Algorithmica (1991) 6: 419–436

Algorithmica

Simulated Annealing Type Algorithms for Multivariate Optimization[1]

Saul B. Gelfand[2] and Sanjoy K. Mitter[3]

Abstract. We study the convergence of a class of discrete-time continuous-state simulated annealing type algorithms for multivariate optimization. The general algorithm that we consider is of the form $X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k$. Here $U(\cdot)$ is a smooth function on a compact subset of $\mathbb{R}^d$, $\{\xi_k\}$ is a sequence of $\mathbb{R}^d$-valued random variables, $\{W_k\}$ is a sequence of independent standard d-dimensional Gaussian random variables, and $\{a_k\}$, $\{b_k\}$ are sequences of positive numbers which tend to zero. These algorithms arise by adding decreasing white Gaussian noise to gradient descent, random search, and stochastic approximation algorithms. We show under suitable conditions on $U(\cdot)$, $\{\xi_k\}$, $\{a_k\}$, and $\{b_k\}$ that X_k converges in probability to the set of global minima of $U(\cdot)$. A careful treatment of how X_k is restricted to a compact set and its effect on convergence is given.

Key Words. Simulated annealing, Random search, Stochastic approximation.

1. Introduction. It is desired to select a parameter value x^* which minimizes a smooth function $U(x)$ over $x \in D$, where D is a compact subset of $\mathbb{R}^d$. The stochastic descent algorithm

$$Z_{k+1} = Z_k - a_k(\nabla U(Z_k) + \xi_k) \tag{1.1}$$

is often used where $\{\xi_k\}$ is a sequence of $\mathbb{R}^d$-valued random variables and $\{a_k\}$ is a sequence of positive numbers with $a_k \to 0$ and $\sum a_k = \infty$. An algorithm of this type might arise in several ways. The sequence $\{Z_k\}$ could correspond to a stochastic approximation [1], where the sequence $\{\xi_k\}$ arises from noisy or imprecise measurements of $\nabla U(\cdot)$ or $U(\cdot)$. The sequence $\{Z_k\}$ could also correspond to a random search [2], where the sequence $\{\xi_k\}$ arises from randomly selected search directions. Now since D is compact it is necessary to ensure the trajectories of $\{Z_k\}$ are bounded; this may be done either by projecting Z_k back into D if it ever leaves D, or by fixing the dynamics in (1.1) so that Z_k never leaves D or only leaves D finitely many times with probability 1 (w.p.1). Let S be the set of local minima of $U(\cdot)$ and let S^* be the set of global minima of $U(\cdot)$. Under suitable conditions on $U(\cdot)$, $\{\xi_k\}$, and $\{a_k\}$, and assuming that $\{Z_k\}$ is bounded, it is well known that $Z_k \to S$ as $k \to \infty$ w.p.1. In particular, if $U(\cdot)$ is well behaved, $a_k = A/k$ for k large,

[1] Research supported by the Air Force Office of Scientific Research contract 89-0276B, and the Army Research Office contract DAAL03-86-K-0171 (Center for Intelligent Control Systems).
[2] School of Electrical Engineering, Purdue University, West Lafayette, IN 47907, USA.
[3] Department of Electrical Engineering and Computer Science, and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Received November 26, 1988; revised December 6, 1989. Communicated by Alberto Sangiovanni-Vincentelli.

$\{\xi_k\}$ are independent with $E\{|\xi_k|^2\} = O(a_k^\alpha)$, and $|E\{\xi_k\}| = O(a_k^\beta)$ where $\alpha > -1$, $\beta > 0$, and $\{Z_k\}$ is bounded by a suitable device, then $Z_k \to S$ as $k \to \infty$ w.p.1. However, if $U(\cdot)$ has a strictly local minima, then in general $Z_k \not\to S^*$ as $k \to \infty$ w.p.1.

The analysis of the convergence w.p.1 of $\{Z_k\}$ is usually based on the convergence of an *associated ordinary differential equation* (ODE)

$$\dot{z}(t) = -\nabla U(z(t)).$$

This approach was pioneered by Ljung [3] and further developed by Kushner and Clark [4], Metivier and Priouret [5], and others. Kushner and Clark also analyzed the convergence in probability of $\{Z_k\}$ by this method. However, although their theory yields much useful information about the asymptotic behavior of $\{Z_k\}$ under very weak assumptions, it fails to obtain $Z_k \to S^*$ as $k \to \infty$ in probability unless $S = S^*$ is a singleton: see p. 125 of [4].

Consider a modified stochastic descent algorithm

$$X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k, \tag{1.2}$$

where $\{W_k\}$ is a sequence of independent Gaussian random variables with zero-mean and identity covariance matrix, and $\{b_k\}$ is a sequence of positive numbers with $b_k \to 0$. The $b_k W_k$ term is added in artificially by Monte Carlo simulation so that $\{X_k\}$ can avoid getting trapped in a strictly local minimum of $U(\cdot)$. In general, $X_k \not\to S^*$ as $k \to \infty$ w.p.1 (for the same reasons that $Z_k \not\to S^*$ as $k \to \infty$ w.p.1). However, under suitable conditions on $U(\cdot)$, $\{\xi_k\}$, $\{a_k\}$, and $\{b_k\}$, and assuming that $\{X_k\}$ is bounded, we shall show that $X_k \to S^*$ as $k \to \infty$ in probability. In particular, if $U(\cdot)$ is well behaved, $a_k = A/k$ and $b_k^2 = B/k \log \log k$ for k large where $B/A > C_0$ (a constant which depends on $U(\cdot)$), $\{\xi_k\}$ are independent with $E\{|\xi_k|^2\} = O(a_k^\alpha)$ and $|E\{\xi_k\}| = O(a_k^\beta)$ where $\alpha > -1$, $\beta > 0$, and $\{X_k\}$ is bounded by a suitable device, then $X_k \to S^*$ as $k \to \infty$ in probability. We actually require a weaker condition than the independence of the $\{\xi_k\}$; see Section 2.

Our analysis of the convergence in probability of $\{X_k\}$ is based on the convergence of what we call the *associated stochastic differential equation* (SDE)

$$dx(t) = -\nabla U(x(t))\, dt + c(t)\, dw(t), \tag{1.3}$$

where $w(\cdot)$ is a standard d-dimensional Wiener process and $c(\cdot)$ is a positive function with $c(t) \to 0$ as $t \to \infty$ (take $t_k = \sum_{n=0}^{k-1} a_n$ and $b_k = \sqrt{a_k}c(t_k)$ to see the relationship between (1.2) and (1.3)). The simulation of the Markov diffusion $x(\cdot)$ for the purpose of global optimization has been called continuous simulated annealing. In this context, $U(x)$ is called the energy of state x and $T(t) = c^2(t)/2$ is called the temperature at time t. This method was first suggested by Grenender [6] and Geman and Hwang [7] for image processing applications with continuous grey levels. We remark that the discrete simulated annealing algorithm for combinatorial optimization based on simulating a Metropolis-type Markov chain

[8], and the continuous simulated annealing algorithm for multivariate optimization based on simulating the Langevin-type Markov diffusion discussed above both have a Gibbs invariant distribution $\propto \exp(-U(x)/T)$ when the temperature is fixed at T. The invariant distributions concentrate on the global minima of $U(\cdot)$ as $T \to 0$. The idea behind simulated annealing is that if $T(t)$ decreases slowly enough, then the distribution of the annealing process remains close to the Gibbs distribution $\propto \exp(-U(x)/T(t))$ and hence also concentrates on the global minima of $U(\cdot)$ as $t \to \infty$ and $T(t) \to 0$. Now the asymptotic behavior of $x(\cdot)$ has been studied intensively by a number of researchers [7], [10]–[12]. Our work is based on the analysis of $x(\cdot)$ developed by Chiang *et al.* [11] who prove the following result: if $U(\cdot)$ is well behaved and $c^2(t) = C/\log t$ for t large where $C > C_0$ (the same C_0 as above), then $x(t) \to S^*$ as $t \to \infty$ in probability.

