SIAM J. CONTROL AND OPTIMIZATION
Vol. 29, No. 5, pp. 999-1018, September 1991

RECURSIVE STOCHASTIC ALGORITHMS FOR GLOBAL OPTIMIZATION IN $\mathbb{R}^d$*

SAUL B. GELFAND† AND SANJOY K. MITTER‡

Abstract. An algorithm of the form $X_{k+1}=X_k-a_k(\nabla U(X_k)+\xi_k)+b_kW_k$, where $U(\cdot)$ is a smooth function on $\mathbb{R}^d$, $\{\xi_k\}$ is a sequence of $\mathbb{R}^d$-valued random variables, $\{W_k\}$ is a sequence of independent standard d-dimensional Gaussian random variables, $a_k=A/k$ and $b_k=\sqrt{B}/\sqrt{k\log\log k}$ for k large, is considered. An algorithm of this type arises by adding slowly decreasing white Gaussian noise to a stochastic gradient algorithm. It is shown, under suitable conditions on $U(\cdot)$, $\{\xi_k\}$, A, and B, that X_k converges in probability to the set of global minima of $U(\cdot)$. No prior information is assumed as to what bounded region contains a global minimum. The analysis is based on the asymptotic behavior of the related diffusion process $dY(t)=-\nabla U(Y(t))dt+c(t)dW(t)$, where $W(\cdot)$ is a standard d-dimensional Wiener process and $c(t)=\sqrt{C}/\sqrt{\log t}$ for t large.

Key words. global optimization, random optimization, simulated annealing, stochastic gradient algorithms, diffusions

AMS(MOS) subject classifications. 65K10, 90C30, 60J60

1. Introduction. In this paper we consider a class of algorithms for finding a global minimum of a smooth function $U(x)$, $x\in\mathbb{R}^d$. Specifically, we analyze the convergence of a modified stochastic gradient algorithm

$$X_{k+1}=X_k-a_k(\nabla U(X_k)+\xi_k)+b_kW_k, \tag{1.1}$$

where $\{\xi_k\}$ is a sequence of $\mathbb{R}^d$-valued random variables, $\{W_k\}$ is a sequence of standard d-dimensional independent Gaussian random variables, and $\{a_k\}$, $\{b_k\}$ are sequences of positive numbers with $a_k, b_k\to 0$. An algorithm of this type arises by artificially adding the b_kW_k term (via a Monte Carlo simulation) to a standard stochastic gradient algorithm,

$$Z_{k+1}=Z_k-a_k(\nabla U(Z_k)+\xi_k). \tag{1.2}$$

Algorithms like (1.2) arise in a variety of optimization problems including adaptive filtering, identification, and control; here the sequence $\{\xi_k\}$ is due to noisy or imprecise measurements of $\nabla U(\cdot)$ (cf. [1]). The asymptotic behavior of $\{Z_k\}$ has been extensively studied. Let S and S^* be the set of local and global minima of $U(\cdot)$, respectively. It can be shown, for example, that if $U(\cdot)$ and $\{\xi_k\}$ are suitably behaved, $a_k=A/k$ for k large, and $\{Z_k\}$ is bounded, then $Z_k\to S$ as $k\to\infty$ with probability one. However, in general, $Z_k\not\to S^*$ (unless of course $S=S^*$). The idea behind the additional b_kW_k term in (1.1) compared with (1.2) is that if b_k tends to zero slowly enough, then possibly $\{X_k\}$ (unlike $\{Z_k\}$) will avoid getting trapped in a strictly local minimum of $U(\cdot)$. In fact, we will show that if $U(\cdot)$ and $\{\xi_k\}$ are suitably behaved, $a_k=A/k$ and $b_k^2=B/k\log\log k$ for k large with $B/A>C_0$ (where C_0 is a positive constant that depends only on $U(\cdot)$), and $\{X_k\}$ is tight, then $X_k\to S^*$ as $k\to\infty$ in probability. We also give a condition for the tightness of $\{X_k\}$. We remark that in [1] both probability one and

* Received by the editors January 10, 1990; accepted for publication (in revised form) July 26, 1990. This research was supported by National Science Foundation grant ECS-8910073, Air Force Office of Scientific Research contract 89-0276B, and Army Research Office contract DAAL03-86-K-0171 (Center for Intelligent Control Systems).

