^\beta)$ *is a simple eigenvalue of* T_1^β *separated from the rest of the spectrum and its eigenfunction* e_β *belongs to the cone* $C^+(M)$. *Moreover, function* $r(T_1^\beta)$ *is smooth (of* C^∞*) in* β *and the function* e_β *is bounded and separated away from zero uniformly in* $|\beta| < b$ *and any* x', x.

LEMMA 3. *Let* $\beta \in E^d$ *and let assumptions* (A_f), (A_B), (A_C) *be satisfied. Then there exists a limit*

$$H(x', x, \beta) = \lim_{t \to 0} t^{-1} \log E_y \exp\left(\beta \int_0^t f(x', y_s^x)\, ds \right);$$

moreover, $H(x', x, \beta) = \log r(T_1^{x', x, \beta})$. *The function* $H(x', x, \beta)$ *is of* C^∞ *in* β *and convex in* β. *For any* $b > 0$ *there exists* $C(b)$ *such that, for any* y, $|\beta| < b$,

$$\left| t^{-1} \log E_y \exp\left(\beta \int_0^t f(x', y_s^x)\, ds \right) - H(x', x, \beta) \right| \leq C(b) t^{-1}. \tag{5}$$

Notice that $|H(x', x, \beta)| \leq \|f\|_C |\beta|$.

LEMMA 4. *Let assumptions* (A_f), (A_B), (A_C) *be satisfied. Then for any* $b > 0$ *the functions H and* $\nabla_\beta H$ *are uniformly continuous in* (x', x, β), $|\beta| < b$.

Lemmas 1–4 are standard [cf. Veretennikov (1994) or (1992)]. They are based on Frobenius-type theorems for positive compact operators [see Krasnosel'skii, Lifshitz and Sobolev (1989)] and the theory of perturbations of linear operators [see Kato (1976), Chapter 2]. Denote $\tilde{F}_t = F_{t\varepsilon^2}$.

LEMMA 5. *Let assumptions* (A_f), (A_B), (A_C), $b > 0$, $t(\varepsilon) \to \infty$ *and* $t(\varepsilon) = o(\log \varepsilon^{-1})$ *as* $\varepsilon \to 0$. *Then for any* $\nu > 0$ *there exist* $\delta(\nu) > 0$, $\varepsilon(\nu) > 0$ *such that for* $\varepsilon \le \varepsilon(\nu)$ *uniformly w.r.t.* t_0, x', x, x_0, y_0 *and* $|\beta| \le b$, *the inequality holds on the set* $\{|x_{t_0} - x| < \delta(\nu)\}$,

$$(6)\qquad \left| \log E\left(\exp\left(\beta \int_{t_0}^{t_0 + t(\varepsilon)} f(x', y_s)\, ds \right) \Big| \tilde{F}_{t_0} \right) - t(\varepsilon) H(x', x, \beta) \right| \le \nu t(\varepsilon).$$

Moreover, *if* $\Delta \le \Delta(\nu) = (1 + \|f\|_C)^{-1}\delta(\nu)/2$ *and* ε *is small enough*, *then uniformly w.r.t.* $T_0 \ge 0$, t_0, x', x, x_0, y_0 *and* $|\beta| \le b$,

$$(7)\qquad \begin{aligned} &\exp\left(\varepsilon^{-2}\Delta H(x', x, \beta) - \nu\Delta\varepsilon^{-2}\right) \\ &\quad \le E\left(\exp\left(\beta\varepsilon^{-2} \int_{T_0}^{T_0+\Delta} f(x', Y_s)\, ds \right) \Big| F_{T_0} \right) \\ &\quad \le \exp\left(\varepsilon^{-2}\Delta H(x', x, \beta) + \nu\Delta\varepsilon^{-2}\right). \end{aligned}$$

PROOF. *Step* 1. It is sufficient to prove (6) and (7) for $T_0 = 0$. Moreover, since H is continuous, it suffices to check both inequalities for $x = x_0$. Indeed, the bound

$$\left| \log E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s)\, ds \right) - t(\varepsilon) H(x', x_0, \beta) \right| \le \nu t(\varepsilon)$$

implies

$$\begin{aligned} &\left| \log E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s)\, ds \right) - t(\varepsilon) H(x', x, \beta) \right| \\ &\quad \le t(\varepsilon)\left(\nu + |H(x', x, \beta) - H(x', x_0, \beta)|\right). \end{aligned}$$

The same arguments are applicable to the second inequality of the assertion of the lemma. So, in the sequel we consider the case $x_0 = x$.