The actual implementation of (1.3) on a digital computer requires some type of discretization or numerical integration, such as (1.2). Aluffi-Pentini *et al.* [13] describe some numerical experiments performed with (1.2) for a variety of test problems. Kushner [12] was the first to analyze (1.2) but for the case of $a_k = b_k = A/\log k$, k large. Our work differs from [12] in the following ways. First, we give a careful treatment of how the trajectories of $\{X_k\}$ are bounded and its effect on the convergence analysis. Although bounded trajectories are assumed in [12], a thorough discussion is not included there. Second, although a detailed asymptotic description of X_k as $k \to \infty$ is obtained in [12], in general, $X_k \not\to S^*$ as $k \to \infty$ in probability unless $\xi_k = 0$. The reason for this is intuitively clear: even if $\{\xi_k\}$ is bounded, $a_k \xi_k$ and $a_k W_k$ can be of the same order and hence can interfere with each other. On the other hand, we get $X_k \to S^*$ as $k \to \infty$ in probability for $\xi_k \neq 0$ and in fact for ξ_k with $E\{|\xi_k|^2\} = O(k^\gamma)$ and $\gamma < 1$. This has practical implications when $\nabla U(\cdot)$ is not measured exactly: we give a simple example. Finally, our method of analysis is different from [12] in that we obtain the asymptotic behavior of X_k as $k \to \infty$ from the corresponding behavior of $x(t)$ as $t \to \infty$.

2. Main Results and Discussion. In the following $|\cdot|$ and $\langle \cdot, \cdot \rangle$ are the Euclidean norm and inner product, respectively. $\|\cdot\|$ is the supremum norm.

Our analysis, like Kushner's [12], requires that we bound the trajectories of $\{X_k\}$. We proceed as follows. Take D to be a closed ball in $\mathbb{R}^d$, say $D = \{x : |x| \le r\}$. We modify (1.2), (1.3) in a thin annulus $\{x : r_0 \le |x| \le r\}$ and make suitable assumptions to ensure that $\{X_k\}$ and $x(\cdot)$ remain in D. The actual algorithm is

$$\begin{aligned} \tilde{X}_{k+1} &= X_k - a_k(\nabla U(X_k) + \xi_k) + b_k \sigma(X_k) W_k, \\ X_{k+1} &= \tilde{X}_{k+1} 1_D(\tilde{X}_{k+1}) + X_k 1_{D^c}(\tilde{X}_{k+1}), \end{aligned} \tag{2.1}$$

and the associated SDE is

$$dx(t) = -\nabla U(x(t))\, dt + c(t)\sigma(x(t))\, dw(t). \tag{2.2}$$

In what follows we make the following assumptions. Let $0 < r_0 < r_1 < r$ (typically $r - r_0 \ll 1$).

(A1) $U(\cdot)$ is a twice continuously differentiable function from D to $[0, \infty)$ with $\min U(x) = 0$ and $\langle \nabla U(x), x \rangle > 0$ for $|x| \geq r_0$.
(A2) $\sigma(\cdot)$ is a Lipshitz continuous function from D to $[0, 1]$ with $\sigma(x) = 1$ for $|x| \leq r_1$, $\sigma(x) \in (0, 1]$ for $r_1 \leq |x| < r$, and $\sigma(x) = 0$ for $|x| = r$.
(A3) $a_k = A/k$, $b_k^2 = B/k \log\log k$, k large, where $A, B > 0$.
(A4) $c^2(t) = C/\log t$, t large, where $C > 0$.

For $k = 0, 1, \ldots$ let $\mathscr{F}_k$ be the σ-field generated by

$$\{X_0, \xi_0, \ldots, \xi_{k-1}, W_0, \ldots, W_{k-1}\}.$$

(A5) $E\{|\xi_k|^2 \,|\mathscr{F}_k\} = O(a_k^\alpha)$, $E\{\xi_k|\mathscr{F}_k\} = O(a_k^\beta)$ as $k \to \infty$ uniformly w.p.1; $\xi_k = 0$ when $|X_k| \geq r_1$ w.p.1; W_k is independent of $\mathscr{F}_k$ for all k.

For $\varepsilon > 0$ let

$$d\pi^\varepsilon(x) = \frac{1}{Z^\varepsilon} \exp\left(-\frac{2U(x)}{\varepsilon^2}\right) dx, \qquad Z^\varepsilon = \int_D \exp\left(-\frac{2U(x)}{\varepsilon^2}\right) dx.$$

(A6) π^ε has a unique weak limit π as $\varepsilon \to 0$.

We remark that π concentrates on S^*, the global minima of $U(\cdot)$. The existence of π and a simple characterization in terms of the Hessian of $U(\cdot)$ is discussed in [14]. We also remark that under the above assumptions, it is clear that $x(t)$ always stays in D, and it can be shown (see the remark following Proposition 1) that $\tilde{X}_k$ eventually stays in D.

For a process $u(\cdot)$ and function $f(\cdot)$, let $E_{t_1,u_1}\{f(u(t))\}$ denote conditional expectation given $u(t_1) = u_1$ and let $E_{t_1,u_1;t_2,u_2}\{f(u(t))\}$ denote conditional expectation given $u(t_1) = u_1$ and $u(t_2) = u_2$ (more precisely, these are suitable fixed versions of conditional expectations). Also for a measure $\mu(\cdot)$ and a function $f(\cdot)$ let $\mu(f) = \int f \, d\mu$.

By a modification of the main result of [11] there exists constants C_0, C_1 such that for $C_0 < C < C_1$ and any continuous function $f(\cdot)$ on D

$$\lim_{t \to \infty} E_{0,x}\{f(x(t))\} = \pi(f) \tag{2.3}$$

uniformly for $x \in D$ (this modification follows easily from Lemma 3 below). The modification is needed here because [11] deals with a nondegenerate diffusion ($\sigma(x) = 1$ for all x in (2.2)) while we are concerned with a degenerate diffusion ($\sigma(x) \to 0$ as $|x| \uparrow r$ in (2.2)). The constant C_0 depends only on $U(x)$ for $|x| < r_0$ and is defined in [11] in terms of the action functional for the dynamical system

$\dot{z}(t) = -\nabla U(z(t))$. The constant C_1 depends only on $U(x)$ for $|x| \geq r_0$ and is given by

$$C_1 = \tfrac{3}{2}\left(\inf_{|x|=r_1} U(x) - \sup_{|x|=r_0} U(x)\right).$$

In [11] only $C > C_0$ and not $C < C_1$ is required; however, $U(x)$ and $\nabla U(x)$ must satisfy certain growth conditions as $|x| \to \infty$. Note that a penalty function can be added to $U(\cdot)$ so that C_1 is as large as desired. Here is our theorem on the convergence of $\{X_k\}$.