† School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907.

‡ Department of Electrical Engineering and Computer Science, and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

weak convergence of $\{Z_k\}$ are treated. Furthermore, convergence of Z_k to S is established under very weak conditions on $\{\xi_k\}$ assuming that $\{Z_k\}$ is bounded. Here the convergence of X_k to S^* is established under somewhat stronger conditions on $\{\xi_k\}$ assuming that $\{X_k\}$ is tight (which is weaker than boundedness).

An algorithm like (1.1) was first proposed and analyzed by Kushner [2]. However, the analysis in [2] required that the trajectories of $\{X_k\}$ lie within a fixed ball (which was achieved by modifying (1.1) near the boundary of the ball). Hence, the version of (1.1) in [2] is only suitable for optimizing $U(\cdot)$ over a compact set. Some other differences between the results presented here and in [2] include conditions on $\{a_k\}$, $\{b_k\}$, and $\{\xi_k\}$, and also the method of analysis; these are discussed below.

The analysis of the convergence of $\{Z_k\}$ is usually based on the asymptotic behavior of the *associated ordinary differential equation* (ODE)

$$\dot{z}(t) = -\nabla U(z(t)) \tag{1.3}$$

(cf. [1], [3]). This motivates our analysis of the convergence of $\{X_k\}$ based on the asymptotic behavior of the *associated stochastic differential equation* (SDE)

$$dY(t) = -\nabla U(Y(t))\,dt + c(t)\,dW(t), \tag{1.4}$$

where $W(\cdot)$ is a standard d-dimensional Wiener process and $c(\cdot)$ is a positive function with $c(t)\to 0$ as $t\to\infty$. The diffusion $Y(\cdot)$ 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. Continuous simulated annealing was first suggested in [4] and [5] for global optimization problems that arise in image processing applications with continuous grey levels. Now the asymptotic behavior of $Y(t)$ as $t\to\infty$ has been studied intensively by a number of researchers. In [2] and [5], convergence results were obtained by considering a version of (1.4) with a reflecting boundary; in [6] and [7] the reflecting boundary was removed. Our analysis of $\{X_k\}$ is based on the analysis of $Y(\cdot)$ developed by Chiang, Hwang, and Sheu [7] who prove the following result: if $U(\cdot)$ is well behaved and $c^2(t)=C/\log t$ for t large with $C>C_0$ then $Y(t)\to S^*$ as $t\to\infty$ in probability. The main difficulty associated with using $Y(\cdot)$ to analyze $\{X_k\}$ is that we must deal with long time intervals and slowly decreasing (unbounded) Gaussian noise.

We make some further remarks on the differences between the results and methods in this paper as compared with [2]. We first note that in [2] the modified version of (1.1), which constrains the trajectories of $\{X_k\}$ to lie within a fixed ball, is analyzed for $a_k=b_k=A/\log k$, k large. Although a detailed asymptotic description of $\{X_k\}$ is obtained for this case, in general, $X_k \not\to S^*$ unless $\xi_k=0$. The reason for this is intuitively clear: even if $\{\xi_k\}$ is bounded, $a_k\xi_k$ and a_kW_k can be of the same order, and hence can interfere with each other. On the other hand, here we allow $\{\xi_k\}$ with unbounded variance, in particular, $E\{|\xi_k|^2\}=O(k^\gamma)$ and $\gamma<1$. This has important implications when $\nabla U(\cdot)$ is not measured exactly. We also note that the analysis in [2] is different from that done here, in that in [2] the behavior of $\{X_k\}$ is obtained by deriving various large deviations estimates of Donsker–Varadhan type, whereas here we obtain the behavior of $\{X_k\}$ directly from the corresponding behavior of $Y(\cdot)$. It should be pointed out that in a certain sense the results in [2] are also stronger than those presented here, because the large deviation approach in [2] treats the whole tail of the process $\{X_k\}$, while only "local" type results are discussed here. However, from our point of view the most significant difference between our work and that done in [2] (and more generally in other work on global optimization such as [8]) is that we deal with unbounded processes and establish the convergence of an algorithm that finds a global minimum of a function when it is not known a priori what bounded region contains such a point.

