The Annals of Probability
1988, Vol. 16, No. 4, 1509–1536

LARGE DEVIATIONS ANALYSIS OF SOME RECURSIVE ALGORITHMS WITH STATE DEPENDENT NOISE[1]

BY PAUL DUPUIS

Brown University

Consider the recursive, stochastic algorithm $X_{n+1}^{\varepsilon} = X_n^{\varepsilon} + \varepsilon b(X_n^{\varepsilon}, \xi_n)$, where $\{\xi_n\}$ is a random process and X_n^{ε} lives in $\mathbb{R}^d$. Algorithms of this type arise frequently in applications in control and communications, as well as elsewhere. In the study of the important long term behavior of such recursive algorithms the "large deviations" behavior of the system, which describes the asymptotics of the order 1 deviations of the system from its "mean" trajectory as ε tends to 0, plays a central role. Typical systems arising in communication theory and control often use complicated forcing terms involving correlated and state dependent noises and forcing terms with discontinuities. This paper presents a general approach for proving large deviations type theorems for such systems. The problem of proving such a theorem is considered first for the general case of a stochastic process with Lipschitz continuous sample paths. The assumptions are stated in terms of the conditional distribution of time increments of the process. After giving the proof in this general framework, we give several examples (in both continuous and discrete time) of driving terms that satisfy the hypotheses. The results are subsequently extended to a "projected" version of the discrete time model. The paper concludes with an application of the results to an automatic routing mechanism.

1. Introduction. We shall concern ourselves in this paper with proving useful extensions of the large deviations results given by Freidlin [5] and Azencott and Ruget [2]. In [5], the author considered the problem of proving a large deviations principle for the continuous time version of the discrete time dynamical system (living in $\mathbb{R}^d$)

$$X_{n+1}^{\varepsilon} = X_n^{\varepsilon} + \varepsilon b\big(X_n^{\varepsilon}, \xi_n\big), \qquad X_0^{\varepsilon} = x. \tag{1.1}$$

Given any system of the same general form as (1.1), we define $x^{\varepsilon}(\cdot)$ to be the piecewise linear version of X_n^{ε} with interpolation interval ε. Let $C_x[0, T]$ denote the set of continuous paths with the values in $\mathbb{R}^d$ that start at x. We then say that (1.1) obeys a large deviations principle (L.D.P.) if there exists a functional $S_x(T, \phi)$ defined for $\phi \in C_x[0, T]$ and a real valued function $f(\varepsilon)$ tending to 0 as $\varepsilon \to 0$, such that

(i) $S_x(T, \cdot)$ is lower semicontinuous (l.s.c.) on $C_x[0, T]$,
(ii) $\Phi_x(s) \equiv \{\phi \in C_x[0, T]\colon\ S_x(T, \phi) \le s\}$ is compact for $s \in (0, \infty)$,

and if for any Borel set $A \subset C_x[0, T]$,

Received January 1987; revised March 1988.
[1]Work supported in part by National Science Foundation Grant DMS-85-11470 and Office of Naval Research Contract N00014-83-K-0542. Part of the research for this paper was carried out while the author was visiting the IMA at the University of Minnesota, Minneapolis.
AMS 1980 *subject classifications*. Primary 60F10; secondary 62L20.
Key words and phrases. Large deviations, recursive algorithms, asymptotic analysis, state dependent noises.

(iii) $\limsup_\varepsilon f(\varepsilon)\log P_x\{x^\varepsilon \in A\} \le -\inf_{\phi \in \bar{A}} S_x(T, \phi)$,
(iv) $\liminf_\varepsilon f(\varepsilon)\log P_x\{x^\varepsilon \in A\} \ge -\inf_{\phi \in A^0} S_x(T, \phi)$.

Here $\bar{A}$ and A^0 denote the closure and interior of A, respectively, and P_x (E_x, P_ξ, etc.) denotes probability (or expectation) given $X_0^\varepsilon = x$ ($X_0^\varepsilon = x$, $\xi_0 = \xi$, etc.). We note that in all our cases as well as in [5, 2], it turns out that $f(\varepsilon) = \varepsilon$. We note also that a L.D.P. typically implies more than (i)–(iv). Besides the asymptotic relations (iii) and (iv), a L.D.P. can oftentimes be used to obtain asymptotic expressions for many interesting and complicated quantities, such as mean exit times and the locations of exit from a stable region, transition probabilities between neighborhoods of stable points, and invariant measures (all of which are discussed in [6]). In the absence of exact results, these asymptotic formulas may provide the only objective measure of system performance. However, in this paper we shall concern ourselves simply with formulating conditions under which (i)–(iv) hold.

Notation. If the metric space involved is obvious, for a given set S and $\gamma \ge 0$ we denote

$$S^\gamma = \{y: \inf\{d(y,x), x \in S\} \le \gamma\}$$

and

$$S^{-\gamma} = \{y: \inf\{d(y,x), x \notin S\} \ge \gamma\}.$$

Hence $\{x\}^\gamma = \{y: d(y,x) \le \gamma\}$.

Remark. We point out here the equivalence under (ii) of conditions (iii) and (iv) with the following:

(iii′) for any $h > 0$ and $s > 0$, there is $\varepsilon_0 > 0$ such that for $\varepsilon \le \varepsilon_0$

$$\varepsilon \log P_x\{x^\varepsilon \notin \Phi_x(s)^h\} \le -s + h;$$

(iv′) for any $h > 0$ and $\phi \in C_x[0, T]$, there is $\varepsilon_0 > 0$ such that for $\varepsilon \le \varepsilon_0$

$$\varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^h\} \ge -S_x(T, \phi) - h.$$

See [6, page 85]. We will in fact prove (i), (ii), (iii′) and (iv′), which then imply (iii) and (iv).

Important assumptions used in [5] were that each component of $b(x, \xi)$ was bounded and possessed bounded first derivatives, and that the driving noise ξ_n was "state independent" in the sense that for any n and set $A \in \sigma(\xi_i, i > n)$ we have

$$P\{A|\xi_i, i \le n\} = P\{A|\xi_i, X_i^\varepsilon, i \le n\}.$$

We shall refer to such a process ξ_n as *exogenous*. If the process ξ_n is not exogenous, then it will be referred to as *state dependent*.

From the point of view of applications to typical problems in communications, etc., it would be nice to extend these results in the following directions:

(1) *Allow for state dependent noise processes*. In many problems, the statistics of the future evolution of what we consider to be the driving noise ξ_i, $i > n$, depends on X_n^ε, the state of the system.
(2) *Allow $b(x, \xi)$ to be discontinuous*. This extension is of particular importance when dealing with problems arising in communications and control. For example, the "state" may represent some variable used to control a system (such as a gain) and small (ε-sized) adjustments are made depending on whether or not some combination of the state and noise satisfy a given criteria (an event of the correct nature has occurred). In such an application, b might be given by some combination of indicator functions. Often this case can be reduced to the preceding case (1) or the subsequent case (3). For a simple example, suppose the ξ_n are i.i.d. and have a density with respect to Lebesgue measure, and $b(x, \xi) = b(x - \xi)$ with $b(\cdot)$ discontinuous. Then we can consider this to be of the class of state dependent noise processes.
(3) *Allow $b(x, \xi)$ itself to be random*. In many cases b needs to be replaced by a sequence $\{b_n(x, \xi)\}$ of i.i.d. random vector fields defined on $\mathbb{R}^r \times \xi$-space. The system used in [2] may be thought of as being in this form with $b_n(x, \xi) = b_n(x)$. See in particular Section 4.4 and the example in Section 6.

Our principle assumptions A2 and A3 allow problems with all of these features by being defined solely in terms of the conditional statistics of the time increments of the process.

The model. Our system will take the form of a collection $\{x^\varepsilon(\cdot), \varepsilon > 0\}$ of stochastic processes living in $C_x[0, T]$ and defined on a probability space $(\Omega_x, \mathscr{F}_x, P_x)$.

ASSUMPTION A1. For every $\varepsilon > 0$, $x^\varepsilon(\cdot)$ has Lipschitz (with constant K) sample paths (a.s.).

We point out that as far as the distribution of $\{x^\varepsilon(\cdot)\}$ is concerned, it does not matter if we assume $|\dot{x}^\varepsilon(\cdot)| \le K$ a.s. or $|\dot{x}^\varepsilon(\cdot)| \le K$ for all ω, and we choose the latter to avoid writing w.p.1 statements.

The paper is organized as follows:

In Section 2 sufficient hypotheses for (i), (ii) and (iii) in the definition of a L.D.P. are stated and the upper bound is proved. In Section 3 additional assumptions are made and the lower bound (iv) is proved. These hypotheses are in no way trivial to verify for a given class of noise models and, therefore, Section 4 is devoted to proving they hold for certain models of interest. In Section 5 we comment on an extension and then follow with an application in Section 6. We conclude with a few remarks on the l.s.c. of the functionals we use in the Appendix.

To simplify the notation, we take $T = 1$. The results carry over to an arbitrary interval $[0, T]$ in the obvious way.

NOTATION. For a convex function $f(x)$ mapping $\mathbb{R}^d$ to $\mathbb{R}$, we say that z is a subgradient of f at y if

$$f(x) \geq f(y) + \langle x - y, z \rangle$$

for all $x \in \mathbb{R}^d$. We then define

$$\partial f(x) = \{\text{subgradients of } f \text{ at } x\}.$$

For $\phi \in C_x[0,1]$ and $0 \leq t \leq 1 - \Delta$, we define $D^\Delta \phi(t) = \phi(t + \Delta) - \phi(t)$.

2. Upper bounds. In this section and the next we will consider our system $x^\varepsilon(\cdot)$ and assume that the starting value x of the system lies in F, some fixed compact subset of $\mathbb{R}^d$. The resulting upper and lower bounds will be shown to be uniform for such initial conditions. By this we mean that given any $s_0 \geq 0$ and $h > 0$, there is $\varepsilon_0 > 0$ such that for any $x \in F$, any $s \leq s_0$, any $\phi \in C_x[0,1]$ such that $S_x(\phi) \leq s_0$ and any $\varepsilon \leq \varepsilon_0$,

$$\varepsilon \log P_x\{x^\varepsilon \notin \Phi_x(s)^h\} \leq -s + h,$$

$$\varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^h\} \geq -S_x(\phi) - h.$$

Let $\tilde{F} = F^K$. By Assumption A1, $x^\varepsilon(t) \in \tilde{F}$ for all $0 \leq t \leq 1$.

ASSUMPTION A2. There exist a real valued $H(x, \alpha)$, convex in $\alpha \in \mathbb{R}^d$, and upper semicontinuous in x, and a family of σ-algebras $\mathscr{F}^\varepsilon(t) \supset \sigma(x^\varepsilon(s), 0 \leq s \leq t)$, with the following property. Given $\gamma > 0$ there exists $\Delta_0 > 0$ such that for all $\Delta_0 > \Delta > 0$, $t \in [0, 1 - \Delta]$ (uniformly in $x \in F$ and ω a.s.),

$$\begin{aligned} &\limsup_{\varepsilon} (\varepsilon/\Delta) \log E_x\left[\exp\langle \alpha, D^\Delta x^\varepsilon(t)\rangle/\varepsilon \mid \mathscr{F}^\varepsilon(t)\right] \\ &\quad \leq H(x^\varepsilon(t), \alpha) + \gamma(1 + |\alpha|). \end{aligned} \tag{2.1}$$

REMARKS. By the Lipschitz assumption on $x^\varepsilon(\cdot)$, we may always assume $H(x, \alpha) \leq K|\alpha|$. We will denote $H(x, \alpha) + \gamma(1 + |\alpha|)$ by $H_\gamma(x, \alpha)$. Although it is not indicated by the notation, the family $\mathscr{F}^\varepsilon(t)$ may of course depend on x.

For a function $H(\alpha)$, $\alpha \in \mathbb{R}^d$, we define its Legendre (or Cramér) transform as

$$L(\beta) = \sup_\alpha [\langle \alpha, \beta \rangle - H(\alpha)].$$

We will denote this relationship by $L = H^*$. For convenience we collect several properties of the Legendre transform.