THEOREM. *Let $\alpha > -1$, $\beta > 0$, and $C_0 < B/A < C_1$. Then for any continuous function $f(\cdot)$ on D*

$$\lim_{k\to\infty} E_{0,x}\{f(X_k)\} = \pi(f) \tag{2.4}$$

uniformly for $x \in D$.

Since π concentrates on S^*, (2.3) and (2.4) imply $x(t) \to S^*$ as $t \to \infty$ and $X_k \to S^*$ as $k \to \infty$ in probability, respectively.

The proof of the theorem requires the following three lemmas. Let $\{t_k\}$ and $\beta(\cdot)$ be defined by

$$t_k = \sum_{n=0}^{k-1} a_n, \qquad k = 0, 1, \ldots,$$

$$\int_s^{\beta(s)} \frac{\log s}{\log u}\, du = s^{2/3}, \qquad s > 1.$$

It is easy to check that $\beta(s)$ is well defined by this expression and in fact satisfies $s + s^{2/3} \leq \beta(s) \leq s + 2s^{2/3}$.

LEMMA 1. *Let $\alpha > -1$, $\beta > 0$, and $B/A = C$. Then there exists $\gamma > 1$ such that for any continuous function $f(\cdot)$ on D*

$$\lim_{n\to\infty} \sup_{k: t_n \leq t_k \leq \gamma t_n} (E_{0,x;n,y}\{f(X_k)\} - E_{t_n,y}\{f(x(t_k))\}) = 0$$

uniformly for $x, y \in D$.

LEMMA 2. *Let $T > 0$. Then for any continuous function $f(\cdot)$ on D*

$$\lim_{n\to\infty} \sup_{s: t_n \leq s \leq t_{n+1}} (E_{t_n,y}\{f(x(\beta(s+T)))\} - E_{s,y}\{f(x(\beta(s+T)))\}) = 0$$

uniformly for $y \in D$.

LEMMA 3. *Let $C_0 < C < C_1$. Then there exists $T > 0$ such that for any continuous function $f(\cdot)$ on D*

$$\lim_{s \to \infty} (E_{s,y}\{f(x(\beta(s+T)))\} - \pi^{c(s+T)}(f)) = 0$$

uniformly for $y \in D$.

The proofs of Lemmas 1 and 2 are given in Section 3 and Lemma 3 is proved in Section 4. Note that the lemmas are concerned with approximations on intervals of increasing length ($\gamma t_n - t_n \to \infty$ as $n \to \infty$, $\beta(s) - s \to \infty$ as $s \to \infty$). Lemma 3 is a modification of results obtain in [11] for a nondegenerate diffusion ($\sigma(x) = 1$ for all x in (2.2)).

We now show how the lemmas may be combined to prove the theorem.

PROOF OF THE THEOREM. Choose T as in Lemma 3. Note that $\beta(s)$ is a strictly increasing function and $s + s^{2/3} \le \beta(s) \le s + 2s^{2/3}$ for s large enough. Hence for k large enough we can choose s such that $t_k = \beta(s+T)$. Clearly, $s < t_k$ and $s \to \infty$ as $k \to \infty$. Furthermore, for k and hence s large enough we can choose n such that $t_n \le t_k \le \gamma t_n$ and $t_n \le s \le t_{n+1}$. Clearly, $n < k$ and $n \to \infty$ as $k \to \infty$. We can write

$$E_{0,x}\{f(X_k)\} - \pi(f) = \int_D P_{0,x}\{X_n \in dy\}(E_{0,x;n,y}\{f(X_k)\} - \pi(f)). \tag{2.5}$$

Now

$$\begin{aligned} E_{0,x;n,y}\{f(X_k)\} - \pi(f) = {} & E_{0,x;n,y}\{f(X_k)\} - E_{t_n,y}\{f(x(t_k))\} \\ & + E_{t_n,y}\{f(x(\beta(s+T)))\} - E_{s,y}\{f(x(\beta(s+T)))\} \\ & + E_{s,y}\{f(x(\beta(s+T)))\} - \pi^{c(s+T)}(f) \\ & + \pi^{c(s+T)}(f) - \pi(f) \to 0 \qquad \text{as} \quad k \to \infty \end{aligned} \tag{2.6}$$

uniformly for $x, y \in D$ by Lemmas 1–3 and (A6). Combining (2.5) and (2.6) completes the proof. □

As an illustration of our theorem, we examine the random directions version of (1.2) that was implemented in [13]. If we could make noiseless measurements of $\nabla U(X_k)$, then we could use the algorithgm

$$X_{k+1} = X_k - a_k \nabla U(X_k) + b_k W_k \tag{2.7}$$

(modified as in (2.1)). Suppose that $\nabla U(X_k)$ is not available but we can make noiseless measurements of $U(\cdot)$. If we replace $\nabla U(X_k)$ in (2.7) by a forward finite difference aproximation of $\nabla U(X_k)$, then $d + 1$ evaluations of $U(\cdot)$ would be required at each iteration. As an alternative, suppose that at each iteration a direction D_k is chosen at random and we replace $\nabla U(X_k)$ in (2.7) by a finite difference approximation of the directional derivative $\langle \nabla U(X_k), D_k \rangle D_k$ in the direction D_k, which only requires two evaluations of $U(\cdot)$. Conceivably, fewer evalutions of $U(\cdot)$ would be required by such a random directions algorithm to converge. Now assume that the $\{D_k\}$ are random vectors each distributed uniformly over the surface of the $(d-1)$-dimensional sphere and that D_k is independent of $X_0, W_0, \ldots, W_{k-1}, D_0, \ldots, D_{k-1}$. By analysis similar to that on pp. 58–60 of [4] it can be shown that such a random directions algorithm can be written in the form of (1.2) with $E\{\xi_k | \mathscr{F}_k\} = O(h_k)$ and $\xi_k = O(1)$ where $\{h_k\}$ are the finite difference intervals ($h_k \to 0$). Hence the conditions of the theorem will be satisfied and convergence will be obtained provided $h_k = O(k^{-\gamma})$ for some $\gamma > 0$.[4]

Our theorem, like Kushner's [12], requires that the trajectories of $\{X_k\}$ be bounded. However, there is a version of Lemma 3 in [11] which applies with $D = \mathbb{R}^d$ assuming certain growth conditions on $U(\cdot)$. We are currently trying to obtain versions of Lemmas 1 and 2 which also hold for $D = \mathbb{R}^d$. On the other hand, we have found that bounding the trajectories of $\{X_k\}$ seems useful and even necessary in practice. The reason is that even with the specified growth conditions $|X_k|$ tends occasionally to very large values which leads to numerical problems in the simulation.