The paper is organized as follows. In § 2 we state our assumptions and main result. In § 3 we take up the proof of this result. In § 4 we prove a general tightness criterion, which is then used in § 5 to establish tightness and ultimately convergence for two example algorithms.

2. Main result. In this section we present our main result on the convergence of the discrete time algorithm

$$X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k, \qquad k \geqq 0, \tag{2.1}$$

which is closely related to the continuous time algorithm

$$dY(t) = -\nabla U(Y(t))\, dt + c(t)\, dW(t), \qquad t \geqq 0. \tag{2.2}$$

Here $U(\cdot)$ is a smooth function on $\mathbb{R}^d$, $\{\xi_k\}$ is a sequence of $\mathbb{R}^d$-valued random variables, $\{W_k\}$ is a sequence of independent d-dimensional Gaussian random variables with $E\{W_k\} = 0$ and $E\{W_k \otimes W_k\} = I$ (the identity matrix), $W(\cdot)$ is a standard d-dimensional Wiener process, and

$$a_k = \frac{A}{k}, \quad b_k^2 = \frac{B}{k \log\log k}, \quad k \text{ large},$$

$$c^2(t) = \frac{C}{\log t}, \quad t \text{ large},$$

where A, B, and C are positive constants with $C = B/A$. Further conditions on $U(\cdot)$, $\{\xi_k\}$, and $\{W_k\}$ will be discussed below. It will be useful to define a continuous-time interpolation of $\{X_k\}$. Let

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

and

$$X(t) = X_k, \quad t \in [t_k, t_{k+1}), \quad k \geqq 0.$$

In the sequel we assume some or all of the following conditions (α and β are constants whose values will be specified later):

(A1) $U(\cdot)$ is a C^2 function from $\mathbb{R}^d$ to $[0, \infty)$ such that

$$\min U(x) = 0,$$

$$U(x) \to \infty \quad \text{and} \quad |\nabla U(x)| \to \infty \quad \text{as } |x| \to \infty,$$

$$\inf\,(|\nabla U(x)|^2 - \Delta U(x)) > -\infty.$$

(A2) For $\varepsilon > 0$ let

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

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

(A3) $$\varliminf_{|x| \to \infty} \left\langle \frac{\nabla U(x)}{|\nabla U(x)|}, \frac{x}{|x|} \right\rangle > L(d), \qquad L(d) = \left(\frac{4d-4}{4d-3}\right)^{1/2}.$$

(A4) For $k=0,1,\cdots$ let $\mathcal{F}_k=\sigma(X_0, W_0,\cdots, W_{k-1}, \xi_0,\cdots,\xi_{k-1})$. Let K be a compact subset of $\mathbb{R}^d$. There exists $L>0$ such that

$$E\{|\xi_k|^2|\mathcal{F}_k\}\leqq La_k^\alpha, \quad |E\{\xi_k|\mathcal{F}_k\}|\leqq La_k^\beta, \quad \forall X_k\in K, \quad \text{w.p.1.}$$

W_k is independent of $\mathcal{F}_k$.

We note that the measure 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 [9]. In [7], (A1) and (A2) were needed for the analysis of $Y(t)$ as $t\to\infty$; here we also need (A3) and (A4) for the analysis of X_k as $k\to\infty$. Assumption (A3) asserts that $\nabla U(x)$ has a sufficiently large radial component for $|x|$ large. This condition will be used to extend an escape time estimate for $\{X_k\}$ from a bounded region in the $d=1$ case to the $d>1$ case (see Lemma 4). It may be possible to replace $L(d)$ by 0 in (A3) but we have not been able to do so (except of course for $d=1$). Note that (A3) is implied by (A1) when $d=1$.

For a process $Z(\cdot)$ and a function $f(\cdot)$, let $E_{t_1,z_1}\{f(Z(t))\}$ denote conditional expectation given $Z(t_1)=z_1$, and let $E_{t_1,z_1;t_2,z_2}\{f(Z(t))\}$ denote conditional expectation given $Z(t_1)=z_1$ and $Z(t_2)=z_2$ (more precisely, these are suitable fixed versions of conditional expectations). Also for a measure $\mu(\cdot)$ and a function $f(\cdot)$ let $\mu(f)=\int f\,d\mu$.