LEMMA 2.1. *Let H be a convex function that is finite on $\mathbb{R}^d$. Let $L = H^*$. Then*:

(i) *$L(\beta)$ is convex and l.s.c.*
(ii) $L(\beta) \geq -H(0)$.

(iii) *$L(\beta)$ achieves its minimum of $-H(0)$ at $\bar{\beta}$ if and only if $\bar{\beta} \in \partial H(0)$ and, furthermore, such a $\bar{\beta}$ exists.*

(iv) *If $L(\beta)$ is finite in a neighborhood of β', then $\partial L(\beta')$ is nonempty and $L(\beta') = \langle \alpha, \beta' \rangle - H(\alpha)$ if and only if $\alpha \in \partial L(\beta')$.*

(v) *There is uniqueness in the sense that there is only one convex and l.s.c. function H such that $L = H^*$.*

(vi) *If L achieves its* min *at $\beta = 0$, then $H(\alpha) \geq H(0)$.*

(vii) *$L(\beta) \to \infty$ as $|\beta| \to \infty$.*

PROOF. (i) [11, Theorem 12.2].

(ii) Obvious.

(iii) Existence follows from [11, Theorem 23.4]; the rest follows from the fact that by the definition of the subgradient $H(\alpha) \geq \langle \alpha, \bar{\beta} \rangle + H(0)$ if and only if $\bar{\beta} \in \partial H(0)$.

(iv) [11, Theorem 23.5, Corollary 23.5.1, Theorem 23.4].

(v) [11, Theorem 12.2].

(vi) By [11, Theorem 23.5], $0 \in \partial H(0)$ implies $0 \in \partial L(0)$, so by (iii) the min of H is at $\alpha = 0$.

(vii) This follows from the fact that H is finite in a neighborhood of the origin. □

THEOREM 2.1. *Assume* A1 *and* A2 *and let $L(x, \beta) = (H(x, \alpha))^*$. Define*

$$S_x(\phi) = \int_0^1 L(\phi, \dot{\phi})\, ds$$

if $\phi(0) = x$ and ϕ is absolutely continuous and set $S_x(\phi) = \infty$ otherwise. Then a large deviations upper bound holds in the form of:

(i) *$S_x(\cdot)$ is l.s.c.*

(ii) *$\Phi_x(s)$ is compact for every $s \in (0, \infty)$.*

(iii) *For any Borel set $A \subset C_x[0,1]$,*

$$\limsup_{\varepsilon} \varepsilon \log P_x\{x^\varepsilon \in A\} \leq -\inf_{\phi \in \bar{A}} S_x(\phi).$$

The proofs of (i) and (ii) are given in the Appendix.

We first present several lemmas and then tie them together to obtain an upper bound on $P_x\{x^\varepsilon \in \{\phi\}^\delta\}$ in terms of $S_x(\phi)$ for an appropriately chosen δ. We then show that the theorem follows from this.

The lemma which follows is an obvious generalization of a theorem of Gärtner [6, 7]. In conjunction with A2 it gives a large deviations upper bound in terms of H_γ^*.

LEMMA 2.2. *Let μ^N be a sequence of probability measures on $\mathbb{R}^d$ and let $H(\alpha)$ be a finite valued convex function such that*

$$\limsup_{N \to \infty} \frac{1}{N} \log \int_{\mathbb{R}^d} \exp N\langle \alpha, x \rangle \mu^N(dx) \leq H(\alpha).$$

Let $L = H^*$ and $\Phi(s) = \{\beta\colon L(\beta) \le s\}$. Then, given $h > 0$ and $s > 0$, there is $N_0 < \infty$ such that for $N \ge N_0$,

$$\frac{1}{N}\log \mu^N\big(\mathbb{R}^d \setminus \Phi(s)^h\big) \le -s + h. \tag{2.2}$$

REMARKS. As mentioned previously, the usual form of the upper bound

$$\limsup \frac{1}{N}\log \mu^N(A) \le -\inf_{\beta \in \bar{A}} L(\beta)$$

follows from (2.2) and the compactness of the level sets $\Phi(s)$ [Lemma 2.1(i) and (vii)]. If we consider a collection of sequences $\{\mu^N\}$ that is parametrized in some way, then the estimate (2.2) may be taken as uniform for all values of the parameter for which the convergence in the lim sup is uniform.

Since our goal is upper bounds in terms of $S_x(\phi)$, we must be able to rewrite, in terms of $L(x,\beta) = (H(x,\alpha))^*$, the upper bounds given directly by Assumption A2 and Lemma 2.2, which are in terms of $L_\gamma(x,\beta) = (H_\gamma(x,\alpha))^*$. $H_\gamma \ge H$ implies $L_\gamma \le L$, which does not give a useful bound. It will later be seen that it is sufficient to prove that we can perturb β by a small amount (say β^*) and achieve $L(x, \beta + \beta^*) \le L_\gamma(x,\beta) + \gamma$. We prove this first for the one dimensional case and then generalize. At the same time, we prove an analogous result for the lower bound, for use in Section 3.

LEMMA 2.3. *Let $H_1(\alpha)$ and $H_2(\alpha)$ be convex functions on $\mathbb{R}$. Assume $H_2(\alpha) \le H_1(\alpha) + \gamma|\alpha|$. Let $L_i = H_i^*$, $i = 1,2$. Then, given β there are β^* and β^{**} such that $|\beta^*| \vee |\beta^{**}| \le \gamma$, $L_1(\beta + \beta^*) \le L_2(\beta)$ and $L_1(\beta) \le L_2(\beta + \beta^{**})$.*

PROOF. Let $L_1(\beta)$ achieve its min at $\bar{\beta}$. Assume for now that $\bar{\beta} = 0$. Since $H_2(0) \le H_1(0)$,

$$L_2(\beta) \ge -H_2(0) \ge -H_1(0) = L_1(0)$$

[Lemma 2.1(ii) and (iii)]. Hence for $\beta \in [-\gamma, \gamma]$ we take $\beta^* = -\beta$. Now suppose $\beta > \gamma$. Then

$$\begin{aligned} L_2(\beta) &= \sup_\alpha \left[\alpha\beta - H_2(\alpha)\right] \ge \sup_\alpha \left[\alpha\beta - H_1(\alpha) - \gamma|\alpha|\right] \\ &\ge \sup_{\alpha \ge 0}\left[\alpha\beta - H_1(\alpha) - \gamma|\alpha|\right]. \end{aligned}$$

By Lemma 2.1(vi), $H_1(\alpha) \ge H_1(0)$, and since $\alpha(\beta - \gamma) < 0$ if $\alpha < 0$,

$$L_2(\beta) \ge \sup_\alpha\left[\alpha(\beta-\gamma) - H_1(\alpha)\right] = L_1(\beta - \gamma).$$

Thus we may take $\beta^* = -\gamma$. An analogous argument shows $\beta^* = \gamma$ works when $\beta < -\gamma$.

Now consider the case $\bar{\beta} \ne 0$. We define $\bar{H}_i(\alpha) = H_i(\alpha) + \alpha\bar{\beta}$, $i = 1,2$. Then

$$\bar{L}_i(\beta) = \sup_\alpha\left[\alpha\beta - \bar{H}_i(\alpha)\right] = \sup_\alpha\left[\alpha\beta - H_i(\alpha) - \alpha\bar{\beta}\right] = L_i(\beta - \bar{\beta})$$

and we can now apply the preceding argument since $\bar{L}_i(\beta)$ achieves its min at $\beta = 0$. It is also clear that for $\beta \geq \bar{\beta}$ we can take $\beta^{**} = \gamma$ and for $\beta \leq \bar{\beta}$ we can use $\beta^{**} = -\gamma$. □

LEMMA 2.4. *Let $H_1(\alpha)$ and $H_2(\alpha)$ be convex functions defined on $\mathbb{R}^d$. Assume $H_2(\alpha) \leq H_1(\alpha) + \gamma(1 + |\alpha|)$. Let $L_i = H_i^*$, $i = 1, 2$. Then given β there is β^* such that $|\beta^*| \leq \gamma$ and $L_1(\beta + \beta^*) \leq L_2(\beta) + \gamma$, and there is β^{**} such that $|\beta^{**}| \leq \gamma$ and $L_1(\beta) \leq L_2(\beta + \beta^{**}) + \gamma$.*

PROOF. The first assertion will obviously be true if we show that by taking $H_2(\alpha) \leq H_1(\alpha) + \gamma|\alpha|$ we can obtain β^* such that $|\beta^*| \leq \gamma$ and $L_1(\beta + \beta^*) \leq L_2(\beta)$. Let $v \in \mathbb{R}^d$ have norm 1. Define $h_i^v(a) = H_i(av)$ and note that $h_2^v(a) \leq h_1^v(a) + \gamma|a|$. Let $l_i^v(b) = (h_i^v(a))^*$. We first show that

$$\inf_{\{\beta:\, \langle v, \beta\rangle = b\}} L_i(\beta) = l_i^v(b). \tag{2.3}$$

We consider $i = 1$ and note that the same proof works for $i = 2$. First observe

$$\begin{aligned}\sup_b \left[ab - \inf_{\{\beta:\, \langle v, \beta\rangle = b\}} L_1(\beta)\right] &= \sup_b \sup_{\{\beta:\, \langle v, \beta\rangle = b\}} [ab - L_1(\beta)] \\ &= \sup_\beta [a\langle v, \beta\rangle - L_1(\beta)] \\ &= H_1(av).\end{aligned}$$

It is easily checked that the left-hand side of (2.3) is convex and l.s.c., so that by uniqueness [Lemma 2.1(iv)] (2.3) is true, since both sides have Legendre transform $h_i^v(a) = H_i(av)$.

The first claim of Lemma 2.4 is the same as showing that $H_2(\alpha) \leq H_1(\alpha) + \gamma|\alpha|$ implies that for any $I \in [-H_1(0), \infty]$,

$$\{\beta\colon L_2(\beta) \leq I\} \subset \{\beta\colon L_1(\beta) \leq I\}^\gamma. \tag{2.4}$$

Assume this is not true. Then there are $I, \tilde{\beta}$ such that $\tilde{\beta} \in \{\beta\colon L_2(\beta) \leq I\}$, but $\tilde{\beta} \notin \{\beta\colon L_1(\beta) \leq I\}^\gamma$. Let $\Gamma = \{\beta\colon L_1(\beta) \leq I\}^\gamma$ and let β' be that point in Γ closest to $\tilde{\beta}$. Since Γ is convex and closed, there is an outward normal $\tilde{v}$ (with respect to Γ) pointing from β' to $\tilde{\beta}$ (see Figure 1). From (2.3) we see

$$\langle \tilde{v}, \tilde{\beta}\rangle \in \{b\colon l_2^{\tilde{v}}(b) \leq I\},$$

but that

$$\langle \tilde{v}, \tilde{\beta}\rangle \notin \{b\colon l_1^{\tilde{v}}(b) \leq I\}^\gamma,$$

contradicting Lemma 2.3. To prove the second assertion, we note that in this case the claim is equivalent to showing $H_2(\alpha) \leq H_1(\alpha) + \gamma|\alpha|$ implies that for any $I \in [-H_1(0), \infty]$,

$$\{\beta\colon L_1(\beta) \geq I\} \subset \{\beta\colon L_2(\beta) \geq I\}^\gamma,$$

or that

$$\{\beta\colon L_2(\beta) \leq I\}^{-\gamma} \subset \{\beta\colon L_1(\beta) \leq I\}.$$

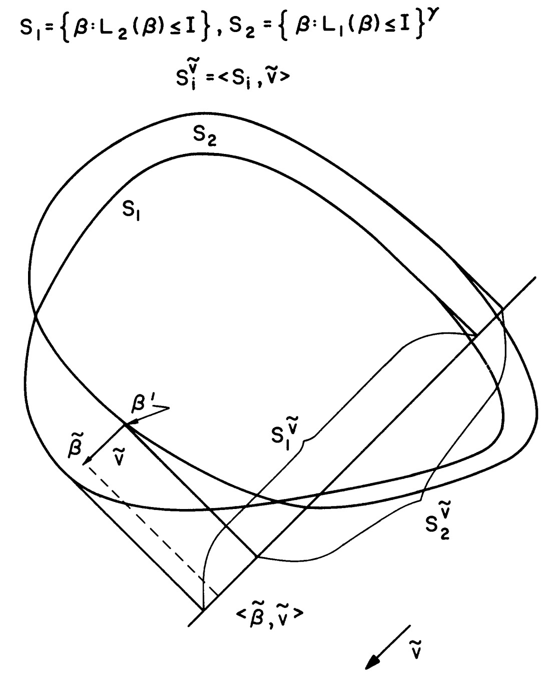

FIG. 1.