3. Proofs of Lemmas 1 and 2. Throughout this section it is convenient to make the following assumption in place of (A4):

(A4′) $c^2(t_k) = C/\log\log k$, k large, where $C > 0$, and $c^2(\cdot)$ is a piecewise linear interpolation of $\{c^2(t_k)\}$.

Note that under (A4′) $c^2(t) \sim C/\log t$ as $t \to \infty$, and if $B/A = C$, then $b_k = \sqrt{a_k}\,c(t_k)$ for k large enough. The results are unchanged whether we assume (A4) or (A4′). We also assume that a_k, b_k, and $c(t)$ are all bounded above by 1. In the following $c_1, c_2, \ldots,$ denote positive constants whose value may change from proof to proof.

We start with several propositions.

PROPOSITION 1.

$$P\{\tilde{X}_{k+1} \notin D | \mathscr{F}_k\} = O(a_k^{2+\alpha}) \qquad \textit{as} \quad k \to \infty,$$

*uniformly w.p.*1.

PROOF. We shall show that for k large enough (and w.p.1)

$$P\{\tilde{X}_{k+1} \notin D, |W_k| \geq \sqrt{k} | \mathscr{F}_k\} \leq c_1 a_k^{2+\alpha}, \tag{3.1}$$

$$P\{\tilde{X}_{k+1} \notin D, |W_k| \leq \sqrt{k} | \mathscr{F}_k\} \leq c_2 a_k^{2+\alpha}, \qquad |X_k| < r_1, \tag{3.2}$$

$$P\{\tilde{X}_{k+1} \notin D, |W_k| \leq \sqrt{k} | \mathscr{F}_k\} = 0, \qquad |X_k| \geq r_1. \tag{3.3}$$

Combining (3.1)–(3.3) gives the proposition.

[4] Note that we are assuming that $\nabla U(\cdot)$ is known exactly (and points outward) in a thin annulus near the boundary of D so that assumptions (A1) and (A5) are satisfied; this could be accomplished by using a penalty function in a straightforward manner.

Using a standard estimate for the tail probability of a Gaussian random variable we have

$$P\{\tilde{X}_{k+1} \notin D, |W_k| \geq \sqrt{k} | \mathscr{F}_k\} \leq d \exp\left(-\frac{k}{2d}\right) \leq c_1 a_k^{2+\alpha} \qquad \text{w.p.1}$$

and (3.1) is proved.

Assume $|X_k| < r_1$. Let $0 < \varepsilon < r - r_1$. Using the fact that $\sqrt{k} b_k \to 0$ and also the Chebyshev inequality we have for k large enough

$$\begin{aligned} P\{\tilde{X}_{k+1} \notin D, |W_k| \leq \sqrt{k} | \mathscr{F}_k\} \\ &\leq P\{|-a_k(\nabla U(X_k) + \xi_k) + b_k W_k| > r - r_1, |W_k| \leq \sqrt{k} | \mathscr{F}_k\} \\ &\leq P\{a_k |\xi_k| > r - r_1 - \varepsilon | \mathscr{F}_k\} \leq \frac{a_k^2 E\{|\xi_k|^2 | \mathscr{F}_k\}}{(r - r_1 - \varepsilon)^2} \leq c_2 a_k^{2+\alpha} \qquad \text{w.p.1} \end{aligned}$$

and (3.2) is proved.

Assume $|X_k| \geq r_1$. By assumption $\langle \nabla U(X_k), X_k \rangle > c_3 > 0$ and $\xi_k = 0$. Let $\bar{X}_k = X_k + b_k \sigma(X_k) W_k 1_{\{|W_k| \leq \sqrt{k}\}}$. Since $\sigma(\cdot)$ is Lipshitz, $\sigma(x) > 0$ for $|x| < r$, and $\sigma(x) = 0$ for $|x| = r$, we have $\sigma(x) \leq c_4 \inf_{|y|=r} |x - y|$. Hence $|\bar{X}_k - X_k| \leq c_4 \sqrt{k} b_k \inf_{|y|=r} |X_k - y|$, and since $\sqrt{k} b_k \to 0$ as $k \to \infty$, we get $\bar{X}_k - X_k \to 0$ as $k \to \infty$ and also $\bar{X}_k \in D$ for k large enough. Now $\bar{X}_k - X_k \to 0$ as $k \to \infty$ implies $\langle \nabla U(X_k), \bar{X}_k \rangle > c_5 > 0$ for k large enough. Hence $\bar{X}_k \in D$ for k large implies $\bar{X}_k - a_k \nabla U(X_k) \in D$ for k large. Hence for k large enough

$$P\{\tilde{X}_{k+1} \notin D, |W_k| \leq \sqrt{k} | \mathscr{F}_k\} \leq P\{\bar{X}_k - a_k \nabla U(X_k) \notin D | \mathscr{F}_k\} = 0 \qquad \text{w.p.1}$$

and (3.3) is proved. □

REMARK. By Proposition 1 and the Borel–Cantelli lemma $P \bigcup_n \bigcap_{k \geq n} \{\tilde{X}_k \in D\} = 1$ when $\alpha > -1$.

PROPOSITION 2. *For each n let $\{u_{n,k}\}_{k \geq n}$ be a sequence of nonnegative numbers such that*

$$\begin{aligned} u_{n,k+1} &\leq (1 + M a_k) u_{n,k} + M a_k^{\delta}, \qquad k \geq n, \\ u_{n,n} &\leq M a_n^{\varepsilon}, \end{aligned}$$

where $\delta > 1$, $\varepsilon > 0$, and $M > 0$. Then there exists a $\gamma > 1$ such that

$$\lim_{n \to \infty} \sup_{k : t_n \leq t_k \leq \gamma t_n} u_{n,k} = 0.$$

PROOF. We may set $M = 1$ since $a_k = A/k$ for k large and the proof is for arbitrary $A > 0$. Now

$$u_{n,k} \le u_{n,n} \prod_{l=n}^{k-1} (1 + a_l) + \sum_{m=n}^{k-1} a_m^\delta \prod_{l=m+1}^{k-1} (1 + a_l)$$
$$\le \left(u_{n,n} + \sum_{m=n}^{k-1} a_m^\delta \right) \cdot \exp\left(\sum_{m=n}^{k-1} a_m \right),$$

since $1 + x \le e^x$. Also $\sum_n^{k-1} a_m \le A(\log(k/n) + 1/n)$ and $\sum_n^{k-1} a_m^\delta \le A(1/(\delta - 1)n^{\delta-1} + 1/n^\delta)$, and if $t_k \le \gamma t_n$, then $k \le c_1 n^\gamma$. Choose γ such that $1 < \gamma < 1 + \min\{\delta - 1, \varepsilon\}/A$. It follows that

$$\sup_{k : t_n \le t_k \le \gamma t_n} u_{n,k} \le c_2 \left(\frac{1}{n^\varepsilon} + \frac{1}{n^{\delta-1}} \right) n^{(\gamma-1)A} \to 0 \qquad \text{as} \quad n \to \infty. \qquad \square$$

REMARK. Proposition 2 is used to make the long time comparisons in the proofs of Lemmas 1 and 2. Proposition 2 does *not* hold if we take $a_k = A/k^\eta$ for $\eta < 1$.