In [7] it was shown that there exists a constant C_0 (denoted there by c_0) that plays a critical role in the convergence of $Y(t)$ as $t\to\infty$. C_0 has an interpretation in terms of the action functional [10] for the perturbed dynamical systems

$$dY^\varepsilon(t)=-\nabla U(Y^\varepsilon(t))\,dt+\varepsilon\,dW(t). \tag{2.3}$$

Now for $\phi(\cdot)$ an absolutely continuous function on $\mathbb{R}^d$, the (normalized) action functional for (2.3) is given by

$$I(t,x,y)=\inf_{\substack{\phi(0)=x\\ \phi(t)=y}}\frac{1}{2}\int_0^t|\dot\phi(s)+\nabla U(\phi(s))|^2\,ds.$$

According to [7]

$$C_0=\tfrac{3}{2}\sup_{x,y\in S_0}(V(x,y)-2U(y)),$$

where $V(x,y)=\lim_{t\to\infty}I(t,x,y)$ and S_0 is the set of all the stationary points of $U(\cdot)$, i.e., $S_0=\{x\colon \nabla U(x)=0\}$; see [7] for a further discussion of C_0 including some examples. Here is the Chiang–Hwang–Sheu theorem on the convergence of $Y(t)$ as $t\to\infty$.

THEOREM 1 [7]. *Assume* (A1) *and* (A2) *hold. Then for* $C>C_0$ *and any bounded continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

$$\lim_{t\to\infty}E_{0,y_0}\{f(Y(t))\}=\pi(f) \tag{2.4}$$

uniformly for y_0 *in a compact set.*

Let $K_1\subset\mathbb{R}^d$ and let $\{X_k^{x_0}\}$ denote the solution of (2.1) with $X_0=x_0$. We say that $\{X_k^{x_0}\colon k\geqq 0, x_0\in K_1\}$ is tight if given $\varepsilon>0$ there exists a compact $K_2\subset\mathbb{R}^d$ such that $P_{0,x_0}\{X_k\in K_2\}>1-\varepsilon$ for all $k\geqq 0$ and $x_0\in K_1$. Here is our theorem on the convergence of X_k as $k\to\infty$.

THEOREM 2. *Assume* (A1)–(A4) *hold with* $\alpha>-1$ *and* $\beta>0$. *Also assume that* $\{X_k^{x_0}\colon k\geqq 0, x_0\in K\}$ *is tight for* K *a compact set. Then for* $B/A>C_0$ *and any bounded* *continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

$$\lim_{k\to\infty} E_{0,x_0}\{f(X_k)\} = \pi(f) \tag{2.5}$$

uniformly for x_0 *in a compact set.*

Remark. We specifically separate the question of tightness from convergence in Theorem 2. It is appropriate to do this because sometimes it is convenient to first prove tightness and then to put an algorithm into the form of (2.1) to prove convergence. In § 4, we actually give a condition for tightness of a class of algorithms somewhat more general than (2.1), and then use it in § 5 to prove tightness and ultimately convergence for two example algorithms.

Since π concentrates on S^*, we have, of course, that (2.4) and (2.5) imply $Y(t)\to S^*$ as $t\to\infty$ and $X_k \to S^*$ as $k\to\infty$ in probability, respectively.

The proof of Theorem 2 requires the following two lemmas. Let $\beta(\cdot)$ be defined by

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

Note that $s + s^{2/3} \leqq \beta(s) \leqq s + 2s^{2/3}$ for s large.

LEMMA 1 [7]. *Assume the conditions of Theorem* 1. *Then for any bounded continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

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

uniformly for x in a compact set.

LEMMA 2. *Assume the conditions of Theorem* 2. *Then for any bounded continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

$$\lim_{s\to\infty} (E_{0,x_0;s,x}\{f(X(\beta(s)))\} - E_{s,x}\{f(Y(\beta(s)))\}) = 0$$

uniformly for x_0 *in a compact set and all x.*

Lemma 1 is proved in Lemmas 1–3 of [7]. Lemma 2 is proved in § 3. Note that these lemmas involve approximation on increasingly large time intervals: $\beta(s) - s \geqq s^{2/3} \to \infty$ as $s\to\infty$. We now show how these lemmas may be combined to prove Theorem 2.