Let us show first that

$$(8)\qquad \sup_{x', x_0} \left| t(\varepsilon)^{-1} \log E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s)\, ds \right) - H(x', x, \beta) \right| \le \nu$$

if ε is small enough. Due to Lemma 3, it would be correct if y_s were replaced by y_s^x and $t(\varepsilon) \ge \nu^{-1}C(b)$. We will also use the bounds

$$(9)\qquad \sup_{0 \le s \le t} |x_s - x_0| \le \|f\|_C \varepsilon^2 t \quad \forall\, C, \qquad \exp(Ct(\varepsilon))t(\varepsilon)^2 \varepsilon^2 \to 0, \qquad \varepsilon \to 0.$$

Let $|f(x', y) - f(x', y')| \le L_f |y - y'|$ for all y, y', x', $L_f > 0$, $C_f = \|f\|_C$. We estimate for $t(\varepsilon) > \nu^{-1} C(b)/4$,

$$
\begin{aligned}
&E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s)\, ds \right) \\
&\qquad \times \left\{ I\left(\sup_{0 \le t \le t(\varepsilon)} |y_t - y_t^x| \le \nu/(4L_f b) \right) \right. \\
&\qquad\qquad \left. + I\left(\sup_{0 \le t \le t(\varepsilon)} |y_t - y_t^x| > \nu/(4L_f b) \right) \right\} \\
&\qquad \le E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s^x)\, ds + t(\varepsilon)\nu/4 \right) \\
&\qquad\qquad \times I\left(\sup_{0 \le t \le t(\varepsilon)} |y_t - y_t^x| \le \nu/(4L_f b) \right) \\
&\qquad\qquad + \exp\left(C_f b t(\varepsilon) \nu \right) E I\left(\sup_{0 \le t \le t(\varepsilon)} |y_t - y_t^x| > \nu/(4L_f b) \right) \\
&\qquad \le E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s^x)\, ds \right) \exp(t(\varepsilon)\nu/4) \\
&\qquad\qquad + \exp\left(C_f b t(\varepsilon) \nu \right) \nu^{-2} E \sup_{t \le t(\varepsilon)} |y_t - y_t^x|^2 .
\end{aligned}
\tag{10}
$$

By virtue of Lemma 3 we have

$$
E \exp\left(\beta \int_0^{t(\varepsilon)} f(x', y_s^x)\, ds \right) \le \exp\left(t(\varepsilon)\left(H(x', x, \beta) + \nu/4 \right) \right)
$$

if ε is small enough. A similar lower bound holds true also.

Let us estimate the second term. By virtue of the inequalities for the Itô and Lebesgue integrals, we have

$$
\begin{aligned}
&E \sup_{t' \le t} |y_{t'} - y_{t'}^x|^2 \\
&\qquad \le C E \int_0^t \left| C(x_s, y_s) - C(x, y_s^x) \right|^2 ds \\
&\qquad\qquad + C t E \int_0^t \left| B(x_s, y_s) - B(x, y_s^x) \right|^2 ds \\
&\qquad \le C \int_0^t E |x_s - x|^2\, ds + C \int_0^t E \sup_{u \le s} |y_s - y_s^x|^2\, ds \\
&\qquad \le C t^2 \varepsilon^2 + C \int_0^t E \sup_{u \le s} |y_u - y_u^x|^2\, ds .
\end{aligned}
$$

By virtue of Gronwall's lemma, one gets

$$
E \sup_{t' \le t} |y_{t'} - y_{t'}^x|^2 \le C t^2 \varepsilon^2 \exp(Ct).
$$

In particular,

$$E \sup_{t' \le t(\varepsilon)} |y_{t'} - y^x_{t'}|^2 \le Ct(\varepsilon)^2 \varepsilon^2 \exp(Ct(\varepsilon)). \tag{11}$$

So the second term in (10) does not exceed the value $\exp(C_f bt(\varepsilon)\nu)\nu^{-2} Ct(\varepsilon)^2\varepsilon^2$ which is $o(\exp(Ht(\varepsilon)))$ for any H. Indeed, $\exp(t(\varepsilon)(C_f b\nu - H))\nu^{-2} Ct(\varepsilon)^2\varepsilon^2 \to 0$ due to the assumption $t(\varepsilon) = o(\log \varepsilon^{-1})$, $\varepsilon \to 0$. This proves (8).

Notice that the bound (8) is uniform w.r.t. $|\beta| \le b$ and x', x, y_0. Since the function H is continuous, we get on the set $\{|x_{t_0} - x| < \delta(\nu)\}$,

$$\sup_{x', x, y_0} \sup_{|\beta| \le b} \sup_{t_0} \left| \log E\left(\exp\left(\beta \int_{t_0}^{t_0 + t(\varepsilon)} f(x', y^x_s)\, ds\right) \middle| \tilde F_{t_0}\right) - t(\varepsilon) H(x', x, \beta) \right| \le \nu t(\varepsilon) \tag{12}$$

if $\delta(\nu)$ is small enough.