Define $\xi(\cdot, \cdot)$ by

$$x(t) = x(s) - (t - s)(\nabla U(x(s)) + \xi(s, t)) + c(s)\sigma(x(s))(w(t) - w(s))$$

for $t \ge s \ge 0$.

PROPOSITION 3.

$$E\{|\xi(t, t + h)|^2 \,|x(t)\} = O(1),$$
$$E\{\xi(t, t + h)|x(t)\} = O(h^{1/2})$$

as $h \to 0$, *uniformly for a.e.* $x(t) \in D$ *and* $t \ge 0$.

PROOF. We use some elementary facts about stochastic integrals and martingales (see [15]). First write

$$h\xi(t, t + h) = \int_t^{t+h} (\nabla U(x(u)) - \nabla U(x(t)))\, du - \int_t^{t+h} (c(u)\sigma(x(u)) - c(t)\sigma(x(t)))\, dw(u). \tag{3.4}$$

Now a standard result is that

$$E_{t,x}\{|x(t + h) - x(t)|^2\} = O(h)$$

as $h \to 0$, uniformly for $x \in D$ and t in a finite interval. In fact, under our assumptions the estimate is uniform here for $x \in D$ and all $t \ge 0$. Let K_1, K_2 be

Lipshitz constants for $\nabla U(\cdot)$, $\sigma(\cdot)$, respectively. Also note that $c(\cdot)$ is piecewise continuously differentiable with bounded derivative (where it exists) and hence is also Lipshitz continuous, say with constant K_3. Hence

$$\begin{aligned}(3.5)\qquad E_{t,x}&\left\{\left|\int_t^{t+h} (\nabla U(x(u)) - \nabla U(x(t)))\,du\right|^2\right\}\\ &\le K_1^2 E_{t,x}\left\{\left(\int_t^{t+h} |x(u)-x(t)|\,du\right)^2\right\}\\ &\le 2K_1^2 h\int_t^{t+h} E_{t,x}\{|x(u)-x(t)|^2\}\,du = O(h^3)\end{aligned}$$

and

$$\begin{aligned}(3.6)\quad E_{t,x}&\left\{\left|\int_t^{t+h} (c(u)\sigma(x(u)) - c(t)\sigma(x(t)))\,dw(u)\right|^2\right\}\\ &= \int_t^{t+h} E_{t,x}\{|c(u)\sigma(x(u)) - c(t)\sigma(x(t))|^2\}\,du\\ &\le 3K_2^2\int_t^{t+h} E_{t,x}\{|x(u)-x(t)|^2\}\,du + 3K_3^2\int_t^{t+h}(u-t)^2\,du = O(h^2)\end{aligned}$$

as $h \to 0$, uniformly for $x \in D$ and all $t \ge 0$. The proposition follows easily from (3.4)-(3.6) and the fact that the second (stochastic) integral in (3.4) defines a martingale as h varies. □

Now in Lemma 1 we compare the distributions of X_k and $x(t_k)$. This is done most easily by comparing X_k and $x(t_k)$ to Y_k and $\bar{Y}_k$ (defined below), respectively, which are equal in distribution.

Let

$$\begin{aligned}\tilde{Y}_{k+1} &= Y_k - a_k\nabla U(Y_k) + b_k\sigma(Y_k)W_k,\\ Y_{k+1} &= \tilde{Y}_{k+1}1_D(\tilde{Y}_{k+1}) + Y_k 1_{D^c}(\tilde{Y}_{k+1}).\end{aligned}$$

LEMMA 1.1. *There exists $\gamma > 1$ such that for any continuous function $f(\cdot)$ on D*

$$\lim_{n\to\infty}\ \sup_{k:t_n\le t_k\le \gamma t_n}\ (E_{0,x;n,y}\{f(X_k)\} - E_{n,y}\{f(Y_k)\}) = 0$$

uniformly for x, $y \in D$.

PROOF. Assume all quantities are conditioned on $X_0 = x$ and $X_n = Y_n = y$, with x, $y \in D$. Let $\Delta_k = X_k - Y_k$. Write

$$\begin{aligned}(3.7)\qquad E\{|\Delta_{k+1}|^2\} &= E\{|\Delta_{k+1}|^2 1_{\{\tilde{X}_{k+1}\notin D\}\cup\{\tilde{Y}_{k+1}\notin D\}}\}\\ &\quad + E\{|\Delta_{k+1}|^2 1_{\{\tilde{X}_{k+1}\in D\}\cap\{\tilde{Y}_{k+1}\in D\}}\}.\end{aligned}$$

We estimate the first term in (3.7) as follows. We have by Proposition 1 that

$$(3.8)\qquad E\{|\Delta_{k+1}|^2 1_{\{\tilde{X}_{k+1}\notin D\}\cup\{\tilde{Y}_{k+1}\notin D\}}\} \le c_1(P\{\tilde{X}_{k+1}\notin D\} + P\{\tilde{Y}_{k+1}\notin D\}) \le c_2 a_k^{2+\alpha}, \qquad k \ge n.$$

We estimate the second term in (3.7) as follows. If $\tilde{X}_{k+1} \in D$ and $\tilde{Y}_{k+1} \in D$, then

$$\Delta_{k+1} = \Delta_k - a_k(\nabla U(Y_k + \Delta_k) - \nabla U(Y_k)) + b_k(\sigma(Y_k + \Delta_k) - \sigma(Y_k))W_k - a_k\xi_k.$$

Hence

$$\begin{aligned}(3.9)\ E\{|\Delta_{k+1}|^2 1_{\{\tilde{X}_{k+1}\in D\}\cap\{\tilde{Y}_{k+1}\in D\}}\} &\le E\{|\Delta_k - a_k(\nabla U(Y_k+\Delta_k) - \nabla U(Y_k)) \\ &\qquad + b_k(\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k - a_k\xi_k|^2\} \\ &\le E\{|\Delta_k|^2\} + a_k^2 E\{|\nabla U(Y_k+\Delta_k) - \nabla U(Y_k)|^2\} \\ &\quad + a_k E\{|(\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k|^2\} \\ &\quad + a_k^2 E\{|\xi_k|^2\} \\ &\quad + 2a_k|E\{\langle\Delta_k, \nabla U(Y_k+\Delta_k) - \nabla U(Y_k)\rangle\}| \\ &\quad + 2a_k^{1/2}|E\{\langle\Delta_k, (\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k\rangle\}| \\ &\quad + 2a_k|E\{\langle\Delta_k, \xi_k\rangle\}| \\ &\quad + 2a_k^{3/2}|E\{\langle\nabla U(Y_k+\Delta_k) - \nabla U(Y_k), (\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k\rangle\}| \\ &\quad + 2a_k^2|E\{\langle\nabla U(Y_k+\Delta_k) - \nabla U(Y_k), \xi_k\rangle\}| \\ &\quad + 2a_k^{3/2}|E\{\langle(\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k, \xi_k\rangle\}|, \qquad k \ge n.\end{aligned}$$