Proof of Theorem 2. Since $\beta(s)$ is continuous and $\beta(s)\to\infty$ as $s\to\infty$, it is enough to show that

$$\lim_{s\to\infty} E_{0,x_0}\{f(X(\beta(s)))\} = \pi(f) \tag{2.6}$$

uniformly for x_0 in a compact set. We have for $r > 0$

$$\begin{aligned} &|E_{0,x_0}\{f(X(\beta(s)))\} - \pi(f)| \\ &\leqq \int P_{0,x_0}\{X(s)\in dx\}|E_{0,x_0;s,x}\{f(X(\beta(s)))\} - \pi(f)| \\ &\leqq \int_{|x|\leqq r} P_{0,x_0}\{X(s)\in dx\}|E_{0,x_0;s,x}\{f(X(\beta(s)))\} - \pi(f)| + 2\|f\| P_{0,x_0}\{|X(s)| > r\}. \end{aligned} \tag{2.7}$$

Now by the tightness assumption

$$\sup_{s\geqq 0} P_{0,x_0}\{|X(s)| > r\} \to 0 \quad \text{as } r\to\infty. \tag{2.8}$$

Also by Lemmas 1 and 2 and assumption (A2)

$$
\begin{aligned}
&\sup_{|x|\leqq r} |E_{0,x_0;s,x}\{f(X(\beta(s)))\} - \pi(f)| \\
&\quad \leqq \sup_{|x|\leqq r} |E_{0,x_0;s,x}\{f(X(\beta(s)))\} - E_{s,x}\{f(Y(\beta(s)))\}| \\
&\qquad + \sup_{|x|\leqq r} |E_{s,x}\{f(Y(\beta(s)))\} - \pi^{c(s)}(f)| \\
&\qquad + |\pi^{c(s)}(f) - \pi(f)| \to 0 \quad \text{as } s \to \infty.
\end{aligned}
\tag{2.9}
$$

Combining (2.7)-(2.9) and letting $s \to \infty$ and then $r \to \infty$ gives (2.6) and hence the theorem. □

3. Proof of Lemma 2. Before proceeding with the proof of Lemma 2 we address the following technical issue. Observe that Lemma 2 is not concerned with the joint probability law of $X(\cdot)$ and $Y(\cdot)$. Hence without loss of generality we can and will assume that

$$W_k = a_k^{-1/2}(W(t_{k+1}) - W(t_k)),$$

and that the following assumption holds in place of (A4):

(A4′) For $k = 0, 1, \cdots$ let $\mathcal{F}_k = \sigma(X_0, Y_0, \xi_0, \cdots, \xi_{k-1}, W(s), 0 \leqq s \leqq t_k)$. Let K be a compact subset of $\mathbb{R}^d$. There exists $L > 0$ such that

$$E\{|\xi_k|^2 | \mathcal{F}_k\} \leqq L a_k^{\alpha}, \quad |E\{\xi_k | \mathcal{F}_k\}| \leqq L a_k^{\beta} \quad \forall X_k \in K, \quad \text{w.p.1.}$$

$W(t) - W(t_k)$ is independent of $\mathcal{F}_k$ for $t > t_k$.

It will also be convenient to define

$$c^2(t_k) = \frac{C}{\log \log k}, \quad k \text{ large},$$

and to let $c^2(\cdot)$ be a piecewise linear interpolation of $\{c^2(t_k)\}$. Note that $c^2(t) \sim C/\log t$, and since $C = B/A$ we have $\sqrt{a_k}\, c(t_k) = b_k$.

In the sequel, $c_1, c_2, \cdots$ denote positive constants whose value may change from proof to proof.

The proof of Lemma 2 is based on the following three lemmas. For $s, R > 0$ define the exit times

$$\sigma(s, R) = \inf\{t \geqq s : |X(t)| > R\},$$

$$\tau(s, R) = \inf\{t \geqq s : |Y(t)| > R\}.$$

LEMMA 3 [7, p. 745]. *Assume the conditions of Theorem 1. Then given $r > 0$ there exists $R > r$ such that*

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

uniformly for $|x| \leqq r$.