But this inclusion is implied by (2.4) and the fact that $\{\beta: L_1(\beta) \le I\}$ is closed and convex. □

We now derive our first upper estimate for $P_x\{x^\varepsilon \in \{\phi\}^\delta\}$. Let $\gamma > 0$ be given. By Assumption A1, $D^t x^\varepsilon(i\Delta) \le K\Delta$ for all $0 \le t \le \Delta$. Let $\Omega_i = \{d(x^\varepsilon(i\Delta), \phi(i\Delta)) \le \gamma\Delta\}$ and $\Pi_i = P_x\{\bigcap_{k \le i} \Omega_k\}$. We have the bound $P_x\{x^\varepsilon \in \{\phi\}^{\gamma\Delta}\} \le \Pi_{1/\Delta}$. We will use

$$\Pi_i = P_x\Big\{\Omega_i \Big| \bigcap_{k \le i-1} \Omega_k\Big\} \Pi_{i-1}.$$

An upper bound for $P_x\{\Omega_i | \bigcap_{k \le i-1} \Omega_k)$ is given by

$$(*) = P_x\Big\{d\big(D^\Delta x^\varepsilon(i\Delta - \Delta), D^\Delta \phi(i\Delta - \Delta)\big) \le 2\gamma\Delta \Big| \bigcap_{k \le i-1} \Omega_k\Big\}.$$

Choose $\Delta > 0$ small enough so that by A2 [with $x_i = x^\varepsilon(i\Delta - \Delta)$],

$$\limsup_{\varepsilon} (\varepsilon/\Delta) \log E_x\Big[\exp\langle \alpha, D^\Delta x^\varepsilon(i\Delta - \Delta)\rangle/\varepsilon \Big| \bigcap_{k \le i-1} \Omega_k\Big] \le H_\gamma(x_i, \alpha).$$

From Lemma 2.2 we obtain [with $N = 1/\varepsilon$, $H(\alpha) = \Delta H_\gamma(x_i, \alpha)$, $L(\beta) = \Delta L_\gamma(x_i, \beta/\Delta)$ and μ^N the measure induced on $\mathbb{R}^d$ by $D^\Delta x^\varepsilon(i\Delta - \Delta)$, conditioned on $\bigcap_{k \le i-1} \Omega_k$]

$$\limsup_{\varepsilon} \varepsilon \log(*) \le - \inf_{\phi_i^\Delta \in \{\phi(i\Delta) - \phi(i\Delta - \Delta)\}^{2\gamma\Delta}} \Delta L_\gamma(x_i, \phi_i^\Delta/\Delta).$$

By Lemma 2.4,

$$\limsup_{\varepsilon} \varepsilon \log(*) \le - \inf_{\phi_i^\Delta \in \{\phi(i\Delta) - \phi(i\Delta - \Delta)\}^{3\gamma\Delta}} \Delta L(x_i, \phi_i^\Delta/\Delta) + \gamma\Delta,$$

which implies

$$\limsup_{\varepsilon} \varepsilon \log \Pi_{1/\Delta} \le -\inf \int_0^1 L(\bar{\phi}, \dot{\phi}^\Delta)\, ds + \gamma, \tag{2.5}$$

where the infimum is over

$$\begin{aligned}\phi^\Delta \in A^\gamma(\phi) = \{\psi \in C_x[0,1] &\text{ such that } \psi \text{ is linear on } [i\Delta - \Delta, i\Delta) \text{ and} \\ &d((\phi(i\Delta) - \phi(i\Delta - \Delta))/\Delta, \dot{\psi}(s)) \le 3\gamma \\ &\text{for } s \in (i\Delta - \Delta, i\Delta),\ i\Delta \le 1\},\end{aligned}$$

$$\bar{\phi} \in B^\gamma(\phi) = \{\psi\colon \psi \text{ is constant on } [i\Delta - \Delta, i\Delta) \text{ and } |\psi(i\Delta - \Delta) - \phi(i\Delta - \Delta)| \le \gamma,\ i\Delta \le 1\}.$$

Since an upper bound for $H(x, \alpha)$ is $K|\alpha|$, $|\beta| > K$ implies $L(x, \beta) = \infty$. It follows that $S_x(\phi) < \infty$ implies ϕ is Lipschitz with constant K. Hence for small enough $\Delta > 0$, $S_x(\phi) < \infty$ implies

$$\sup\Big\{ \sup_{0 \le t \le 1} |\bar{\phi}(t) - \phi(t)| \colon \bar{\phi} \in B^\gamma(\phi)\Big\} \le \gamma, \qquad A^\gamma(\phi) \subset \{\phi\}^{4\gamma}. \tag{2.6}$$

Before tying these estimates together to obtain the proof of Theorem 2.1, we will need the following lemma, which simply formalizes a discussion in Freidlin [5, page 142].

LEMMA 2.5. *Fix $s > 0$ and $\delta > 0$. Then there exists $\eta > 0$ such that if*

(i) $S_x(\phi) \ge s$,
(ii) $\sup_{0 \le t \le 1} d(\psi(t), \phi(t)) \le \eta$ (*here ψ need not be continuous*),

then $\int_0^1 L(\psi, \dot{\phi})\, ds \ge s - \delta$.

PROOF OF THEOREM 2.1. As mentioned in the remark in the Introduction, given that the level sets $\Phi_x(s)$ are compact, the upper bound will follow if we show that given $s > 0$ and $h > 0$ there is $\varepsilon_0 > 0$ such that for $\varepsilon \le \varepsilon_0$,

$$\varepsilon \log P_x\{x^\varepsilon \notin \Phi_x(s)^h\} \le -s + h.$$

Now fix $h > 0$. Choose η according to Lemma 2.5 with $\delta = h$. Take $\gamma = h \wedge (\eta/8)$ and pick Δ small enough so that (2.6) holds. Define R' to be the compact subset of $C_x[0,1]$ containing all paths with Lipschitz constant less than or equal to K and let $R = R' \setminus \Phi_x(s)^h$. Pick a finite $\gamma\Delta$-net for R: $\{\phi_i, 1 \le i \le N\}$. We then use the estimate

$$P_x\{x^\varepsilon \notin \Phi_x(s)^h\} \le \sum_1^N P_x\{x^\varepsilon \in \{\phi_i\}^{\gamma\Delta}\}. \tag{2.7}$$

As shown previously, for each $1 \le i \le N$,

$$\limsup_\varepsilon \varepsilon \log P_x\{x^\varepsilon \in \{\phi_i\}^{\gamma\Delta}\} \le -\inf \int_0^1 L(\bar{\phi}, \dot{\phi})\, ds + \gamma,$$

where the inf is over $\phi \in A^\gamma(\phi_i)$, $\bar{\phi} \in B^\gamma(\phi_i)$. Our choice of γ, Δ implies:

(i) For any $\phi \in A^\gamma(\phi_i)$, $\bar{\phi} \in B^\gamma(\phi_i)$, $\sup_{0 \le t \le 1} d(\bar{\phi}(t), \phi(t)) \le \eta$.
(ii) $\phi \in A^\gamma(\phi_i)$ implies $\phi \notin \Phi_x(s)$, so $S_x(\phi) > s$.
(iii) $\gamma \le h$.

It follows that

$$\limsup_\varepsilon \varepsilon \log P_x\{x^\varepsilon \notin \Phi_x(s)^h\} \le -s + 2h$$

and the theorem is proved. The upper bound is in fact uniform in x and $s \le s_0$, since we can choose a finite $\gamma\Delta$-net for the compact set $\{\phi\colon |\dot{\phi}| \le K$ a.s., $\phi(0) \in F\}$, which contains all possible sample paths starting in F. □

3. Lower bounds. In obtaining the upper bound we were able to *assume* that each sample path of $x^\varepsilon(\cdot)$ of interest was close to *some* element ϕ_i of our $\gamma\Delta$-net of the set R. This fact, together with A2, allowed estimates on a "sampled" system $\{x^\varepsilon(i\Delta),\ i \le 1/\Delta\}$ to be used in lieu of $x^\varepsilon(\cdot)$ in order to obtain the bound. Consider (1.1) and, for (deterministic) ψ, define $x^{\varepsilon,\psi}(\cdot)$ as $x^\varepsilon(\cdot)$ was but with $\psi(n\varepsilon)$ replacing X_N^ε in the b term. In [5], Freidlin used the exogenous nature of the driving noise and a Lipschitz condition on $b(x, \xi)$ in x to obtain a lower bound for $P_x\{x^\varepsilon \in \{\phi\}^\delta\}$, where $\delta > 0$, in terms of $P_x\{x^{\varepsilon,\psi} \in \{\phi\}^{\delta/2}\}$, for a suitably chosen ψ near ϕ. He essentially used the fact that if $X_n^{\varepsilon,\psi}$ is near to $\phi(n\varepsilon)$ and hence $\psi(n\varepsilon)$, then we must have

$$X_{n+1}^{\varepsilon,\psi} = X_n^{\varepsilon,\psi} + \varepsilon b(X_n^{\varepsilon,\psi}, \xi_n) + \text{small error}.$$

This fact and Gronwall's inequality complete the argument, and the particular nature of the noise process never enters into the estimating procedure.

This is not possible in our case. Following Azencott and Ruget [2], the technique we use is to show that it is sufficient to obtain the lower bound

$$\liminf_\varepsilon \varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^\delta\} \ge -S_x(\phi) \tag{3.1}$$

for a restricted class of "nice" paths for which (3.1) can be proved.

We start by proving an analog of Lemma 2.2 for lower bounds (Theorem 3.1). Since the proof is not such an obvious adaptation of Gärtner's proof [7] as was the case for that lemma, we provide the details. Before stating and proving Theorem 3.1, we develop a few needed results in Lemmas 3.1 and 3.2. It will be seen that in order to make the "errors" δ_1 and δ_2 in the lower bound (3.3) as small as desired, it is necessary that we be able to control the jumps in the derivatives of all convex functions $\overline{H}(\alpha)$ that are sufficiently close in a certain sense to a given smooth convex function $H(\alpha)$. The sizes of the jumps at α are given by diam $\partial\overline{H}(\alpha)$ and Lemma 3.2 bounds this quantity for all α of interest.

LEMMA 3.1. *Let a convex function $H(\alpha)$ be given. Let $L = H^*$ and assume $\{\beta\colon L(\beta) < \infty\} = \mathbb{R}^d$. Let $F_1 \subset \mathbb{R}^d$ be compact and let $\varepsilon > 0$ be given. Then there is $\gamma > 0$ depending only on ε, H and F_1, such that if $\overline{H}$ is any convex function satisfying*

$$|H(\alpha) - \overline{H}(\alpha)| \leq \gamma(1 + |\alpha|)$$

and $\overline{L} = \overline{H}^*$, then

$$\sup\{|L(\beta) - \overline{L}(\beta)|\colon \beta \in F_1\} \leq \varepsilon.$$

PROOF. We argue by contradiction. If not, we find $\bar{\varepsilon} > 0$, $\gamma_i \downarrow 0$, $\overline{H}_i$ such that

$$|H(\alpha) - \overline{H}_i(\alpha)| \leq \gamma_i(1 + |\alpha|)$$

and $\beta_i \in F_1$ such that for all i

$$|L(\beta_i) - \overline{L}(\beta_i)| \geq \bar{\varepsilon}.$$

We may assume $\beta_i \to \bar{\beta} \in F_1$. By Lemma 2.4, we may choose β_i^* and β_i^{**} such that $\beta_i^* \to \bar{\beta}$, $\beta_i^{**} \to \bar{\beta}$ and

$$L(\beta_i^*) - \gamma_i \leq \overline{L}_i(\beta_i) \leq L(\beta_i^{**}) + \gamma_i.$$

The continuity of L then implies

$$\lim L(\beta_i) = L(\bar{\beta}) = \lim \overline{L}_i(\beta_i),$$

a contradiction. □

LEMMA 3.2. *Let a continuously differentiable convex function $H(\alpha)$ be given. Let $L = H^*$ and assume that $\{\beta\colon L(\beta) < \infty\} = \mathbb{R}^d$. Let $F_1 \subset \mathbb{R}^d$ be compact and let $\varepsilon > 0$ be given. Then there is $\gamma > 0$ depending only on ε, H and F_1 such that if $\overline{H}$ is any convex function satisfying*

$$|H(\alpha) - \overline{H}(\alpha)| \leq \gamma(1 + |\alpha|)$$

and $\overline{L} = \overline{H}^*$, then

$$\sup\left\{\operatorname{diam} \partial\overline{H}(\alpha)\colon \alpha \in \bigcup_{\beta \in F_1} \partial\overline{L}(\beta)\right\} \leq \varepsilon.$$

There are $\gamma_0 > 0$ and $M < \infty$ (we may take, for example, $\gamma_0 = \gamma$ when $\varepsilon = 1$ in the preceding equation) such that $\gamma \leq \gamma_0$ implies

$$\sup\left\{|\alpha|\colon \alpha \in \bigcup_{\beta \in F_1} \partial\overline{L}(\beta)\right\} \leq M.$$

PROOF. The lemma is a consequence of [11, Theorem 24.5], which states the following. Let f be a convex function that is finite on an open convex set $C \subset \mathbb{R}^d$ and let f_i be convex functions that are finite on C and converge pointwise to f.