Step 2. Let $\Delta \le (1 + \|f\|_C)^{-1}\delta(\nu)/2$ and $N = \Delta\varepsilon^{-2}t(\varepsilon)^{-1}$. Then $\sup_{0 \le s \le Nt(\varepsilon)}|x_s - x_0| \le \delta(\nu)/2$. Let $|x - x_0| < \delta(\nu)/2$. So, $\sup_{0 \le s \le Nt(\varepsilon)}|x_s - x| < \delta(\nu)$. In particular, $|x_{kt(\varepsilon)} - x| < \delta(\nu)$ for any $1 \le k \le N$. By induction, we get from (12) for such k,

$$\begin{aligned} &\exp(kt(\varepsilon)H(x', x, \beta) - \nu kt(\varepsilon)) \\ &\quad\le E \exp\left(\beta \int_0^{kt(\varepsilon)} f(x', y_s)\, ds\right) \\ &\quad\le \exp(kt(\varepsilon)H(x', x, \beta) + \nu kt(\varepsilon)), \end{aligned}$$

or, after the time change,

$$\begin{aligned} &\exp(kt(\varepsilon)H(x', x, \beta) - \nu kt(\varepsilon)) \\ &\quad\le E \exp\left(\beta\varepsilon^{-2} \int_0^{kt(\varepsilon)\varepsilon^{-2}} f(x', Y_s)\, ds\right) \\ &\quad\le \exp(kt(\varepsilon)H(x', x, \beta) + \nu kt(\varepsilon)). \end{aligned}$$

Since H is continuous, we obtain for $k = N$,

$$\begin{aligned} &\exp\left(\varepsilon^{-2}\Delta H(x', x, \beta) - \nu\Delta\varepsilon^{-2}\right) \\ &\quad\le E \exp\left(\beta\varepsilon^{-2} \int_0^{\Delta} f(x', Y_s)\, ds\right) \\ &\quad\le \exp\left(\varepsilon^{-2}\Delta H(x', x_0, \beta) + \nu\Delta\varepsilon^{-2}\right). \end{aligned} \tag{13}$$

Lemma 5 is proved. □

In the sequel we denote $\Delta = \Delta(\nu)$. An important point is that $\delta(\nu)/\Delta(\nu) = \text{const} > 0$.

LEMMA 6 [Freidlin (1978), Freidlin and Wentzell (1984)]. *Let $S(\varphi) < \infty$. If ψ^n is a sequence of step functions tending uniformly to φ in $C[0,T]; R^d)$ as*

$n \to \infty$, then there exists a sequence of piecewise linear functions χ^n which also tend uniformly to φ and such that

$$\limsup_{n\to\infty} \int_0^T L(\psi_s^n, \dot{\chi}_s^n)\, ds \le S(\varphi).$$

Moreover, one may assume without loss of generality that for any s there exists a value

$$\beta_s = \operatorname*{argmax}_{\beta} \left(\beta \dot{\chi}_{s+}^n - H(\psi_s^n, \psi_s^n, \beta) \right)$$

and

$$L(\psi_s^n, \alpha) > L(\psi_s^n, \dot{\chi}_{s+}^n) + (\alpha - \dot{\chi}_{s+}^n)\beta_s \quad \forall\, \alpha \ne \dot{\chi}_s^n.$$

If $\hat{\psi}$ is close enough to ψ_s^n then there exists a value

$$\hat{\beta}_s = \operatorname*{argmax}_{\beta} \left(\beta \dot{\chi}_{s+}^n - H(\psi_s^n, \hat{\psi}, \beta) \right),$$

$$L(\psi_s^n, \hat{\psi}, \alpha) < L(\psi_s^n, \hat{\psi}, \dot{\chi}_{s+}^n) + (\alpha - \dot{\chi}_{s+}^n)\hat{\beta}_s \quad \forall \alpha \ne \dot{\chi}_s^n$$

and

$$L(\psi_s^n, \hat{\psi}, \dot{\chi}_{s+}^n) \to L(\psi_s^n, \psi_s^n, \dot{\chi}_{s+}^n), \quad \hat{\psi} \to \psi_s^n.$$

We added to the original assertion the property which is used in the next section, that is, that χ_t may be chosen piecewise linear. Indeed, such functions are used in the proof; see Freidlin and Wentzell (1984), Section 7.5. The existence of β_s asserted in the lemma also follows from the proof; see Freidlin and Wentzell (1984) or Freidlin (1978). Assertions about $\hat{\psi}$ and $\hat{\beta}_s$ also added to the original assertion can be deduced from the proof using similar arguments.