LEMMA 4. *Assume the conditions of Theorem 2. Then given $r > 0$ there exists $R > r$ such that*

$$\lim_{s \to \infty} P_{0,x_0;s,x}\{\sigma(s, R) > \beta(s)\} = 1$$

uniformly for $|x| \leqq r$ and all x_0.

LEMMA 5. *Assume the conditions of Theorem* 2. *Then for* $0<r<R$

$$\lim_{s\to\infty} E_{0,x_0;s,x}\{|X(\beta(s))-Y(\beta(s))|^2, \sigma(s,R)\wedge\tau(s,R)>\beta(s)\}=0$$

uniformly for $|x|\leqq r$ *and all* x_0.

The proofs of Lemmas 4 and 5 are given below. We now show how these lemmas may be combined to prove Lemma 2.

Proof of Lemma 2. Given $r>0$, choose $R>r$ as in Lemmas 3 and 4. Fix $s>0$ for the moment and let $\sigma=\sigma(s,R)$ and $\tau=\tau(s,R)$. Henceforth assume all quantities are conditioned on $X(0)=x_0$, $X(s)=Y(s)=x$, and $|x|\leqq r$. We have

$$\begin{aligned}&|E(\{f(X(\beta(s)))\}-E\{f(Y(\beta(s)))\}|\\ &\qquad\leqq E\{|f(X(\beta(s)))-f(Y(\beta(s)))|,\sigma\wedge\tau>\beta(s)\}+2\|f\|P\{\sigma\wedge\tau\leqq\beta(s)\}.\end{aligned}\tag{3.1}$$

Now by Lemmas 3 and 4

$$P\{\sigma\wedge\tau\leqq\beta(s)\}\to 0 \quad \text{as } s\to\infty.\tag{3.2}$$

Also, since $f(\cdot)$ is uniformly continuous on a compact, given $\varepsilon>0$ there exists $\delta>0$ such that $|f(u)-f(v)|<\varepsilon$ whenever $|u-v|<\delta$ and $|u|,|v|\leqq R$. Hence using the Chebyshev inequality and Lemma 5

$$\begin{aligned}&E\{|f(X(\beta(s)))-f(Y(\beta(s)))|,\sigma\wedge\tau>\beta(s)\}\\ &\quad\leqq 2\|f\|P\{|X(\beta(s))-Y(\beta(s))|\geqq\delta,\sigma\wedge\tau>\beta(s)\}+\varepsilon\\ &\quad\leqq\frac{2\|f\|}{\delta^2}E\{|X(\beta(s))-Y(\beta(s))|^2,\sigma\wedge\tau>\beta(s)\}+\varepsilon\to\varepsilon \quad \text{as } s\to\infty.\end{aligned}\tag{3.3}$$

Combining (3.1)-(3.3) and letting $s\to\infty$ and then $\varepsilon\to 0$ gives the lemma. □

The proofs of Lemmas 4 and 5 involve comparisons between $X(\cdot)$ and $Y(\cdot)$. Define $\zeta(\cdot,\cdot)$ by

$$Y(t)=Y(s)-(t-s)(\nabla U(Y(s))+\zeta(s,t))+c(s)(W(t)-W(s))$$

for $t\geqq s\geqq 0$. To compare $X(\cdot)$ and $Y(\cdot)$ we will need statistics for $\zeta(\cdot,\cdot)$.

PROPOSITION 1. *For every* $R>0$

$$E_{s,y}\{|\zeta(s,t\wedge\tau(s,R))|^2\}=O(|t-s|)$$

as $t\downarrow s$, *uniformly for* $s\geqq 0$ *and all* y.