Then for any $\varepsilon > 0$, $\bar{x} \in C$ and $x_i \in C$ such that $x_i \to \bar{x}$, there exists $i_0 < \infty$ such that for $i \geq i_0$, $\partial f_i(x_i) \subset \partial f(\bar{x})^\varepsilon$.

To start, we take C to be an open, bounded, convex neighborhood of F_1 and claim that if γ is sufficiently small, then for any $\beta \in F_1$, there is $\bar{\beta} \in F_1$ (depending on $\bar{L}$) such that

$$\partial \bar{L}(\beta) \subset \partial L(\bar{\beta})^\varepsilon,$$

where $\bar{L}$ can be *any* convex function such that $\bar{L} = \bar{H}^*$, $|H - \bar{H}| \leq \gamma(1 + |\alpha|)$. If not, we obtain $\bar{\varepsilon} > 0$, $\gamma_i \downarrow 0$, convex $\bar{H}_i$ such that $|\bar{H}_i - H| \leq \gamma_i(1 + |\alpha|)$ and $\beta_i \in F_1$ (which we may assume converges to $\bar{\beta} \in F_1$) such that for all $\beta \in F_1$,

$$\partial \bar{L}_i(\beta_i) \not\subset \partial L(\beta)^{\bar{\varepsilon}}$$

and, in particular,

$$\partial \bar{L}_i(\beta_i) \not\subset \partial L(\bar{\beta})^{\bar{\varepsilon}}.$$

But $\bar{L}_i$ is finite on C and by the preceding argument (Lemma 3.1), $\bar{L}_i \to L$ on C, contradicting [11, Theorem 24.5].

The second assertion of the lemma is now a consequence of the preceding fact and the boundedness of the *fixed* set $\{|\alpha|: \alpha \in \bigcup_{\beta \in F_1} \partial L(\beta)\}$ [11, Theorem 24.1]. (We may take $M = \sup\{|\alpha|: \alpha \in \bigcup_{\beta \in F_1} \partial L(\beta)\} + 1$.)

Now take $C = \mathbb{R}^d$ and use the fact that $\bar{H} \to H$ on C as $\gamma \to 0$. We have established that the set $\bigcup_{\beta \in F_1} \partial \bar{L}(\beta)$ is contained in the fixed compact set $\{0\}^M$, independent of $\bar{H}$. By precisely the same argument as before, for small enough γ we know that for any $\bar{H}$ such that $|H(\alpha) - \bar{H}(\alpha)| \leq \gamma(1 + |\alpha|)$ and any $\alpha \in \{0\}^M$, there is $\bar{\alpha} \in \{0\}^M$ such that

$$\partial \bar{H}(\alpha) \subset \partial H(\bar{\alpha})^\varepsilon.$$

The differentiability of H implies diam $\partial H(\bar{\alpha}) = 0$ and the lemma is proved. □

From Lemma 3.2 we can obtain our "robust" version of Gärtner's theorem.

THEOREM 3.1. *Let $H(\alpha)$ be a continuously differentiable convex function (and thus finite on $\mathbb{R}^d$). Let $L = H^*$ and assume $\{\beta: L(\beta) < \infty\} = \mathbb{R}^d$. Let $F_1 \subset \mathbb{R}^d$ be compact and let μ^N be a sequence of probability measures. Define*

$$H^N(\alpha) = \frac{1}{N} \log \int_{\mathbb{R}^d} \exp N\langle \alpha, x \rangle \mu^N(dx).$$

Let $\delta_1 > 0$, $\delta_2 > 0$ be given. Then there is $\gamma > 0$ such that if

$$H_{-\gamma}(\alpha) \leq \liminf H^N(\alpha) \leq \limsup H^N(\alpha) \leq H_\gamma(\alpha), \tag{3.2}$$

then for all $\beta \in F_1$,

$$\liminf \frac{1}{N} \log \mu^N(\{\beta\}^{\delta_1}) \geq -L(\beta) - \delta_2. \tag{3.3}$$

Proof. We first prove that (3.2) implies $\{H^N\}$ is precompact in the topology of uniform convergence on compact subsets of $\mathbb{R}^d$. Let $F_2 \subset \mathbb{R}^d$ be compact and let R be such that $F_2 \subset \{0\}^R$. By (3.2) we can find $\bar{h}$, $\underline{h}$ such that (at least for large N_0)

$$(3.4) \qquad -\infty < \underline{h} \le \inf_{N \ge N_0,\, \alpha \in \{0\}^{2R}} H^N(\alpha) \le \sup_{N \ge N_0,\, \alpha \in \{0\}^{2R}} H^N(\alpha) \le \bar{h} < \infty.$$

The convexity of H^N and (3.4) imply that H^N is Lipschitz on $\{0\}^R$ with a constant no bigger than $(\bar{h} - \underline{h})/R$, and precompactness follows from Ascoli's theorem.

Choose $M < \infty$ such that for $|H - \overline{H}| \le \gamma(1 + |\alpha|)$, $\overline{L} = \overline{H}^*$, $\beta \in F_1^{\delta_1}$ and $\alpha \in \partial \overline{L}(\beta)$, we have $|\alpha| \le M$, and such that M is independent of γ for small γ (Lemma 3.2). Pick $\gamma \le \delta_1 \wedge \delta_2$ according to Lemma 3.2 so that for any convex H such that $H_{-\gamma} \le H \le H_\gamma$ and any $\alpha \in \{0\}^M$ we have

$$\operatorname{diam} \partial \overline{H}(\alpha) \le \delta_1 \wedge (\delta_2/M).$$

Take any subsequence of N and extract a further subsequence, again denoted by N, such that H^N converges. If we prove (3.3) for this subsequence, then it will be true in general. Let $\overline{H}(\alpha) = \lim H^N(\alpha)$. Being a limit of convex functions, $\overline{H}$ is itself convex. By (3.2),

$$H_{-\gamma}(\alpha) \le \overline{H}(\alpha) \le H_\gamma(\alpha).$$

Define $\overline{L} = \overline{H}^*$ and let $\beta^* \in F_1$ be given. By Lemma 2.4, there is β' such that $|\beta^* - \beta'| \le \delta_1$ and $\overline{L}(\beta') \le L(\beta^*) + \delta_2$. As β' is interior to $\{\beta\colon \overline{L}(\beta) < \infty\}$, there exists $\alpha' \in \partial \overline{L}(\beta')$ so that

$$\overline{L}(\beta') = \sup_\alpha \left[\langle \alpha, \beta' \rangle - \overline{H}(\alpha)\right] = \langle \alpha', \beta' \rangle - \overline{H}(\alpha')$$

and $\beta' \in \partial \overline{H}(\alpha')$ [Lemma 2.1(iv)]. We fix β' henceforth.

Following Gärtner, we now define the probability measures

$$\mu^{N,\alpha'}(\Gamma) = \int_\Gamma \exp N(\langle \alpha', y \rangle - H^N(\alpha'))\mu^N(dy).$$

By defining

$$H^{N,\alpha'}(\alpha) = \frac{1}{N} \log \int_{\mathbb{R}^d} \exp N\langle \alpha, x \rangle \mu^{N,\alpha'}(dx),$$

we see

$$H^{N,\alpha'}(\alpha) = H^N(\alpha + \alpha') - H^N(\alpha').$$

This implies

$$\overline{H}^{\alpha'}(\alpha) \equiv \lim H^{N,\alpha'}(\alpha) = \overline{H}(\alpha + \alpha') - \overline{H}(\alpha'),$$
$$\overline{L}^{\alpha'}(\beta) = \left(\overline{H}^{\alpha'}(\alpha)\right)^* = \overline{L}(\beta) - \left[\langle \alpha', \beta \rangle - \overline{H}(\alpha')\right].$$

Since

$$\overline{L}(\beta) > \left[\langle \alpha', \beta \rangle - \overline{H}(\alpha')\right]$$

when $\beta \notin \partial \overline{H}(\alpha')$ [Lemma 2.1(iv)], we have $\overline{L}^{\alpha'}(\beta) = 0$ if and only if $\beta \in \partial \overline{H}(\alpha')$. It follows from this and Lemma 2.2 that for any $\delta' > 0$, $\mu^{N,\alpha'}(\mathbb{R}^d \setminus \partial \overline{H}(\alpha')^{\delta'}) \to 0$ [and therefore $\mu^{N,\alpha'}(\partial \overline{H}(\alpha')^{\delta'}) \to 1$] as $N \to \infty$.

We now use the equality

$$\mu^N\big(\partial \overline{H}(\alpha')^{\delta}\big) = \int_{\partial \overline{H}(\alpha')^{\delta}} \exp N\big(-\langle \alpha', y\rangle + H^N(\alpha')\big)\mu^{N,\alpha'}(dy). \tag{3.5}$$

Since we have bounded the diameter of $\partial \overline{H}(\alpha')$ by δ_2/M and since $\beta \in \partial \overline{H}(\alpha')$ implies $\overline{L}(\beta) = \langle \alpha', \beta\rangle - \overline{H}(\alpha')$, we have

$$\overline{L}(\beta) = \overline{L}(\beta') + \langle \alpha', \beta - \beta'\rangle \le \overline{L}(\beta') + \delta_2$$

for $\beta \in \partial \overline{H}(\alpha')$. Now let $\delta' = (\delta_2/|\alpha'|) \wedge \delta_1$. Then from (3.5)

$$\mu^N\big(\partial \overline{H}(\alpha')^{\delta_1}\big) \ge \mu^{N,\alpha'}\big(\partial \overline{H}(\alpha')^{\delta'}\big)\exp - N \sup_{\beta \in \partial \overline{H}(\alpha')} \big[\langle \alpha', \beta\rangle - H^N(\alpha')\big]\exp - N\delta_2.$$

We thus obtain [since diam $\partial \overline{H}(\alpha') \le \delta_1$]

$$\liminf \frac{1}{N} \log \mu^N\big(\{\beta'\}^{2\delta_1}\big) \ge -\overline{L}(\beta') - 2\delta_2$$

and, therefore,

$$\liminf \frac{1}{N} \log \mu^N\big(\{\beta^*\}^{3\delta_1}\big) \ge -L(\beta^*) - 3\delta_2,$$

and the theorem is proved. □

REMARKS. Each of Lemmas 3.1 and 3.2 and Theorem 3.1 has an obvious analog in the case where

$$S = \{\beta \colon L(\beta) < \infty\} \ne \mathbb{R}^d.$$

In this case, however, we must choose our compact set to satisfy $F_1 \subset S^0$. Since we can possibly have paths ϕ such that $S_x(\phi) < \infty$ and yet for which $\dot{\phi}(t)$ lies on the boundary of $\{\beta \colon L(\phi(t), \beta) < \infty\}$, we encounter technical difficulties. This will be circumvented in what follows by the introduction of the $x^{\varepsilon,\delta}$ process.