In fact, there is a little gap in the original proof; namely, an additional assumption was used which was not formulated explicitly. This is why we have to present a precise statement and give necessary comments on it in the Appendix.

4. Proof of Theorem 1. *Step* 1. Denote $H(x, \beta) = H(x, x, \beta)$. The existence of the limit $H(x, \cdot)$ and its differentiability and continuity are asserted in Lemmas 3 and 4.

Step 2. Let $\Delta = T/m$, $m > 0$ an integer. Let ψ, χ be two functions close to φ with the following properties: ψ is a step function and χ is piecewise linear function in accordance with the partition of $[0, T]$ by points $k\Delta$, $k = 1, 2, \dots, m$, and $\rho(\varphi, \psi) + \rho(\varphi, \chi) < \lambda$ where $\lambda > 0$ is small enough.

Denote

$$X_t^{\psi} = x_0 + \int_0^t f(\psi_s, Y_s)\, ds.$$

We have, due to the Lipschitz condition on f,

$$\{\rho(X, \varphi) < \delta\} \supset \{\rho(X^{\psi}, \chi) < \delta'\} \tag{14}$$

if δ' is small enough w.r.t. δ.

Denote $\varphi^{\Delta} = (\varphi_{\Delta}, \varphi_{2\Delta}, \dots, \varphi_{m\Delta})$. Since $\|f\|_C < \infty$, we have

$$\{\rho(X^{\psi}, \chi) < \delta'\} \supset \left\{\rho\big((X^{\psi})^{\Delta}, \chi^{\Delta}\big) < \delta''\right\} \tag{15}$$

if δ'' and Δ are small enough. We can and will assume $\delta'' \geq cm\delta(\nu)$, $c > 0$.

Step 3. We assume $S(\varphi) < \infty$. By virtue of Lemma 6, we can choose a step function ψ and a linear piecewise χ so that

$$\int_0^T L(\psi_s, \dot{\chi}_s)\, ds \leq \int_0^T L(\varphi_s, \dot{\varphi}_s)\, ds + \nu$$

and

$$\sup_s |\dot{\chi}_s| \leq C < \infty.$$

Moreover, for any s there exists β_s such that

$$\beta_s = \operatorname*{argmax}_{\beta} \big(\beta\dot{\chi}_{s+} - H(\psi_s, \psi_s, \beta)\big)$$

and

$$L(\psi_s, \alpha) > L(\psi_s, \dot{\chi}_{s+}) + (\alpha - \dot{\chi}_{s+})\beta_s \quad \forall\, \alpha \neq \dot{\chi}_s.$$

Let us show that we may assume

$$\sup_s |\beta_s| \leq b < \infty$$

with some $b > 0$. Indeed, we can find b such that

$$\int_0^T L(\psi_s, \dot{\chi}_s) I(|\beta_s| \leq b)\, ds \leq \int_0^T L(\varphi_s, \dot{\varphi}_s)\, ds + 2\nu. \tag{16}$$

For any χ_s the function $L(\chi_s, \cdot)$ is equal to zero at some point, which we denote by $\hat{\alpha}_s$. Since $|\nabla_{\beta} H(\psi_s, \cdot)| \leq \|f\|_C$ then $|\hat{\alpha}_s| \leq \|f\|_C$.

Consider a new curve

$$\hat{\chi}_t = \int_0^t \big(\dot{\chi}_s I(|\beta_s| \leq b) + \hat{\alpha}_s I(|\beta_s| > b)\big)\, ds.$$

Let $|\beta_s| > b$. Recall that $L(\psi_s, \cdot) \geq 0$ and this function is equal to zero at a unique point which follows from the differentiability of H. Hence, we can put $\hat{\beta}_s = 0$ (here we use the same notation β_s as above with a "hat" for the curve $\hat{\chi}$). If $|\beta_s| \leq b$, then we take $\hat{\beta}_s = \beta_s$.

If b is large enough, then $\hat{\chi}$ is arbitrarily close to χ. By construction, we have for this new curve

$$\sup_s |\hat{\beta}_s| \leq b < \infty.$$

Moreover, $\hat{\chi}$ is also piecewise linear. Denote this new curve again by χ.

Step 4. Now, let us estimate from below the value

$$E \prod_{k=1}^{m} I\big(|X^{\psi}_{k\Delta} - \chi_{k\Delta}| < \delta'_k\big)$$

rather than $P(\rho((X^{\psi})^{\Delta}, \chi^{\Delta}) < \delta')$. We choose $\delta'' \leq 2T(1 + \|f\|_C)$ and $\delta'_i = (i/m)\delta''$, $i = 1, \dots, m$. Notice that $c\delta(\nu) \leq \delta''/m \leq \delta(\nu)$ $(c > 0)$.