Proof. In this proof we can and will assume that $\nabla U(\cdot)$ is a bounded and Lipschitz function on $\mathbb{R}^d$ (since $|Y(u)|\leqq R$ for $s\leqq u\leqq t\wedge\tau(s,R)$ we can modify $U(x)$ for $|x|>R$ without loss of generality). Fix $s\geqq 0$ and let $\tau=\tau(s,R)$. Henceforth assume all quantities are conditioned on $Y(s)=y$. Now for $t\geqq s$ we can write

$$(t-s)\zeta(s,t\wedge\tau)=\int_s^{t\wedge\tau}(\nabla U(Y(u))-\nabla U(Y(s)))\,du-\int_s^{t\wedge\tau}(c(u)-c(s))\,dW(u).\tag{3.4}$$

Let d_1 and d_2 be Lipschitz constants for $\nabla U(\cdot)$, $c(\cdot)$, respectively. Under our assumptions on $\nabla U(\cdot)$ and $c(\cdot)$ it is well known (cf. [11]) that $E\{|Y(u)-Y(s)|^2\}=O(|u-s|)$ as $u\downarrow s$, uniformly for $s\geqq 0$ and all y. Hence

$$\begin{aligned}E\left\{\left|\int_s^{t\wedge\tau}(\nabla U(Y(u))-\nabla U(Y(s)))\,du\right|^2\right\}&\leqq d_1^2E\left\{\left(\int_s^t|Y(u)-Y(s)|\,du\right)^2\right\}\\ &\leqq 2d_1^2(t-s)\int_s^t E\{|Y(u)-Y(s)|^2\}\,du\\ &=O((t-s)^3)\end{aligned}\tag{3.5}$$

and

$$E\left\{\left|\int_s^{t\wedge\tau}(c(u)-c(s))\,dW(u)\right|^2\right\}\leqq\int_s^t(c(u)-c(s))^2\,du \tag{3.6}$$
$$\leqq d_2^2\int_s^t(u-s)^2\,du=O((t-s)^3)$$

as $t\downarrow s$, uniformly for $s\geqq 0$ and all y. The proposition follows from (3.4)-(3.6). □

COROLLARY 1. *Given $R>0$, let $\zeta_k=\zeta(t_k, t_{k+1}\wedge\tau(t_k, R))$. Then there exists $M>0$ such that*

$$E\{|\zeta_k|^2|\mathscr{F}_k\}\leqq Ma_k,\quad |E\{\zeta_k|\mathscr{F}_k\}|\leqq Ma_k^{1/2}\quad \text{w.p.1.}$$

Proof. Observe that ζ_k is $\{Y(t_k), W(t)-W(t_k), t_k<t\leqq t_{k+1}\}$ measurable. Since $Y(t_k)$ is $\mathscr{F}_k$ measurable and $\{W(t)-W(t_k), t_k<t\leqq t_{k+1}\}$ is independent of $\mathscr{F}_k$, we must have $P\{\zeta_k\in\cdot\,|\mathscr{F}_k\}=P\{\zeta_k\in\cdot\,|Y(t_k)\}$ w.p.1. The corollary now follows from Proposition 1 and Holder's inequality. □

3.1. Proof of Lemma 4. The idea behind this proof is to compare $X(t)$ and $Y(t)$ in such a way as to eliminate the slowly decreasing Gaussian noise (i.e., the b_kW_k term) between them. Once the decreasing Gaussian noise is eliminated, we can control the deviation of $X(t)$ from $Y(t)$ over increasingly large time intervals and ultimately obtain the escape time estimate for $X(t)$ from a bounded region from that for $Y(t)$ in Lemma 3. It seems very difficult to work directly with the continuous-time interpolation $X(t)$.

For each n let k_n be the integer that satisfies $\beta(t_n)\in[t_{k_n}, t_{k_{n+1}})$. We show there exists $R>r$ such that

$$\lim_{n\to\infty}P_{0,x_0;t_n,x}\{\sigma(t_n, R)>t_{k_n}\}=1 \tag{3.7}$$

uniformly for $|x|\leqq r$ and all x_0. The lemma then follows by some minor details that are omitted.