As was the case in Lemma 2.2, for collections of sequences that are parametrized in some way, the estimate (3.3) is uniform for all values of the parameter for which the convergence in (3.2) is uniform.

We turn now to the proof of the lower bound. We first give our assumptions and the statement of the theorem. We then introduce and explain the use of the $x^{\varepsilon,\delta}$ process. Finally we present several lemmas and finish this section with the proof of Theorem 3.2. We recall the definition $H_\gamma(x, \alpha) = H(x, \alpha) + \gamma(1 + |\alpha|)$.

The appropriate assumptions required are:

ASSUMPTION A3. There exist a constant $\overline{K} < \infty$ and real valued $H(x, \alpha)$, convex and continuously differentiable in $\alpha \in \mathbb{R}^d$ and continuous in x, and a family of σ-algebras $\mathscr{F}^\varepsilon(t) \supset \sigma(x^\varepsilon(s), 0 \le s \le t)$ having the following properties. $|H(x_1, \alpha) - H(x_2, \alpha)| \le \overline{K}|x_1 - x_2|(1 + |\alpha|)$ and, for all $\Delta > 0$, $t \in [0, 1 - \Delta]$

(uniformly in $x \in F$ and ω a.s.),

(3.6) $$\limsup(\varepsilon/\Delta)\log E_x\left[\exp\langle\alpha, D^\Delta x^\varepsilon(t)\rangle/\varepsilon|\mathscr{F}^\varepsilon(t)\right] \leq H_{\bar{K}\Delta}(x^\varepsilon(t), \alpha),$$

(3.7) $$\liminf(\varepsilon/\Delta)\log E_x\left[\exp\langle\alpha, D^\Delta x^\varepsilon(t)\rangle/\varepsilon|\mathscr{F}^\varepsilon(t)\right] \geq H_{-\bar{K}\Delta}(x^\varepsilon(t), \alpha).$$

REMARKS. We have made the assumption that

(3.8) $$|H(x,\alpha) - H(\bar{x},\alpha)| \leq \bar{K}|x - \bar{x}|(1 + |\alpha|).$$

Let $L(x,\beta) = (H(x,\alpha))^*$ and $S(x) = \{\beta\colon L(x,\beta) < \infty\}$. It follows from (3.8) and Lemma 2.4 that $S(x)$ is Lipschitz in the Hausdorff topology. We note this fact because it is often easy to characterize $S(x)$ in terms of the statistics of the forcing terms and, hence, provide a simple necessary condition for A3 to hold. For example if the system is the interpolated version of $X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon b_n(X_n^\varepsilon)$, where $\{b_n(\cdot)\}$ is a sequence of i.i.d. random vector fields (Section 4.2), then one can easily prove that $S(x)$ is the convex hull of the support of the distribution of $b_n(x)$.

THEOREM 3.2. *Assume* A1 *and* A3 *and define* $S_x(\phi)$ *as in Theorem* 2.1. *Then, in addition to the conclusions of Theorem* 2.1, *a large deviations lower bound holds in the form of*,

(iv) *for any Borel set* $A \subset C_x[0,1]$,
$$\liminf_\varepsilon \varepsilon \log P_x\{x^\varepsilon \in A\} \geq -\inf_{\phi \in A^0} S_x(\phi).$$

The $x^{\varepsilon,\delta}$ *process.* Let $w(\cdot)$ be a standard $\mathbb{R}^d$ Wiener process starting at 0 at time $t = 0$ that is independent of x^ε. Let

$$x^{\varepsilon,\delta} = x^\varepsilon + \varepsilon^{1/2}\delta w,$$
$$H^\delta(x,\alpha) = H(x,\alpha) + \frac{\delta^2}{2}|\alpha|^2,$$
$$L^\delta(x,\beta) = \left(H^\delta(x,\alpha)\right)^*,$$
$$S_x^\delta(\phi) = \int_0^1 L^\delta(\phi,\dot{\phi})\,ds.$$

Our proof will make use of the $x^{\varepsilon,\delta}$ process. The advantage of dealing with $x^{\varepsilon,\delta}$ is that the presence of the quadratic term in H^δ implies that $L^\delta(x,\beta)$ is finite for all $\beta \in \mathbb{R}^d$ and, as we will see, continuous in $x \in \tilde{F}$, $\beta \in \mathbb{R}^d$. This allows the convergence that will be obtained in (3.12). The "error" due to the introduction of the Wiener process term will be shown to be negligible from the point of view of large deviations.

REMARK. As pointed out previously, $L(x,\beta) = \infty$ if $|\beta| > K$. We shall therefore only have to consider paths ϕ for which $|\dot{\phi}| \leq K$ a.s.

LEMMA 3.3. *Let* $\delta > 0$. *Then* $L^\delta(x,\beta)$ *is continuous for* $x \in \tilde{F}$, $\beta \in \mathbb{R}^d$.

PROOF. The lemma is a consequence of A3, Lemma 3.1 and the continuity of $L^\delta(x,\beta)$ in β (uniformly in $x \in \tilde{F}$ for any fixed β). □

The following lemma shows that in proving a lower bound for the $x^{\varepsilon,\delta}$ process, we can restrict ourselves to a set of "nice" paths.

LEMMA 3.4. *Let $\bar{L}(x,\beta)$ be continuous and let $\bar{S}_x(\phi)$ be defined as usual. Let $h > 0$. Then given ϕ such that $|\dot{\phi}(t)| \le K$ a.s., there is ϕ^h such that*:

(i) *ϕ^h is piecewise differentiable.*
(ii) *$\phi^h \in \{\phi\}^h$.*
(iii) *$\bar{S}_x(\phi^h) \le \bar{S}_x(\phi) + h$.*
(iv) *$|\dot{\phi}^h| \le K$.*

PROOF. The proof is similar to [2, Lemma 4.4]. Let $\gamma > 0$. We can find a finite number of disjoint Borel measurable sets A_i and vectors v_i such that $|v_i| \le K$ and

$$\left|\dot{\phi}(\cdot) - \sum_{i=1}^{n} v_i I_{A_i}(\cdot)\right| \le \gamma.$$

We can then choose B_i that are the union of a finite number of intervals and that are disjoint and satisfy

$$\sum_{i=1}^{n} m(A_i \vartriangle B_i) \le \gamma,$$

where m is Lebesque measure and $A_i \vartriangle B_i = A_i \setminus B_i \cup B_i \setminus A_i$. Define

$$\dot{\phi}^h(t) = \sum_{i=1}^{n} v_i I_{B_i}(t), \qquad \phi^h(0) = \phi(0).$$

Then since $m\{t\colon |\dot{\phi} - \dot{\phi}^h| > \gamma\} \le \gamma$,

$$|\phi(t) - \phi^h(t)| \le \gamma(K+1)$$

on $[0,1]$. Since $\bar{L}(x,\beta)$ is uniformly continuous on $\tilde{F} \times \{0\}^K$, we obtain the desired result when γ is sufficiently small. □

COROLLARY 3.1 (to Theorem 3.1). *Fix $\delta > 0$, $\delta_1 > 0$, $\delta_2 > 0$ and a compact set F_1. Assume* A3 *and let θ be a $N(0,I)$ Gaussian random variable that is independent of $x^\varepsilon(\cdot)$. Then there exists $\Delta_0 > 0$ such that for any $\Delta_0 > \Delta > 0$ and any β such that $\beta \in F_1$,*

$$\begin{aligned}(3.9)\qquad &\liminf(\varepsilon/\Delta)\log P_x\left\{\left(D^\Delta x^\varepsilon(t)/\Delta\right) + \left(\varepsilon^{1/2}\delta\theta/\Delta^{1/2}\right) \in \{\beta\}^{\delta_1} | \mathscr{F}^\varepsilon(t)\right\}\\ &\qquad \ge -L^\delta(x^\varepsilon(t),\beta) - \delta_2\end{aligned}$$

uniformly in ω, $x \in F$ and $0 \le t \le 1 - \Delta$.

PROOF. Since

$$\begin{aligned}&E_x \exp\left[\left\langle \alpha, D^\Delta x^\varepsilon(t)/\varepsilon + \Delta^{1/2}\delta\theta/\varepsilon^{1/2}\right\rangle | \mathscr{F}^\varepsilon(t)\right]\\ &\qquad = E_x\left[\exp\langle \alpha, D^\Delta x^\varepsilon(t)/\varepsilon\rangle | \mathscr{F}^\varepsilon(t)\right] \exp \Delta\delta^2|\alpha|^2/2\varepsilon,\end{aligned}$$

the result is a consequence of Assumption A3, Theorem 3.1 and the definition of $H^{\delta}(x, \alpha)$. □

PROOF OF THEOREM 3.2. Let $\delta_1 > 0$, $\delta_2 > 0$ and ϕ be given. We may assume $S_x(\phi) < \infty$, since otherwise there is nothing to prove. Let $S_x(T, \phi)$ denote the functional defined as $S_x(\phi)$ was, but with $[0, T]$ replacing $[0, 1]$.

The first step is to show that there is $T > 0$ (in the following text, it will be seen that it is sufficient to consider any $T \le 1/10\overline{K}$) such that given $\delta_1' > 0$, $\delta_2' > 0$, there is $\delta_3' > 0$ such that

$$\liminf_{\varepsilon} \varepsilon \log P_x\{x^{\varepsilon} \in \{\phi\}^{\delta_1'}\} \ge -S_{\phi(0)}(T, \phi) - \delta_2',$$

uniformly in ω and x such that $|x - \phi(0)| \le \delta_3'$. (We consider x^{ε} only on the interval $[0, T]$ here.) Pick δ such that $(\delta_1')^2/50\delta^2\, dT \ge S_{\phi(0)}(T, \phi) + 1$. Without loss of generality we may assume ϕ satisfies (i)–(iv) of Lemma 3.4 on $[0, T]$. By Lipschitz continuity, for $\Delta \le \delta_1'/10K$, we have

$$|D^t x^{\varepsilon}(i\Delta)| \le \delta_1'/10, \qquad |D^t \phi(i\Delta)| \le \delta_1'/10$$

for $t \in [0, \Delta]$. Also, since ϕ is piecewise differentiable, by picking Δ smaller (if necessary) we can ensure $|D^t \dot{\phi}(i\Delta)| \le \delta_1'/5T$ for $t \in [i\Delta, i\Delta + \Delta]$.

Now define the sets

$$\Omega_i = \left\{D^{\Delta} x^{\varepsilon, \delta}(i\Delta)/\Delta \in \{\dot{\phi}(i\Delta)\}^{\delta_1'/5T}\right\}$$

and let $\Pi_i = P_x\{\bigcap_{k \le i} \Omega_k\}$. We will again use

$$\Pi_i = P_x\left\{\Omega_i \Big| \bigcap_{k \le i-1} \Omega_k\right\} \Pi_{i-1}.$$

Since $|\phi(i\Delta + \Delta) - \phi(i\Delta) - \Delta\dot{\phi}(i\Delta)| \le \Delta\delta_1'/5T$, it is clear that if $|x - \phi(0)| \le \delta_1'/5$, then on the set $\bigcap_{k \le i} \Omega_k$ we have

$$|x^{\varepsilon, \delta}(i\Delta + \Delta) - \phi(i\Delta + \Delta)| \le 3\delta_1'/5.$$

It follows that off the set where

$$\sup_{0 \le t \le T} |\varepsilon^{1/2} \delta w(t)| > \delta_1'/5, \tag{3.10}$$

the Lipschitz continuity of x^{ε} and ϕ imply

$$\sup_{0 \le t \le T} |x^{\varepsilon}(t) - \phi(t)| \le \delta_1'.$$

The probability of the set given by (3.10) is smaller than

$$2d \exp - (\delta_1')^2/50\delta^2\, dT\varepsilon \le 2d \exp - \left(S_{\phi(0)}(T, \phi) + 1\right)/\varepsilon$$

[12, Theorem 4.2.1] and hence is negligible for small ε. We can therefore use $\varepsilon \log \Pi_{T/\Delta}$ as a lower bound for $\varepsilon \log P_x\{x^{\varepsilon} \in \{\phi\}^{\delta_1'}\}$, when ε is small.