We start with the estimation of the conditional expectation $E(I(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m)|F_{(m-1)\Delta})$ on the set $\{X^{\psi}_{(m-1)\Delta} = \hat{\psi}_{(m-1)\Delta}\}$ with $|\hat{\psi}_{(m-1)\Delta} - \chi_{(m-1)\Delta}| < \delta'_{m-1}$. Let us apply the Cramér transformation of measure. Let $|\beta| \le b$, we will choose this vector a bit later. We get, on the set $\{|X^{\psi}_{(m-1)\Delta} - \chi_{(m-1)\Delta} | < \delta'_{m-1}\}$,

$$E\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)\Big|F_{(m-1)\Delta}\Big)$$
$$= E^{\beta}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)$$
$$\times \exp\Big(-\varepsilon^{-2}\beta\Big(X^{\psi}_{m\Delta} - X^{\psi}_{(m-1)\Delta}\Big) + \varepsilon^{-2}\Delta H^{\varepsilon,\psi}_m(\beta)\Big)\Big|F_{(m-1)\Delta}\Big),$$

where E^{β} is the expectation w.r.t. the measure P^{β} defined on the sigma-field $F_{m\Delta}$ by its density

$$dP^{\beta}/dP = \exp\Big(\varepsilon^{-2}\beta\Big(X^{\psi}_{m\Delta} - X^{\psi}_{(m-1)\Delta}\Big) - \varepsilon^{-2}\Delta H^{\varepsilon,\psi}_m(\beta)\Big)$$

and

$$\varepsilon^{-2}\Delta H^{\varepsilon,\psi}_m(\beta) = \log E\Big(\exp\Big(\varepsilon^{-2}\beta\Big(X^{\psi}_{m\Delta} - X^{\psi}_{(m-1)\Delta}\Big)\Big)\Big|F_{(m-1)\Delta}\Big).$$

By virtue of Lemma 5, we get

$$\begin{aligned} &E\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)\Big|F_{(m-1)\Delta}\Big) \\ &\quad \ge E^{\beta}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big) \\ &\qquad \times \exp\Big(-\varepsilon^{-2}\Delta\beta\big((\chi_{m\Delta} - \chi_{(m-1)\Delta})/\Delta\big) \\ &\qquad - \varepsilon^{-2}\Delta\Big(H\big(\psi_{(m-1)\Delta}, \beta\big) + \nu\Big)\Big)\Big|F_{(m-1)\Delta}\Big). \end{aligned} \tag{17}$$

Step 5. Now choose $\beta(m) = \operatorname{argmax}_{\beta}(\beta\dot{\chi}_{(m-1)\Delta+} - H(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta))$. If δ'' is small enough then it follows from the properties of convex functions and considerations in the proof of Lemma 7.5.2 from Freidlin and Wentzell (1984) that $\hat{\beta}_s$ is uniformly close to β_s; whence we can assume $|\beta(m)| \le b$. Moreover,

$$\beta(m)\dot{\chi}_{(m-1)\Delta+} - H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m)\Big) = L\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \dot{\chi}_{(m-1)\Delta+}\Big).$$

So (17) implies that

$$\begin{aligned} &E\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)\Big|F_{(m-1)\Delta}\Big) \\ &\quad \ge \exp\Big(-\varepsilon^{-2}\Delta\Big(L\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \dot{\chi}_{(m-1)\Delta+}\Big) + \nu\Big)\Big) \\ &\qquad \times E^{\beta(m)}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)\Big|F_{(m-1)\Delta}\Big). \end{aligned} \tag{18}$$

Let us show the bound

$$E^{\beta(m)}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m\big)\Big|F_{(m-1)\Delta}\Big) \ge 1 - \exp(-C\Delta\varepsilon^{-2}) \tag{19}$$

on the set $\{|\hat{\psi}_{(m-1)\Delta} - \chi_{(m-1)\Delta}| < \delta'_{m-1}\}$ if ε is small enough.