Let K_1, K_2 be Lipshitz constants for $\nabla U(\cdot)$, $\sigma(\cdot)$, respectively. Using the fact that X_k, Y_k and hence Δ_k are $\mathscr{F}_k$ measurable, W_k is independent of $\mathscr{F}_k$, $E\{W_k\} = 0$, and

$$E\{|\xi_k|^2|\mathscr{F}_k\} \le c_3 a_k^\alpha, \qquad |E\{\xi_k|\mathscr{F}_k\}| \le c_3 a_k^\beta \quad \text{w.p.1}$$

we have

$$\begin{aligned} E\{|\nabla U(Y_k+\Delta_k) - \nabla U(Y_k)|^2\} &\le K_1^2 E\{|\Delta_k|^2\}, \\ E\{|(\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k|^2\} &\le K_2^2 d E\{|\Delta_k|^2\}, \\ E\{|\xi_k|^2\} &\le c_3 a_k^\alpha, \\ |E\{\langle\Delta_k, \nabla U(Y_k+\Delta_k) - \nabla U(Y_k)\rangle\}| &\le K_1 E\{|\Delta_k|^2\}, \\ |E\{\langle\Delta_k, (\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k\rangle\}| &= 0, \\ |E\{\langle\Delta_k, \xi_k\rangle\}| &\le c_3 a_k^\beta E\{|\Delta_k|\}, \\ |E\{\langle\nabla U(Y_k+\Delta_k) - \nabla U(Y_k), (\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k\rangle\}| &= 0, \\ |E\{\langle\nabla U(Y_k+\Delta_k) - \nabla U(Y_k), \xi_k\rangle\}| &\le K_1 c_3 a_k^\beta E\{|\Delta_k|\}, \\ |E\{\langle(\sigma(Y_k+\Delta_k) - \sigma(Y_k))W_k, \xi_k\rangle\}| &\le K_2\sqrt{dc_3}\, a_k^{\alpha/2} E\{|\Delta_k|\}\end{aligned}$$

for $k \geq n$. Substituting these expressions into (3.9) gives (after some simplification)

$$(3.10)\qquad \begin{aligned} &E\{|\Delta_{k+1}|^2 1_{\{\tilde{X}_{k+1}\in D\}\cap\{\tilde{Y}_{k+1}\in D\}}\} \\ &\leq (1 + c_4 a_k)E\{|\Delta_k|^2\} + c_4 a_k^{\delta_1} E\{|\Delta_k|\} + c_3 a_k^{2+\alpha} \\ &\leq (1 + c_4 a_k)E\{|\Delta_k|^2\} + c_4 a_k^{\delta_1} E\{|\Delta_k|^2\}^{1/2} + c_3 a_k^{2+\alpha} \\ &\leq (1 + c_5 a_k)E\{|\Delta_k|^2\} + c_5 a_k^{\delta_2}, \qquad k \geq n, \end{aligned}$$

where $\delta_1 = \min\{1 + \beta, (3 + \alpha)/2\} > 1$ and $\delta_2 = \min\{\delta_1, 2 + \alpha\} > 1$ since $\alpha > -1$ and $\beta > 0$.

Now combine (3.7), (3.8), and (3.10) to get

$$\begin{aligned} E\{|\Delta_{k+1}|^2\} &\leq (1 + c_6 a_k)E\{|\Delta_k|^2\} + c_6 a_k^{\delta_2}, \qquad k \geq n, \\ E\{|\Delta_n|^2\} &= 0 \end{aligned}$$

for n large enough. Applying Proposition 2 there exists $\gamma > 1$ such that

$$(3.11)\qquad \lim_{n\to\infty} \sup_{k: t_n \leq t_k \leq \gamma t_n} E\{|\Delta_k|^2\} = 0.$$

Finally, let $f(\cdot)$ be a continuous function on D. Since $f(\cdot)$ is uniformly continuous on D, given $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(u) - f(v)| < \varepsilon$ whenever $|u - v| < \delta$ and $u, v \in D$. Hence

$$\begin{aligned} |E\{f(X_k)\} - E\{f(Y_k)\}| &\leq \varepsilon P\{|\Delta_k| < \delta\} + 2\|f\| P\{|\Delta_k| \geq \delta\} \\ &\leq \varepsilon + \frac{2\|f\|}{\delta^2} E\{|\Delta_k|^2\}, \end{aligned}$$

and by (3.11)

$$\overline{\lim_{n\to\infty}} \sup_{k: t_n \leq t_k \leq \gamma t_n} |E\{f(X_k)\} - E\{f(Y_k)\}| \leq \varepsilon,$$

and letting $\varepsilon \to 0$ completes the proof. □

Let

$$\bar{W}_k = (w(t_{k+1}) - w(t_k))/\sqrt{a_k}$$

and

$$\begin{aligned} \tilde{\bar{Y}}_{k+1} &= \bar{Y}_k - a_k \nabla U(\bar{Y}_k) + b_k \sigma(\bar{Y}_k)\bar{W}_k, \\ \bar{Y}_{k+1} &= \tilde{\bar{Y}}_{k+1} 1_D(\tilde{\bar{Y}}_{k+1}) + \bar{Y}_k 1_{D^c}(\tilde{\bar{Y}}_{k+1}). \end{aligned}$$

LEMMA 1.2. *There exists $\gamma > 1$ such that for any continuous function $f(\cdot)$ on D*

$$\lim_{n\to\infty} \sup_{k:t_n\le t_k\le \gamma t_n} (E_{t_n,y}\{f(x(t_k))\} - E_{n,y}\{f(\bar{Y}_k)\}) = 0$$

uniformly for $y \in D$.

PROOF. Define $\{\bar{\xi}_k\}$ by

$$x(t_{k+1}) = x(t_k) - a_k(\nabla U(x(t_k)) + \bar{\xi}_k) + b_k\sigma(x(t_k))\bar{W}_k.$$

Let $\bar{\mathscr{F}}_k$ be the σ-field generated by $\{x(0), \bar{\xi}_0, \dots, \bar{\xi}_{k-1}, \bar{W}_0, \dots, \bar{W}_{k-1}\}$. It can be shown that $\bar{\xi}_k$ is conditionally independent of $\bar{\mathscr{F}}_k$ given $x(t_k)$. Hence by Proposition 3

$$E\{|\bar{\xi}_k|^2|\bar{\mathscr{F}}_k\} \le c_1, \qquad |E\{\bar{\xi}_k|\bar{\mathscr{F}}_k\}| \le c_1 a_k^{1/2} \quad \text{w.p.1.} \tag{3.12}$$

Henceforth assume all quantities are conditioned on $x(t_n) = \bar{Y}_n = y$, $y \in D$. Let $\Delta_k = x(t_k) - \bar{Y}_k$. Using (3.12) and proceeding similarly to the proof of Lemma 1.1 we can show with $\delta = 3/2$ that

$$E\{|\Delta_{k+1}|^2\} \le (1 + c_2 a_k)E\{|\Delta_k|^2\} + c_2 a_k^\delta, \qquad k \ge n,$$
$$E\{|\Delta_n|^2\} = 0.$$

Applying Proposition 2 there exists a $\gamma > 1$ such that

$$\lim_{n\to\infty} \sup_{k:t_n\le k\le \gamma t_n} E\{|\Delta_k|^2\} = 0.$$

The lemma follows from this. □

PROOF OF LEMMA 1. Follows immediately from Lemmas 1.1 and 1.2. □

PROOF OF LEMMA 2. Let $x(\cdot; s, y)$ denote the process $x(\cdot)$ emitted from y at time s.