Using the fact that on $\bigcap_{k \le i-1} \Omega_k$ [and off the set where (3.10) holds] we have $|x^{\varepsilon}(i\Delta) - \phi(i\Delta)| \le \delta_1'$, Assumption A3, and Lemma 2.4, we can obtain $\beta(i\Delta)$ depending on $x^{\varepsilon}(i\Delta)$, $\phi(i\Delta)$ and $\dot{\phi}(i\Delta)$ such that $|\beta(i\Delta) - \dot{\phi}(i\Delta)| \le \overline{K}\delta_1' \le$

$\delta_1'/10T$ and

$$L^\delta(x^\varepsilon(i\Delta), \beta(i\Delta)) \geq L^\delta(\phi(i\Delta), \dot{\phi}(i\Delta)) - \overline{K}\delta_1'.$$

Consider the subset Ω_i' of Ω_i given by $\{D^\Delta x^{\varepsilon,\delta}(i\Delta)/\Delta \in \{\beta(i\Delta)\}^{\delta_1'/10T}\}$. Now choose γ via Corollary 3.1 for our $\delta_1'/10T, \delta_2'$ (equal to δ_1, δ_2 there), and any $\beta \in \{0\}^{K+\overline{K}\delta_1'}$. Again pick Δ smaller to ensure $K\Delta \leq \gamma$. We then have

$$\begin{aligned}\liminf_\varepsilon \varepsilon \log P_x\Big\{\Omega_i \Big| \bigcap_{k\leq i-1} \Omega_k\Big\} &\geq \liminf_\varepsilon \varepsilon \log P_x\Big\{\Omega_i' \Big| \bigcap_{k\leq i-1} \Omega_k\Big\} \\ &\geq \Delta\big(-L^\delta(x^\varepsilon(i\Delta), \beta(i\Delta)) - \delta_2'\big) \\ &\geq \Delta\big(-L^\delta(\phi(i\Delta), \dot{\phi}(i\Delta)) - \delta_2' - \overline{K}\delta_1'\big).\end{aligned}$$

This implies

$$\begin{aligned}&\liminf_\varepsilon \varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^{\delta_1'}\} \\ &\quad\geq \liminf_\varepsilon \varepsilon \log \Pi_{T/\Delta} \\ &\quad\geq -\left(\sum_{i=0}^{(T/\Delta)-1} \Delta L^\delta\big(\phi(i\Delta), \dot{\phi}(i\Delta)\big)\right) - T\big(\delta_2' + \overline{K}\delta_1'\big).\end{aligned} \tag{3.11}$$

By virtue of the special properties of ϕ satisfying (i)–(iv) of Lemma 3.2 and the continuity of L^δ on $\tilde{F} \times \mathbb{R}^d$, the right-hand side of (3.11) converges to

$$-\int_0^T L^\delta(\phi, \dot{\phi})\, ds - T\big(\delta_2' + \overline{K}\delta_1'\big) \tag{3.12}$$

as $\Delta \to 0$. This proves our assertion.

The proof for the interval $[0,1]$ now follows by dividing the interval into $1/T$ subintervals, and using the fact that our estimate on $[0, T]$ holds uniformly for x^ε starting in a neighborhood of $\phi(0)$. Suppose for simplicity that $1/T = 2$. In the interval $[T, 2T]$ let $\delta_2' = \delta_2/2$, $\delta_1' = \delta_1$ and obtain $\delta_3' > 0$. In the interval $[0, T]$ let $\delta_2' = \delta_2/2$, $\delta_1' = \delta_3'$. Then

$$\begin{aligned}&\liminf_\varepsilon \varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^{\delta_1}\} \\ &\quad\geq \liminf_\varepsilon \varepsilon \log P_x\Big\{\sup_{0\leq t\leq T} |x^\varepsilon(t) - \phi(t)| \leq \delta_3' \text{ and } \sup_{T\leq t\leq 2T} |x^\varepsilon(t) - \phi(t)| \leq \delta_1\Big\} \\ &\quad\geq \liminf_\varepsilon \varepsilon \log P_x\Big\{\sup_{0\leq t\leq T} |x^\varepsilon(t) - \phi(t)| \leq \delta_3'\Big\} \\ &\qquad + \liminf_\varepsilon \varepsilon \log P_{x^\varepsilon(T)}\Big\{\sup_{T\leq t\leq 2T} |x^\varepsilon(t) - \phi(t)| \leq \delta_1 \Big| |(x^\varepsilon(T) - \phi(T))| \leq \delta_3'\Big\} \\ &\quad\geq -S_x(T, \phi) - S_{\phi(T)}(T, \phi(\cdot + T)) - \delta_2 \\ &\quad= -S_x(\phi) - \delta_2.\end{aligned}$$

Finally, we consider the uniformity issue. Let $\delta > 0$ be given. From the preceding proof and the l.s.c. of $S_{\phi(0)}(\phi)$ (see the Appendix), we may associate to each $\phi \in \mathscr{K} \equiv \{\phi: S_x(\phi) \le s,\ x \in F\}$ positive numbers $\gamma(\phi) \le \delta$ and $\varepsilon_0(\phi)$ such that for $|\phi(0) - x| \le \gamma(\phi)$ and $0 < \varepsilon < \varepsilon_0(\phi)$ we have $\varepsilon \log P_x\{x^\varepsilon \in \{\phi\}^\delta) \ge -S_{\phi(0)}(\phi) - \delta$, and for every $\psi \in \{\phi\}^{\gamma(\phi)}$ we have $S_{\psi(0)}(\psi) \ge S_{\phi(0)}(\phi) - \delta$. By the compactness of $\mathscr{K}$ there is a finite collection $\{\phi_i,\ 1 \le i \le N\} \subset \mathscr{K}$ such that $\mathscr{K} \subset \bigcup_{i=1}^N \{\phi_i\}^{\gamma(\phi_i)}$. Define $\varepsilon_0 = \bigwedge_{i=1}^N \varepsilon_0(\phi_i)$. Then given $\phi \in \mathscr{K}$, one of the ϕ_i's is within $\gamma(\phi_i)$ of ϕ and, hence, for $0 < \varepsilon < \varepsilon_0$,

$$\begin{aligned} \varepsilon \log P_{\phi(0)}\left\{x^\varepsilon \in \{\phi\}^{2\delta}\right\} &\ge \varepsilon \log P_{\phi(0)}\left\{x^\varepsilon \in \{\phi_i\}^{\delta}\right\} \\ &\ge -S_{\phi_i(0)}(\phi_i) - \delta \\ &\ge -S_{\phi(0)}(\phi) - 2\delta. \end{aligned}$$ □

4. On Assumptions A2 and A3. While the preceding sections provide a good notion of some sufficient properties that a stochastic process with Lipschitz paths must possess in order to have a L.D.P., it is typically difficult to formulate general conditions under which the assumptions hold and it can also be difficult to obtain computable formulas for H and L.

In this section we shall show that the large deviations results presented in [5] and [2] can both be obtained as special cases of the theorems presented here and that the upper bound in fact holds under weaker conditions. We then consider models not covered by either [5] or [2], which involve state dependent noise, and indicate when A2 and A3 hold. The examples presented merely suggest the possibilities, without attempting to be complete.

4.1. In [5] Freidlin considered the model (1.1) [with $|b(\cdot,\cdot)|$ bounded and exogenous noise] and assumed the following: For every pair of piecewise constant functions $\phi(\cdot)$, $\alpha(\cdot)$ mapping $[0,1]$ to $\mathbb{R}^d$, we have

$$\begin{aligned} \lim \frac{1}{N} \log E \exp\left[\sum_{i=1}^{N} \langle \alpha(i/N), b(\phi(i/N), \xi_i)\rangle\right] \\ = \int_0^1 H(\phi(s), \alpha(s))\, ds, \end{aligned} \tag{4.1}$$

for some $H(x, \alpha)$, convex and differentiable in α and continuous in x and that $b(x, \xi)$ and its first derivatives are bounded. For our assumptions we take $\mathscr{F}^\varepsilon(t) = \sigma(\xi_i, i \le [t/\varepsilon] + 1)$. A3 implies (4.1) but the converse is not true and A2 does not imply (4.1).

Our assumptions do not completely cover this case since we assume that the convergence in A2 and A3 is uniform conditioned on $\mathscr{F}^\varepsilon(t)$. Freidlin uses only (4.1) and does not need the uniform convergence in the data. However, it seems likely that a proof that (4.1) itself is true would require this uniformity. It should be noted that in any case such uniformity assumptions are necessary if one wishes to use theorems on large deviations to prove more complicated results, such as asymptotics for the mean exit time and locations of exit from a stable domain.

However, if we add the uniformity assumption, then one can easily show that Freidlin's case fits directly into our scheme. Let K' be the Lipschitz constant for $b(x, \xi)$ and let K bound $|b|$. Then

$$\limsup(\varepsilon/\Delta)\log E\left[\exp\left\langle \alpha, \sum_{i=t/\varepsilon}^{(t+\Delta)/\varepsilon} b(X_i^\varepsilon, \xi_i)\right\rangle | \mathscr{F}^\varepsilon(t)\right]$$

$$\leq \limsup(\varepsilon/\Delta)\log E\left[\exp\left(\left\langle \alpha, \sum_{i=t/\varepsilon}^{(t+\Delta)/\varepsilon} b\left(X_{t/\varepsilon}^\varepsilon, \xi_i\right)\right\rangle + K'K\Delta^2|\alpha|/\varepsilon\right) | \mathscr{F}^\varepsilon(t)\right]$$

$$= H(x^\varepsilon(t), \alpha) + K'K\Delta|\alpha| < H_{K'K\Delta}(x^\varepsilon(t), \alpha)$$

and the analogous bound holds for the lim inf. In fact, it is clear that for A2 the Lipschitz condition is not needed and we can assume simply that $b(x, \xi)$ is continuous in x (uniformly in ξ).

4.2. Let μ_x be a family of probability measures on $\mathbb{R}^d$ that are parametrized by $x \in \mathbb{R}^d$ and let $b_n(x)$ denote a sequence of independent random vector fields satisfying $P\{b_n(x) \in A\} = \mu_x(A)$. Define a dynamical system X_n^ε by setting

$$X_{n+1}^\varepsilon = X_n^\varepsilon + b_n(X_n), \qquad n \geq 0,\ X_0^\varepsilon = x.$$

This is the situation considered in Azencott and Ruget [2]. Actually [2] considers a more general setup in which the state space of the process X_n^ε is allowed to be a connected Riemannian manifold. However, this can be reduced to the case considered here via "localization" arguments (such as those used in [1]) and the "contraction principle" [6, Theorem 3.1]. Define

$$H(x, \alpha) = \log \int_{\mathbb{R}^d} \exp\langle \alpha, \xi\rangle \mu_x(d\xi),$$

$L(x, \beta) = (H(x, \alpha))^*$. An essential hypothesis used in [2] is that the "level sets"

$$\{\beta\colon L(x, \beta) \leq l\}$$

vary in a Lipschitz fashion in the Hausdorff metric. This was in turn shown to follow from the existence of $K' < \infty$ such that

$$|H(x, \alpha) - H(y, \alpha)| \leq K'|x - y|\,|\alpha| \tag{4.2}$$

[2, Lemma 3.7] (at least for the case considered there, where the μ_x were assumed to be supported in a fixed compact set that is independent of x, for x in any compact set). In addition to this assumption, a uniform nondegeneracy condition on

$$S(x) = \{\beta\colon L(x, \beta) < \infty\}$$

was also required. This latter assumption is not required in our approach.