There exists a finite number of vectors $v_1, v_2, \dots, v_N$ s.t. $\|v_k\| = 1\ \forall\ k$, $N \le 2d$ and

$$E^{\beta(m)}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| > \delta'_m\big)\Big|F_{(m-1)\Delta}\Big) \le \sum_{k=1}^{N} E^{\beta(m)}\Big(I\Big(\big(X^{\psi}_{m\Delta} - X^{\psi}_{(m-1)\Delta} - \chi_{m\Delta} + \chi_{(m-1)\Delta}\big)v_k > \delta(\nu)\kappa\Big)\Big|F_{(m-1)\Delta}\Big),$$

where $c(2N)^{-1/2} = \kappa$. We estimate, for any $v := v_k$, $0 \le z \le 1$,

$$E^{\beta(m)}\Big(I\Big(\big(X^{\psi}_{m\Delta} - X^{\psi}_{(m-1)\Delta} - \chi_{m\Delta} + \chi_{(m-1)\Delta}\big)v > \delta(\nu)\kappa\Big)\Big|F_{(m-1)\Delta}\Big)$$

$$\le \exp\big(-\delta(\nu)z\kappa\varepsilon^{-2}\big)\exp\Big(\Delta\varepsilon^{-2}\Big[-zv\dot{\chi}_{(m-1)\Delta+} + H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m) + vz\Big) - H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m)\Big) + 2\nu\Big]\Big) \tag{20}$$

if ε is small enough. Denote

$$h(z) := \delta(\nu)\kappa\Delta^{-1}z + \dot{\chi}_{(m-1)\Delta+}vz - \Big[H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m) + vz\Big) - H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m)\Big)\Big].$$

We have $h(0) = 0$. Moreover,

$$h'(0) = \delta(\nu)\kappa\Delta^{-1} + \dot{\chi}_{(m-1)\Delta+}v - \nabla_{\beta}H\Big(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m)\Big)v = \delta(\nu)\Delta^{-1}\kappa = C > 0,$$

for $\dot{\chi}_{(m-1)\Delta+} - \nabla_{\beta}H(\psi_{(m-1)\Delta}, \hat{\psi}_{(m-1)\Delta}, \beta(m)) = 0$. Finally, since $\nabla_{\beta}H$ is bounded and continuous due to Lemma 4, $h'(z) \ge C/2$ for small z, say, for $0 \le z \le z_0$. Then $h(z_0) \ge Cz_0/2$. Hence, the r.h.s. in (20) with $z = z_0$ does not exceed the value

$$\exp\big(\Delta\varepsilon^{-2}(2\nu - h(z))\big) \le \exp\big(-C\Delta z_0\varepsilon^{-2}/4\big]\big)$$

if we choose

$$\nu < Cz_0/8.$$

This gives the bound

$$E^{\beta(m)}\Big(I\big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| \ge \delta'_m\big)\Big|F_{(m-1)\Delta}\Big) \le \exp\big(-C\Delta\varepsilon^{-2}/4\big)$$

which is equivalent to (19). This implies the estimate

$$P\Big(|X^{\psi}_{m\Delta} - \chi_{m\Delta}| < \delta'_m|F_{(m-1)\Delta}\Big) \ge \exp\Big(-\varepsilon^{-2}\Delta\Big(L\Big(\psi_{(m-1)\Delta}, \dot{\chi}_{(m-1)\Delta+}\Big) + \tilde{\nu}\Big)\Big)$$

where $\tilde{\nu}$ is arbitrarily small if δ'' and ν are small enough. By induction, we get

$$P\left(|X^{\psi}_{\Delta m} - \varphi_{\Delta m}| < \delta'_m, \dots, |X^{\psi}_{\Delta} - \varphi_{\Delta}| < \delta'_1\right)$$
$$\geq \exp\left(-\varepsilon^{-2}\Delta \sum_{i=1}^{m} \left(L\left(\psi_{(m-i)\Delta}, \dot{\chi}_{(m-i)\Delta+}\right) + \tilde{\nu}\right)\right)$$
$$= \exp\left(-\varepsilon^{-2}\left(\int_0^T L(\psi_s, \dot{\chi}_s)\, ds + \tilde{\nu} T\right)\right).$$

Finally, by virtue of (16) this gives one

$$P\left(\rho_{0T}(X^{\varepsilon}, \varphi) < \delta\right) \geq \exp\left(-\varepsilon^{-2}\left(S_{0T}(\varphi) + \tilde{\nu} T\right)\right), \qquad \varepsilon \to 0.$$

[cf. Freidlin and Wentzell (1984)]. This bound is uniform in $x \in E^d$, $|y| \leq r$ and $\varphi \in \Phi_x(s)$ for any $r, x > 0$.

Step 6. If the first inequality is established, the proof of the second one is standard; see the corresponding part of Theorem 7.4.1 from Freidlin and Wentzell (1984). Theorem 1 is proved. □

APPENDIX

A. Comments on Lemma 6. To explain that Lemma 6 is valid without additional assumptions, we have to review very briefly its proof and show those assumptions.