Fix $y \in D$, n, and $s \in [t_n, t_{n+1})$, and let $x_1(\cdot) = x(\cdot; t_n, y)$ and $x_2(\cdot) = x(\cdot; s, y)$. Now recall that

$$E_{t,x}\{|x(t + h) - x(t)|^2\} = O(h) \qquad \text{as} \quad h \to 0$$

uniformly for $x \in D$ and all $t \ge 0$ (this is a standard result except for the uniformity for all t which was remarked on in Proposition 3). Hence

$$E\{|x_1(s) - x_2(s)|^2\} \le c_1(s - t_n) \le c_1 a_n.$$

Let $\Delta_k = x_1(t_k) - x_2(t_k)$ for $k > n$. Similarly to the proofs of Lemmas 1.1 and 1.2 we can show with $\delta = \frac{3}{2}$ that

$$E\{|\Delta_{k+1}|^2\} \le (1 + c_2 a_k) E\{|\Delta_k|^2\} + c_2 a_k^\delta, \qquad k \ge n+1,$$
$$E\{|\Delta_{n+1}|^2\} \le (1 + c_2(t_{n+1} - s)) E\{|x_1(s) - x_2(s)|^2\} + c_2(t_{n+1} - s)^\delta$$
$$\le (1 + c_2 a_n) c_1 a_n + c_2 a_n^\delta \le c_3 a_{n+1}$$

and the same inequalities hold if we take suprema over $s \in [t_n, t_{n+1})$. Applying Proposition 2 there exists $\gamma > 1$ such that

$$\lim_{n \to \infty} \sup_{k: t_{n+1} \le t_k \le \gamma t_n} \sup_{s: t_n \le s \le t_{n+1}} E\{|\Delta_k|^2\} = 0. \tag{3.13}$$

Note that $\beta(s)$ is a strictly increasisng function of s and $s + s^{2/3} \le \beta(s) \le s + 2s^{2/3}$ for s large enough. Hence for n large enough we can choose s such that $t_n \le s \le t_{n+1}$ and m such that $t_m \le \beta(s) \le t_{m+1}$ and $t_{n+1} \le t_m \le \gamma t_n$. Now

$$E\{|x_1(\beta(s)) - x_2(\beta(s))|^2\} \le (1 + c_2(\beta(s) - t_m)) E\{|\Delta_m|^2\} + c_2 a_m^\delta \tag{3.14}$$
$$\le c_4 \sup_{k: t_{n+1} \le t_k \le \gamma t_n} E\{|\Delta_k|^2\} + c_2 a_n^\delta.$$

Combining (3.13) and (3.14) gives

$$\lim_{n \to \infty} \sup_{s: t_n \le s \le t_{n+1}} E\{|x_1(\beta(s)) - x_2(\beta(s))|^2\} = 0.$$

The lemma follows from this. □

4. Proof of Lemma 3. The idea behind the proof of Lemma 3 is roughly as follows. Recall that $D = \{x: |x| \le r\}$ and $r_0 < r_1 < r$. First, we show that no matter where $x(s)$ is ($|x(s)| \le r$), there exists $T > 0$ such that $|x(s+T)| \le r_0$ with large probability for s large. Then we show that $|x(t)| \le r_1$ for all $t \in [s+T, \beta(s+T)]$ with large probability for s large. This allows us to make use of results from [11] which hold for a nondegenerate diffusion ($\sigma(x) = 1$ for all x in (2.2)).

LEMMA 3.1. *Given $\delta > 0$ there exists $T > 0$ such that*

$$\lim_{s \to \infty} P_{s,y}\{|x(s+T)| \le r_0 + \delta\} = 1$$

uniformly for $|y| \le r$.

PROOF. Let

$$\dot{z}(t) = -\nabla U(z(t)),$$

where $z(s) = y$, $|y| \le r$. Then there exists $T > 0$ (where T does not depend on s

or y) such that

$$z(s+T) \le r_0. \tag{4.1}$$

This follows from the fact that

$$|z(t)|^2 - |z(s)|^2 = -2\int_s^t \langle \nabla U(z(u)), z(u)\rangle \, du$$

and $\langle \nabla U(z(t)), z(t)\rangle > c_1 > 0$ when $|z(t)| > r_0$.

Now for $z(s) = x(s) = y$, $|y| \le r$,

$$\begin{aligned}|x(t) - z(t)| &\le \left|\int_s^t (\nabla U(x(u)) - \nabla U(z(u)))\, du\right| + \left|\int_s^t c(u)\sigma(x(u))\, dw(u)\right| \\ &\le K\int_s^t |x(u) - z(u)|\, du + \left|\int_s^t c(u)\sigma(x(u))\, dw(u)\right|,\end{aligned}$$

where K is a Lipshitz constant for $\nabla U(\cdot)$. Hence by Gronwall's inequality

$$|x(t) - z(t)| \le \exp(K(t-s)) \sup_{s\le v\le t}\left|\int_s^v c(u)\sigma(x(u))\, dw(u)\right|.$$

Hence by the Martingale inequality

$$\begin{aligned}&P_{s,y}\{|x(s+T) - z(s+T)| > \delta\} \\ &\le P_{s,y}\left\{\sup_{s\le v\le s+T}\left|\int_s^v c(u)\sigma(x(u))\, dw(u)\right| > \delta e^{-KT}\right\} \\ &\le \frac{e^{2KT}}{\delta^2} E_{s,y}\left\{\left|\int_s^{s+T} c(u)\sigma(x(u))\, dw(u)\right|^2\right\} \\ &= \frac{e^{2KT}}{\delta^2}\int_s^{s+T} E_{s,y}\{|c(u)\sigma(x(u))|^2\}\, du \\ &\le \frac{e^{2KT}}{\delta^2} Tc^2(s) \to 0 \quad \text{as} \quad s\to\infty.\end{aligned} \tag{4.2}$$

Combining (4.1) and (4.2) gives the lemma. □

LEMMA 3.2. *Let*

$$\tau = \inf\{t: |x(t)| > r_1\}.$$

Let $C_0 < C < C_1$. Then there exists $\delta > 0$ such that

$$\lim_{s \to \infty} P_{s,y}\{\tau > \beta(s)\} = 1$$

uniformly for $|y| \le r_0 + \delta$.