To see that these assumptions are a special case of A1 and A3, we first note that the compact support requirement for the μ_x measures gives the Lipschitz constant K required of $x^\varepsilon(\cdot)$. Take $\mathscr{F}^\varepsilon(t) = \sigma(X_i^\varepsilon, 0 < i \leq [t/\varepsilon] + 1)$. Then [with $N = t/\varepsilon$ and $N' = (t + \Delta)/\varepsilon$]

$$\begin{aligned}
&\frac{\varepsilon}{\Delta}\log E\left[\exp\left\langle \alpha, \sum_{i=N}^{N'} b_i(X_i^\varepsilon)\right\rangle | \mathscr{F}^\varepsilon(t)\right] \\
&\quad = \frac{\varepsilon}{\Delta}\log E\left[\exp\left\langle \alpha, \sum_{i=1}^{N'-1} b_i(X_i^\varepsilon)\right\rangle E\left[\exp\langle \alpha, b_{N'}(X_{N'}^\varepsilon)\rangle | X_{N'}^\varepsilon\right] | \mathscr{F}^\varepsilon(t)\right] \\
&\quad \le \frac{\varepsilon}{\Delta}\left(\log E\left[\exp\left\langle \alpha, \sum_{i=1}^{N'-1} b_i(X_i^\varepsilon)\right\rangle | \mathscr{F}^\varepsilon(t)\right] + H(X_N^\varepsilon, \alpha) + K'K\Delta|\alpha|\right) \\
&\quad \le H(X_N^\varepsilon, \alpha) + K'K\Delta|\alpha|.
\end{aligned}$$

Hence

$$\limsup(\varepsilon/\Delta)\log E\left[\exp\left\langle \alpha, \sum_{i=t/\varepsilon}^{(t+\Delta)/\varepsilon} b_i(X_i^\varepsilon)\right\rangle | \mathscr{F}^\varepsilon(t)\right] < H_{K'K\Delta}(x^\varepsilon(t), \alpha).$$

An analogous lower bound holds for the lim inf.

It is not a simple matter to formulate general conditions on μ_x under which (4.2) holds and a good deal of effort is expended in [2] to show (4.2) for some interesting cases. The interested reader is referred to [2, Propositions 3.9 and 3.11] for examples.

We note that it is easier to prove the weaker A3 than (4.2). Again, for the upper bound, the Lipschitz conditions can be weakened to simple continuity.

4.3. Our next example is based on results presented in Iscoe, Ney and Nummelin [9]. The model we work with is the following:

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon b(X_n^\varepsilon, \xi_n), \qquad X_0^\varepsilon = x.$$

Dynamics. We assume that $b(x, \xi)$, $x \in \mathbb{R}^d$, $\xi \in E$ is bounded and uniformly (in ξ) continuous in x and that b is measurable in ξ.

Noise model. We will assume that given ξ_{n-1} and X_n^ε, the value of ξ_n is distributed according to a Markov transition kernel having a density with respect to a fixed measure λ that is independent of x:

$$P^{X_n^\varepsilon}(\xi, d\psi) = p^{X_n^\varepsilon}(\xi, \psi)\lambda(d\psi)$$

(so that if X_n^ε were fixed at x, then the noise process would be a Markov process and the state X_n^ε enters into this process only through the transition density).

In order to get a uniform L.D.P. for functionals of the noise ξ_n we will need to assume that, given a compact set F_1, the transition density $p^x(\xi, \psi)$, or the transition density for some fixed finite number of iterations of $P^x(\xi, d\psi)$, is uniformly (in $\xi \in E$ and $x \in F_1$) bounded above and below away from zero. It is proved in [9, Lemma 3.2] that under these assumptions the following is true: If ξ_n^x is defined as the noise process was but for fixed x, then there exists $H(x, \alpha)$

such that (uniformly in $\xi_0^x = \xi_0$)

$$\lim \frac{1}{N} \log E_{\xi_0} \exp\left\langle \alpha, \sum_{i=1}^{N} b(x, \xi_i^x) \right\rangle = H(x, \alpha). \tag{4.3}$$

It is also proved [9, Lemma 3.4] that $H(x, \alpha)$ is convex and smooth (analytic) in α. Define $\mathscr{F}^\varepsilon(t) = \sigma((X_i^\varepsilon, \xi_i), 0 \le i \le [t/\varepsilon] + 1)$.

LEMMA 4.1. *Consider the dynamics and noise model described previously. Assume given compact $F_1 \subset \mathbb{R}^d$ that:*

(i) *There are $0 < a \le A < \infty$ such that for all $x \in F_1$, $\xi, \psi \in E$, $a \le p^x(\xi, \psi) \le A$ (or that this is true for the transition density after some finite number of steps).*

(ii) *$p^x(\xi, \psi)$ is continuous in x, uniformly in $\xi, \psi \in E$, for $x \in F_1$.*

Then Assumption A2 *holds. If we strengthen the continuity in* (ii) *and $b(\cdot, \xi)$ to Lipschitz continuity, then* A3 *holds.*

PROOF. Let $\gamma > 0$ be given. Using the continuity condition and the lower bound on $p^x(\xi, \psi)$, there is $\delta > 0$ such that for $|x - y| \le \delta$,

$$\begin{aligned} p^y(\xi, \psi) &= p^x(\xi, \psi) + (p^y(\xi, \psi) - p^x(\xi, \psi)) \\ &\le p^x(\xi, \psi)\left[1 + \frac{|p^y(\xi, \psi) - p^x(\xi, \psi)|}{a}\right] \\ &\le p^x(\xi, \psi)\exp(p^y(\xi, \psi) - p^x(\xi, \psi))/a \le p^x(\xi, \psi)\exp \gamma, \end{aligned}$$

and similarly

$$p^x(\xi, \psi) \le p^y(\xi, \psi)\exp \gamma.$$

It follows from the uniform continuity of $b(x, \xi)$ in x for all ξ that for small $\delta > 0$ we can obtain

$$E_{\xi_0}\left[\exp\langle \alpha, b(y, \xi_1^y)\rangle\right] \le E_{\xi_0}\left[\exp\langle \alpha, b(x, \xi_1^x)\rangle\right]\exp \gamma(1 + |\alpha|)$$

uniformly in ξ_0, when $|x - y| \le \delta$. The proof is completed by iterating backwards (as was done in Section 4.2) and using (4.3). The proof of the second assertion is completely analogous. □

4.4. The results presented in [9] actually cover a much broader range of processes than the relatively simple example given in Section 4.3. Our present example will make greater use of the generality of their results, without in any way being complete in this respect.

We consider an extension of the previous example to the case where $b(x, \xi)$ is no longer deterministic but is instead an i.i.d. sequence of a random vector fields whose distribution depends on the pair (x, ξ). We therefore adopt the notation $b_n(x, \xi)$. For simplicity we shall assume that the distribution of $b_n(x, \xi)$ is concentrated on a finite set of points a_i, $1 \le i \le N$, and that the a_i depend only on x. This can be viewed as the Markov version of the example of Section 4.2. We have

$$P\{b_n(x,\xi) \in A\} = \sum_{i=1}^{N} p_i(x,\xi)\delta_{a_i(x)}(A)$$

for some measurable functions $p_i(x,\xi)$ satisfying

$$\sum_{i=1}^{N} p_i(x,\xi) = 1, \qquad 0 \le p_i(x,\xi) \le 1, 1 \le i \le N.$$

(In order to correctly define the process, what really needs to be done is to prove the existence of the sequence of independent random vector fields $\{b_n\}$ described previously. We shall, however, ignore the details on this point, since it is essentially the same situation as encountered in [2, Section 5].)

LEMMA 4.2. *Consider the model just described. Make the same assumptions on the transition kernel as in Lemma* 4.1. *Assume in addition that*:

(i) *The functions $a_i(x)$, $1 \le i \le N$, are continuous.*
(ii) *The functions $p_i(x,\xi)$, $1 \le i \le N$, are continuous in x (uniformly in ξ) and either bounded from below by $c > 0$ or identically zero.*

Then Assumption A2 *holds. If we strengthen the continuity in* (i), (ii) *and $b(\cdot,\xi)$ to Lipschitz continuity, then* A3 *holds.*

PROOF. The results in [9] imply the obvious analog of (4.3):

$$\lim \frac{1}{N} \log E_{\xi_0} \exp\left\langle \alpha, \sum_{j=1}^{N} b_j(x, \xi_j^x) \right\rangle = H(x,\alpha).$$

It follows from the assumptions (i) and (ii) that given $\gamma > 0$, there is $\delta > 0$ such that if $|x - y| \le \delta$, then

$$\begin{aligned} &\sum_{i=1}^{N} (\exp\langle \alpha, a_i(x)\rangle) p_i(x,\xi) \\ &\quad \le \sum_{i=1}^{N} (\exp\langle \alpha, a_i(y)\rangle) p_i(y,\xi) \exp \gamma(1 + |\alpha|) \end{aligned} \tag{4.4}$$

uniformly in ξ. In the Lipschitz case there is $\overline{K}$ such that we may take $\gamma = \overline{K}\delta$. The proof is now the same as that of Lemma 4.1. □

4.5. *Continuous time.* For examples of continuous time processes satisfying the conditions of A1, A2 and A3 we refer to the continuous time examples presented in [5]. For an example involving state dependent noise, one can consider the continuous time version of the process considered in Section 4.3. The proof of an analog of Lemma 4.1 is essentially the same, save that one must quote the continuous time results in [9].

5. Projection algorithms. Oftentimes recursive algorithms must be chosen that will achieve a goal (such as convergence to an optimum point of some sort) subject to constraints, such as requiring some kind of feasibility at each step. In this section we consider a method that is commonly used in such situations, namely projection algorithms. With projection algorithms, the state is mapped after each iteration onto the point in a feasible set G that is closest. We consider this problem in the context of the model of Section 4.4.

The approach we take to proving a L.D.P. for such algorithms is to adapt the technique used by Anderson and Orey [1] for reflected diffusions. For simplicity we confine ourselves to the simplest possible case, where $G = \{x \in \mathbb{R}: x \geq 0\}$. If G is the closure of an open connected subset of $\mathbb{R}^d$ with sufficiently smooth boundary, then the general result follows from this special case [1, Sections 1.3 and 1.4].

Let $\pi(x) = 0 \vee x$. We define the projected algorithm as

$$X_{n+1}^{\varepsilon} = \pi\big(X_n^{\varepsilon} + \varepsilon b_n(X_n^{\varepsilon}, \xi_n)\big), \qquad X_0^{\varepsilon} = x \geq 0. \tag{5.1}$$

Define $x^{\varepsilon}(\cdot)$ as the usual piecewise linearly interpolated version.

THEOREM 5.1. *Make the assumptions of Lemma* 4.2 *and define*

$$\bar{L}(x,\beta) = \begin{cases} L(x,\beta), & \text{if } x > 0 \text{ or } x = 0,\ \beta > 0, \\ \inf\limits_{\beta \leq 0} L(x,\beta), & \text{if } x = 0,\ \beta = 0, \\ \infty, & \text{else}. \end{cases}$$

Define

$$\bar{S}_x(\phi) = \int_0^1 \bar{L}(\phi, \dot{\phi})\, ds$$

if $\phi(0) = x \geq 0$ *and* ϕ *is absolutely continuous, and set* $\bar{S}_x(\phi) = \infty$ *otherwise. Then* x^{ε} *as defined through* (5.1) *satisfies a L.D.P.* (*in the sense of Theorems* 2.1 *and* 3.2) *with functional* $\bar{S}_x(\phi)$.

PROOF. Consider the unrestricted version of X_n^{ε}:

$$\tilde{X}_{n+1}^{\varepsilon} = \tilde{X}_n^{\varepsilon} + \varepsilon b_n(X_n^{\varepsilon}, \xi_n), \qquad \tilde{X}_0^{\varepsilon} = x.$$

Define $\tilde{x}^{\varepsilon}(\cdot)$ as usual and define $\bar{x}^{\varepsilon}(\cdot)$ to be the *piecewise constant* version of $\tilde{X}_n^{\varepsilon}$ having interpolation interval ε:

$$\bar{x}^{\varepsilon}(t) = X_n^{\varepsilon} \quad \text{for } t \in [n\varepsilon, n\varepsilon + \varepsilon).$$

We define a mapping Γ on the set of paths (continuous or not) sending $[0,1]$ into $\mathbb{R}$. Γ maps ψ to ϕ if

$$\phi(t) = \psi(t) - \Big(\inf_{0 \leq s \leq t} \psi(s) \wedge 0\Big). \tag{5.2}$$

It is easy to verify that $X_n^{\varepsilon} = \Gamma(\bar{x}^{\varepsilon})(n\varepsilon)$, which together with the definition of Γ and the fact that $\bar{x}^{\varepsilon}(n\varepsilon) = \tilde{x}^{\varepsilon}(n\varepsilon)$ imply that $X_n^{\varepsilon} = \Gamma(\tilde{x}^{\varepsilon})(n\varepsilon)$. We may therefore write

$$\tilde{X}_{n+1}^{\varepsilon} = \tilde{X}_n^{\varepsilon} + \varepsilon b_n\big(\Gamma(\tilde{x}^{\varepsilon}(n\varepsilon)), \xi_n\big), \qquad \tilde{X}_0^{\varepsilon} = x.$$

A simple calculation shows

$$\sup_{0\le s\le t} |\Gamma(\psi_1)(s) - \Gamma(\psi_2)(s)| \le 2 \sup_{0\le s\le t} |\psi_1(s) - \psi_2(s)|. \tag{5.3}$$

It is now quite easy to adapt the proofs of Theorems 2.1 and 3.2 to yield a L.D.P. for $\tilde{x}^\varepsilon$, with functional

$$S_x(\phi) = \int_0^1 L\big(\Gamma(\phi)(s), \dot{\phi}(s)\big)\, ds.$$

The important fact needed is that the inequality (5.3) ensures that the iterates $X_n^\varepsilon = \Gamma(\tilde{x}^\varepsilon)(n\varepsilon)$ still vary in a sufficiently slow way, i.e., the interpolated paths are still Lipschitz continuous.