Let $0 = t_0 < t_1 < \cdots < t_m = T$ be a partition, $\gamma_k(\beta) := \int_{t_{k-1}}^{t_k} H(\varphi_s, \beta)\, ds$, $l_k(\alpha) = \sup_{\beta}(\alpha\beta - \gamma_k(\beta))$, $A_k = \{\alpha\colon\ l_k(\alpha) < \infty\}$, A_k^o its interior w.r.t. the linear hull L_{A_k}.

The inequality $S(\varphi) = \int_0^T L(\varphi_t, \dot{\varphi}_t)\, dt < \infty$ implies

$$\sum_{k=1}^{m} \sup_{\beta}\left((\varphi_{t_k} - \varphi_{t_{k-1}}) - \gamma_k(\beta)\right) = \sum_{k=1}^{m} l_k(\varphi_{t_k} - \varphi_{t_{k-1}}) \leq S(\varphi).$$

Under the additional assumption $A_k^o \neq \emptyset$, it is proved in Freidlin and Wentzell (1984) using the arguments from Rockafellar (1970) that for any $\nu > 0$, there exists a function $\tilde{\varphi}$ such that $\rho(\varphi, \tilde{\varphi}) < \nu$ and there exist β_k such that

$$l_k\left(\tilde{\varphi}_{t_k} - \tilde{\varphi}_{t_{k-1}}\right) = \left(\tilde{\varphi}_{t_k} - \tilde{\varphi}_{t_{k-1}}\right)\beta_k - \gamma_k(\beta_k) \tag{21}$$

and

$$\tilde{\varphi}_{t_k} - \tilde{\varphi}_{t_{k-1}} = \nabla\gamma_k(\beta_k). \tag{22}$$

The proof goes well if $A_k^o \neq \emptyset\ \forall\ k$.

Let us show that the same is true if $A_k^o = \emptyset$ for some k's. The property $A_k^o = \emptyset$ is equivalent to $\dim L_{A_k} = 0$. In this case, $\gamma_k(\beta) = c_k\beta$ with some $c_k \in R^d$. Hence, $l_k(\alpha_k) < \infty$ means that $l_k(\alpha_k) = 0$ and for any other α, $l_k(\alpha) = +\infty$ and $\gamma_k(\beta) = \alpha_k\beta$. So, we have

$$l_k(\varphi_{t_k} - \varphi_{t_{k-1}}) = 0 = (\varphi_{t_k} - \varphi_{t_{k-1}})\beta - \gamma_k(\beta)$$

for *any* β. Let $\beta_k = 0$. Evidently,

$$\varphi_{t_k} - \varphi_{t_{k-1}}\nabla\gamma_k(\beta_k).$$

Hence, in the case $A_k^o = \emptyset$, one should not just change the curve φ_s on the interval (t_{k-1}, t_k); that is, (21) and (22) are valid in this case also.

The rest of the proof is not changed. For any step function ψ, one defines a piecewise linear χ by the formula

$$\chi_0 = \varphi_0, \qquad \dot{\chi}_s = \nabla_\beta H(\psi_s, \beta_k)), \qquad t_{k-1} < s < t_k, \, k = 1, 2, \dots, m.$$

Then it is shown that $\psi^n \to \varphi$ implies $\chi^n \to \varphi$ due to the property that the convergence of smooth convex functions to the limit implies the convergence of their gradients. Then there exists a partition such that this construction gives one

$$\int_0^T L(\psi_t, \dot{\chi}_t)\, dt \le S(\varphi) + \nu.$$

So, the lemma holds true without additional assumptions. The assertions about $\hat{\psi}$ and $\hat{\beta}_s$ can be shown similarly.

B. Comments on the property $A_k^o \neq \emptyset$. Denote the interior of $A(x)$ w.r.t. $L_{A(x)}$ by $A^o(x)$. Then $A_k^o = \emptyset \Leftrightarrow A^o(\varphi_{t_{k-1}}) = \emptyset$. In this section we show the following equivalence:

$$\operatorname{card}\big(f \in R^d \colon f = f(x, y),\, y \in M\big) = 1 \Leftrightarrow \dim L_{A(x)} = 0 \Leftrightarrow A^o(x) = \emptyset.$$

Since $A(x)$ is convex, clearly the first two conditions are equivalent.

If $\{f(x, \cdot)\}$ contains only one point then $H(x, \beta)$ is linear w.r.t. β; hence, $A(x)$ consists of a unique point and $A^o(x) = \emptyset$.