PROOF. Let $\hat{U}(\cdot)$ be a twice continuously differentiable function from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that for some $R > r$ and $K > 0$

$$\hat{U}(x)\begin{cases} = U(x), & |x| \le r, \\ = K|x|^2, & |x| > R, \end{cases}$$

and $\nabla \hat{U}(x) \ne 0$ for $r < |x| \le R$ (in view of (A1) such a $\hat{U}(\cdot)$ exists). For $\varepsilon > 0$ let

$$d\hat{x}^\varepsilon(t) = -\nabla \hat{U}(\hat{x}^\varepsilon(t))\, dt + \varepsilon\, dw(t)$$

and

$$\hat{\tau}^\varepsilon = \inf\{t: |\hat{x}^\varepsilon(t)| > r_1\}.$$

For $0 < \delta < r_1 - r_0$ let

$$C_2(\delta) = \inf_{|x| = r_1} \hat{U}(x) - \sup_{|x| = r_0 + \delta} \hat{U}(x).$$

On p. 750 of [11] it is shown that for any $\eta > 0$ and $\delta > 0$

$$P_{0,y}\left\{\hat{\tau}^\varepsilon > \exp\left(\frac{1}{\varepsilon^2}(C_2(\delta) - \eta)\right)\right\} \to 1 \quad \text{as} \quad \varepsilon \to 0,$$

uniformly for $|y| \le r_0 + \delta$. Since $C_2(\delta) \to \frac{2}{3}C_1$ as $\delta \to 0$, it follows that for any $\eta > 0$ there exists $\delta > 0$ such that

$$P_{0,y}\left\{\hat{\tau}^\varepsilon > \exp\left(\frac{1}{\varepsilon^2}(\tfrac{2}{3}C_1 - \eta)\right)\right\} \to 1 \quad \text{as} \quad \varepsilon \to 0 \tag{4.3}$$

uniformly for $|y| \le r_0 + \delta$.

Next let

$$d\hat{x}(t) = -\nabla \hat{U}(\hat{x}(t))\, dt + c(t)\, dw(t)$$

and

$$\hat{\tau} = \inf\{t: |\hat{x}(t)| > r_1\}.$$

On p. 745 of [11] it is shown that

$$P_{s,y}\{\hat{\tau} > \beta(s)\} - P_{0,y}\{\hat{\tau}^{c(s)} > s^{2/3}\} \to 0 \qquad \text{as} \quad s \to \infty \tag{4.4}$$

uniformly for $|y| \le r$.

Now choose $\eta > 0$ such that

$$\frac{1}{C}(\tfrac{2}{3}C_1 - \eta) \ge \tfrac{2}{3}$$

and choose $\delta > 0$ such that (4.3) is satisfied. Hence using (4.3) and (4.4)

$$\begin{aligned} P_{s,y}\{\tau > \beta(s)\} &= P_{s,y}\{\hat{\tau} > \beta(s)\} \\ &= P_{0,y}\{\hat{\tau}^{c(s)} > s^{2/3}\} + (P_{s,y}\{\hat{\tau} > \beta(s)\} - P_{0,y}\{\hat{\tau}^{c(s)} > s^{2/3}\}) \\ &\ge P_{0,y}\left\{\hat{\tau}^{c(s)} > \exp\left(\frac{1}{c^2(s)}\left(\frac{2}{3}C_1 - \eta\right)\right)\right\} + o(1) \\ &\to 1 \qquad \text{as} \quad s \to \infty \end{aligned}$$

uniformly for $|y| \le r_0 + \delta$. □

Proof of Lemma 3. Let $\hat{x}(\cdot)$ be defined as in the proof of Lemma 3.2. In Lemmas 1–3 of [11] it is shown that

$$E_{s,y}\{f(\hat{x}(\beta(s)))\} - \pi^{c(s)}(f) \to 0 \qquad \text{as} \quad s \to \infty \tag{4.5}$$

uniformly for $|y| \le r$. By Lemma 3.2 there exists $\delta > 0$ such that

$$\begin{aligned} &|E_{s,y}\{f(x(\beta(s)))\} - E_{s,y}\{f(\hat{x}(\beta(s)))\}| \\ &\quad \le |E_{s,y}\{f(x(\beta(s))) - f(\hat{x}(\beta(s))), \tau > \beta(s)\}| + 2\|f\| P_{s,y}\{\tau \le \beta(s)\} \\ &\quad \to 0 \qquad \text{as} \quad s \to \infty \end{aligned} \tag{4.6}$$

uniformly for $|y| \le r_0 + \delta$. Hence combining (4.5) and (4.6) and using Lemma 3.1 there exists $T > 0$ such that

$$\begin{aligned} &|E_{s,y}\{f(x(\beta(s+T)))\} - \pi^{c(s+T)}(f)| \\ &\quad = |E_{s,y}\{E_{s+T,x(s+T)}\{f(x(\beta(s+T))) - \pi^{c(s+T)}(f)\}\}| \\ &\quad \le |E_{s,y}\{E_{s+T,x(s+T)}\{f(x(\beta(s+T))) - \pi^{c(s+T)}(f)\}1_{\{|x(s+T)| \le r_0 + \delta\}}\}| \\ &\qquad + 2\|f\| P_{s,y}\{|x(s+T)| > r_0 + \delta\} \\ &\quad \to 0 \qquad \text{as} \quad s \to \infty \end{aligned}$$

uniformly for $|y| \le r$. □

References

[1] Wasan, M. T., *Stochastic Approximation*, Cambridge University Press, Cambridge, 1969.

[2] Rubinstein, R. Y., *Simulation and the Monte Carlo Method*, Wiley, New York, 1981.

[3] Ljung, L., Analysis of recursive stochastic algorithms, *IEEE Transactions on Automatic Control*, Vol. 22, pp. 551-575, 1977.

[4] Kushner, H., and Clark, D., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Applied Mathematical Sciences, Vol. 26, Springer Verlag, Berlin, 1978.

[5] Metivier, M., and Priouret, P., Applications of a Kushner and Clark lemma to general classes of stochastic algorithm, *IEEE Transactions on Information Theory*, Vol. 30, pp. 140-151, 1984.

[6] Grenender, U., Tutorial in Pattern Theory, Division of Applied Mathematics, Brown University, Providence, RI, 1984.

[7] Geman, S., and Hwang, C. R., Diffusions for global optimization, *SIAM Journal on Control and Optimization*, Vol. 24, pp. 1031-1043, 1986.

[8] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., Optimization by simulated annealing, *Science*, Vol. 220, pp. 621-680, 1983.

[9] Gelfand, S. B., Analysis of Simulated Annealing Type Algorithms, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1987.

[10] Gidas, B., Global optimization via the Langevin equation, *Proceedings of the IEEE Conference on Decision and Control*, pp. 774-778, Fort Lauderdale, FL, 1985.

[11] Chiang, T. S., Hwang, C. R., and Sheu, S. J., Diffusion for global optimization in $\mathbb{R}^n$, *SIAM Journal on Control and Optimization*, Vol. 25, pp. 737-752, 1987.

[12] Kushner, H. J., Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: Global minimization via Monte Carlo, *SIAM Journal on Applied Mathematics*, Vol. 47, pp. 169-185, 1987,

[13] Aluffi-Pentini, F., Parisi, V., and Zirilli, F., Global optimization and stochastic differential equations, *Journal of Optimization Theory and Applications*, Vol. 47, pp. 1-16, 1985.

[14] Hwang, C.-R., Laplaces method revisited: weak convergence of probability measures, *Annals of Probability*, Vol. 8, pp. 1177-1182, 1980.

[15] Gikhman, I. I., and Skorohod, A. V., *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1972.