Because the restriction of the map Γ to $C[0,1]$ is continuous (with respect to the sup norm) we can employ the contraction principle [6, Theorem 3.1] to deduce that $\Gamma(\tilde{x}^\varepsilon)(\cdot)$ satisfies a L.D.P. with functional

$$S_x^*(\phi) = \inf S_x(\psi),$$

where the inf is over $\psi \in C_x[0,1]$ satisfying $\phi = \Gamma(\psi)$. The infimum over the empty set is defined as $+\infty$. Note that $S_x(\phi) < \infty$ implies $\phi(t) \ge 0$ on $[0,1]$. For absolutely continuous ϕ and ψ we have $\phi = \Gamma(\psi)$ if and only if $\phi(t) \ge 0$ and

$$\dot{\psi}(t) = \dot{\phi}(t) + \eta(t) I_{\{\phi(t)=0\}} \quad \text{a.s.},$$

where $\eta(t)$ is any nonpositive, measurable function [3, Lemma 4.7]. Therefore $S_x^*(\phi) = \bar{S}_x(\phi)$.

To finish the proof we need to take care of the fact that $\Gamma(\tilde{x}^\varepsilon)$ is not necessarily the same as x^ε between the interpolation points $n\varepsilon$. But this does not actually pose a problem since $\Gamma(\tilde{x}^\varepsilon)(n\varepsilon) = x^\varepsilon(n\varepsilon)$ implies

$$\sup_{0\le s\le t} |\Gamma(\tilde{x}^\varepsilon)(s) - x^\varepsilon(s)| \le K\varepsilon,$$

and therefore x^ε satisfies a L.D.P. with the same functional as $\Gamma(\tilde{x}^\varepsilon)$ (this latter assertion follows from the equivalent formulation of a L.D.P. mentioned in the first remark in the Introduction). □

6. Example: Application to a routing problem. We apply the results to study an automatic routing problem considered in [10]. The description of our system is as follows. Calls arrive at a switching station at random at the discrete times $n \in Z^+$. We have

$$P\{\text{one call arrives at time } n\} = \mu, \qquad \mu \in (0,1),$$
$$P\{\text{two or more calls arrive at time } n\} = 0.$$

From the station, there are two possible routings to the destination; the ith route has N_i lines and can thus handle N_i calls simultaneously.

Let $[n, n+1)$ denote the nth interval of time. We assume that the duration of each call is random with a geometric distribution:

$$P\{\text{call is completed in } (n+1)\text{st interval} \mid \text{uncompleted at end of } n\text{th interval}\} = \lambda_i, \qquad \lambda_i \in (0,1).$$

We shall assume that the double sequence of call durations and interarrival times are mutually independent and to obtain an unambiguous formulation we shall assume that calls terminating in the nth interval actually terminate at time $n + \frac{1}{2}$ and that calls come in and are assigned to routes at times $n \in Z^+$. Define ξ_n^i equal to the number of lines of route i in use at time n (this includes those that have just been assigned).

We now describe the routing mechanism. We shall define a recursive algorithm whose state will control how assignments are made. If the state takes the value $x \in [0,1]$ at time n, then a random variable η_n with distribution

$$\eta_n = \begin{cases} 1, & \text{w.p. } x, \\ 2, & \text{w.p. } 1 - x \end{cases}$$

is generated.

If a call arrives at time n, it is assigned to route η_n. If all lines are full, it is reassigned to the alternate route, unless those lines are all full, in which case the call is rejected and disappears. Let J_n^i be the indicator of the event {call is first assigned to route i at time n and is accepted by route i}. We may then update the state by calculating

$$x + \varepsilon(1 - x)J_n^1 - \varepsilon x J_n^2$$

and projecting this value onto some interval $[l, u]$, $0 < l < u < 1$.

The transition probabilities of the route occupancy process $\{\xi_n^i\}$ are thus given by (for $i = 1$)

$$P\{\xi_n^1 = \xi' | \xi_{n-1}^1 = \xi, \text{state} = x\}$$
$$= \begin{cases} \mu x(1 - \lambda_1), & \text{if } \xi' = \xi + 1,\ \xi < N_1, \\ 0, & \text{if } \xi' = \xi + 1,\ \xi = N_1, \\ (1-\mu)(1-\lambda_1) + \mu x \lambda_1 + \mu(1-x)(1-\lambda_1), & \text{if } \xi' = \xi, 0 < \xi < N_1, \\ (1-\mu)(1-\lambda_1) + \mu x + \mu(1-x)(1-\lambda_1), & \text{if } \xi' = \xi,\ \xi = N_1, \\ (1-\mu) + \mu(1-x), & \text{if } \xi' = \xi,\ \xi = 0, \\ (1-\mu)\lambda_1 + \mu(1-x)\lambda_1, & \text{if } \xi' = \xi - 1, 0 < \xi, \\ 0, & \text{if } \xi' = \xi - 1,\ \xi = 0, \\ 0, & \text{else.} \end{cases}$$

It follows that if we define $R_n = (1-x)J_n^1 - xJ_n^2$, then for *fixed* x, $\{(\xi_n^1, \xi_n^2), R_n\}$ is a process satisfying the conditions of Section 4.4. We have

$$P\{R_n = r | (\xi_n^1, \xi_n^2) = (\xi^1, \xi^2), x\}$$
$$= \begin{cases} \mu x, & \text{if } r = (1-x) \text{ and } \xi^1 < N_1, \\ 0, & \text{if } r = (1-x) \text{ and } \xi^1 = N_1, \\ \mu(1-x), & \text{if } r = -x \text{ and } \xi^2 < N_2, \\ 0, & \text{if } r = -x \text{ and } \xi^2 = N_2 \end{cases}$$

and $P\{R_n = 0|(\xi_n^1, \xi_n^2), x\}$ is defined so that the conditional probabilities sum to 1.

The algorithm may then be defined by

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon R_n|_l^u, \qquad R_n = (1 - X_n^\varepsilon)J_n^1 - X_n^\varepsilon J_n^2,$$

where $|_l^u$ denotes the projection.

In this case the action functional is given by $\bar{S}_x(\phi) = \int_0^1 \bar{L}(\phi, \dot{\phi})\,dt$, where $\bar{L}$ is defined in terms of $L = H^*$:

$$\bar{L}(x, \beta) = \begin{cases} L(x, \beta), & \text{if } l < x < u \text{ or } x = l, \beta > 0 \\ & \text{or } x = u, \beta < 0, \\ \inf_{\beta \le 0} L(l, \beta), & \text{if } x = l, \beta = 0, \\ \inf_{\beta \ge 0} L(u, \beta), & \text{if } x = u, \beta = 0, \\ \infty, & \text{else.} \end{cases}$$

REMARKS. If the system just described is replaced by the analogous system with *no* projection (in which case the state will still remain in $[0, 1]$), then Assumption A3 is no longer valid. It is easy to show that

$$S(x) = \begin{cases} [-x, 1 - x], & x \in (0, 1), \\ \{0\}, & x = 0, 1. \end{cases}$$

Since $S(x)$ is not Lipschitz, A3 cannot hold (see the remark after A3).

Although we have presented the simplest possible problem of this type, it is clear that this model can be generalized considerably, for example, in the number of lines and switching nodes or in the input and call length statistics, and still admit an analysis of the type given.

APPENDIX

In this section we concern ourselves with referencing and proving the lower semicontinuity results that were needed in the paper. All results that follow are under assumptions A1 and A2 which imply that $H(x, \alpha)$ is convex and continuous in α (for fixed x), u.s.c. in x (for fixed α) and that $|H(x, \alpha)| \le K|\alpha|$.

LEMMA A.1. *Let* $L(x, \beta) = (H(x, \alpha))^*$ *and define* $S_x(\phi)$ *and* $\Phi_x(s)$ *as in Theorem* 2.1. *Let* ψ *mapping* $[0, 1]$ *into* $\tilde{F}$ *be measurable. Then*:

(i) $L(x, \beta)$ *is l.s.c. in* (x, β).
(ii) *Both* $S_{\phi(0)}(\phi)$ *and* $\int_0^1 L(\psi, \dot{\phi})\,ds$ *are l.s.c. in* ϕ.
(iii) $\Phi_x(s)$ *is compact in* $C_x[0, 1]$ *for any* $s < \infty$.

PROOF. (i) Define $\alpha(x, \beta, n)$ such that $\langle \alpha(x, \beta, n), \beta \rangle - H(x, \alpha(x, \beta, n)) > L(x, \beta) - 1/n$. By the continuity of the inner product and the u.s.c. of $H(x, \alpha)$, there is a neighborhood N of (x, β) such that for all $(x', \beta') \in N$,

$$\langle \alpha(x, \beta, n), \beta' \rangle - H(x', \alpha(x, \beta, n)) > L(x, \beta) - 2/n$$

and hence

$$L(x', \beta') > L(x, \beta) - 2/n,$$

and therefore $L(\cdot, \cdot)$ is jointly l.s.c.

(ii) [8, Theorem 3, Section 9.1.4].

(iii) Since $S_x(\phi) < \infty$ implies $|\dot{\phi}| \leq K$ a.s., compactness follows from Ascoli's theorem and (ii). □

Acknowledgment. The author would like to thank Professor Harold Kushner for both suggesting and discussing the problems considered in this paper.

REFERENCES

[1] Anderson, R. and Orey, S. (1976). Small random perturbations of dynamical systems with reflecting boundary. *Nagoya Math. J.* **60** 189–216.

[2] Azencott, R. and Ruget, G. (1977). Melanges d'equations differentielles et grands ̀ecarts a la loi des grands nombres. *Z. Wahrsch. verw. Gebiete* **38** 1–54.

[3] Dupuis, P. (1987). Large deviations analysis of reflected diffusions and constrained stochastic approximation algorithms in convex sets. *Stochastics* **21** 63–96.

[4] Dupuis, P. and Kushner, H. J. (1985). Stochastic approximations via large deviations: Asymptotic properties. *SIAM J. Control Optim.* **23** 675–696.

[5] Freidlin, M. I. (1978). The averaging principle and theorems on large deviations. *Russian Math. Surveys* **33** 117–176.

[6] Freidlin, M. I. and Wentzell, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, Berlin.

[7] Gärtner, J. (1977). On large deviations from the invariant measure. *Theory Probab. Appl.* **22** 24–39.

[8] Ioffe, A. D. and Tikhomirov, V. M. (1979). *Theory of Extremal Problems.* North-Holland, Amsterdam

[9] Iscoe, I., Ney, P. and Nummelin, E. (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math.* **6** 373–412.

[10] Kushner, H. J. and Shwartz, A. (1981). An invariant measure approach to the convergence of stochastic approximations with state dependent noise. *SIAM J. Control Optim.* **19** 87–105.

[11] Rockafellar, R. T. (1970). *Convex Analysis.* Princeton Univ. Press, Princeton, N.J.

[12] Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

Department of Mathematics and Statistics
Lederle Graduate Research Center Towers
University of Massachusetts
Amherst, Massachusetts 01003