Now, let $\{f(x, \cdot)\}$ contain at least two different points, say, $f(x, y_1) \neq f(x, y_2)$. Then there exists $1 \le k \le d$ such that $(f(x, y_1) - f(x, y_2))_k \neq 0$. Denote $M_k = \sup_y f^k(x, y)$, $m_k = \inf_y f^k(x, y)$. Let $0 < \nu < (f(x, y_1) - f(x, y_2))_k/2$. Take two points y' and y'' such that $f^k(x, y') < m_k + \nu/3$ and $f^k(x, y'') > M_k - \nu/3$. There exist two open sets $B' \subset M$ and $B'' \subset M$ such that $\sup_{y \in B'} f^k(x, y) < m_k + \nu/2$ and $\inf_{y \in B''} f^k(x, y) > M_k - \nu/2$.

Since the process y_t^x is a nondegenerate ergodic diffusion, there exists such $q > 0$ that

$$P(y_s^x \in B', 0 \le s \le t) \ge q^t, \qquad P(y_s^x \in B'', 0 \le s \le t) \ge q^t, \, t \to \infty.$$

Let $\beta = z\beta_k$ where $\beta_k \in E^d$ is a kth unit coordinate vector and $z \in R$. Then for $z > 0$ we have,

$$\begin{aligned} z^{-1}t^{-1} &\log E \exp\left(z\beta_k \int_0^t f(x, y^x)\, ds \right) \\ &\ge z^{-1}t^{-1} \log E \exp\left(z\beta_k \int_0^t f(x, y^x)\, ds \right) I(y_s^x \in B'', 0 \le s \le t) \\ &\ge z^{-1}t^{-1} \log\big\{\exp(z(M_k - \nu/2))q^t\big\} = M_k - \nu/2 + z^{-1}\log q \ge M_k - \nu \end{aligned}$$

if z is large enough. In other words, for large positive z one has $H(x, z\beta_k) \ge$

$z(M_k - \nu)$. Similarly, for large negative z,

$$
\begin{aligned}
|z|^{-1}t^{-1}\log E\exp\left(z\beta_k\int_0^t f(x, y^x)\,ds\right)& \\
&\geq |z|^{-1}t^{-1}\log E\exp\left(z\beta_k\int_0^t f(x, y^x)\,ds\right)I(y_s^x \in B', 0 \leq s \leq t) \\
&\geq |z|^{-1}t^{-1}\log\left\{\exp(z(m_k + \nu/2))q^t\right\} \\
&= -(m_k + \nu/2) + |z|^{-1}\log q \geq -m_k - \nu
\end{aligned}
$$

if $|z|$ is large enough. In other words, for negative z with large absolute values one has $H(x, z\beta_k) \geq z(m_k + \nu)$. Therefore, $\{\alpha\colon\ \alpha = \beta_k\theta,\ m_k + \nu < \theta < M_k - \nu\} \subset A(x)$.

Similar inequalities are valid for any unit vector β_0 from the linear hull $L_f(x)$ of the set $\{f(x, \cdot)\}$. This shows, in particular, that $\dim L_A(x) = \dim L_f(x)$. Since $A(x)$ is convex, it shows also that the interior $A^o(x)$ w.r.t. $L_{A(x)}$ is not empty.

Hence, the third condition is equivalent to the second and the first.

So, the condition $A_k^o \neq \emptyset$ is always satisfied if the set $\{f(x, \cdot)\}$ for any x consists of more than one point. In fact, if $\operatorname{card}\{f(x, \cdot)\} = 1$ for *any* x then f does not depend on y. In this case, one has nothing to average.

REFERENCES

FREIDLIN, M. I. (1976). Fluctuations in dynamical systems with averaging. *Dok. Acad. Nauk SSSR* **226** 273–276 (in Russian).

FREIDLIN, M. I. (1978). Averaging principle and large deviations. *Uspekhi Mat. Nauk.* **33** 107–160 (in Russian).

FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.

GULINSKY, O. V. and VERETENNIKOV, A. YU. (1993). *Large Deviations for Discrete-Time Processes with Averaging*. VSP, Utrecht.

KATO, T. (1976). *Perturbation Theory for Linear Operators*, 2nd ed. Springer, New York.

KRASNOSEL'SKII, M. A., LIFSHIFTZ, E. A. and SOBOLEV, A. V. (1989). *Positive Linear Systems*. Helderman, Berlin.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

VERETENNIKOV, A. YU. (1992). On large deviations in the averaging principle for stochastic differential equations with periodic coefficients 2. *Math. USSR Izvestiya* **39** 677–701.

VERETENNIKOV, A. YU. (1994). Large deviations in averaging principle for stochastic differential equation systems (noncompact case). *Stochastics Stochastics Rep.* **48** 83–96.

WATANABE, S. and IKEDA, N. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. North-Holland, Amsterdam.

INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS
19 BOLSHOY KARETNII
101447 MOSCOW
RUSSIA
E-MAIL: ayu@sci.lpi.ac.ru