By Lemma 3 there exists $R_1>r$ such that

$$\lim_{n\to\infty}P_{t_n,x}\{\tau(t_n, R_1)>t_{k_n}\}=1$$

uniformly for $|x|\leqq r$. Hence (3.7) will follow if we can show that there exists $R>r$ such that

$$\lim_{n\to\infty}P_{0,x_0;t_n,x}\{\sigma(t_n, R)\leqq t_{k_n}, \tau(t_n, R_1)>t_{k_n}\}=0 \tag{3.8}$$

uniformly for $|x|\leqq r$ and all x_0. We first assume $d=1$ (the scalar case) and then generalize to $d>1$. The generalization to $d>1$ requires (A3).

Proof for $d=1$. In view of (A1) there exists $R_2>R_1$ such that

$$\sup_{x\leqq -R_2}U'(x)<\inf_{|x|\leqq R_1}U'(x),\qquad \inf_{x\geqq R_2}U'(x)>\sup_{|x|\leqq R_1}U'(x).$$

Let $R_3=R_2+1$ and $R_4=2R_3+3R_1$. We show that (3.8) holds with $R=R_4$.

Fix n for the moment and let $\sigma=\sigma(t_n, R_4)$, $\tau=\tau(t_n, R_1)$. Let

$$\zeta_k=\zeta(t_k, t_{k+1}\wedge\tau(t_k, R_1))$$

and

$$Y_{k+1}=Y_k-a_k(\nabla U(Y_k)+\zeta_k)+b_kW_k.$$

Note that if $Y(t_n)=Y_n$ and $\tau \geqq t_k \geqq t_n$, then $Y(t_k)=Y_k$. Henceforth assume all quantities are conditioned on $X(0)=X_0=x_0$, $X(t_n)=X_n=Y(t_n)=Y_n=x$, $|x|\leqq r$.

We proceed by observing that if the event $\{\sigma \leqq t_{k_n}\}$ occurs then either

- At some time k, $n\leqq k<k_n$, X_k jumps from $[-R_4, R_2]$ to (R_3, ∞), or from $[-R_2, R_4]$ to $(-\infty, -R_3)$;
- At some time k, $n\leqq k<k_n$, X_k jumps from $[-R_4, R_2]$ to $(R_2, R_3]$, and exits from $(R_2, R_4]$ to (R_4, ∞) at some time l, $k<l\leqq k_n$;
- At some time k, $n\leqq k<k_n$, X_k jumps from $[-R_2, R_4]$ to $[-R_3, -R_2)$, and exits from $[-R_4, -R_2)$ to $(-\infty, -R_4)$ at some time l, $k<l\leqq k_n$.

Now define $\mathscr{F}_k$-stopping times:

$$\mu_1^+ = \inf\{k>n\colon X_{k-1}\leqq R_2,\ R_2<X_k\leqq R_3\},$$

$$\nu_1^+ = \inf\{k>\mu_1^+\colon X_k\leqq R_2\},$$

$$\mu_2^+ = \inf\{k>\nu_1^+\colon X_{k-1}\leqq R_2,\ R_2<X_k\leqq R_3\},$$

$$\nu_2^+ = \inf\{k>\mu_2^+\colon X_k\leqq R_2\},$$

$$\vdots$$

and

$$\mu_1^- = \inf\{k>n\colon X_{k-1}\geqq -R_2,\ -R_3\leqq X_k<-R_2\},$$

$$\nu_1^- = \inf\{k>\mu_1^-\colon X_k\geqq -R_2\},$$

$$\mu_2^- = \inf\{k>\nu_1^-\colon X_{k-1}\geqq -R_2,\ -R_3\leqq X_k<-R_2\},$$

$$\nu_2^- = \inf\{k>\mu_2^-\colon X_k\geqq -R_2\},$$

$$\vdots\ .$$

Note that if $\mu_m^+, \mu_m^- < k_n$, then we must have $m\leqq m_n$ (where $m_n\leqq (k_n-n)/2$). Hence if we let

$$D_n = \bigcup_{k=n}^{k_n-1} \{-R_4\leqq X_k\leqq R_2,\ X_{k+1}>R_3\}\cup\{-R_2\leqq X_k\leqq R_4,\ X_{k+1}<-R_3\},$$

$$E_n^+ = \bigcup_{m=1}^{m_n} \{t_{\mu_m^+}<\sigma<t_{\nu_m^+},\ \sigma\leqq t_{k_n},\ \tau>t_{k_n}\},$$

$$E_n^- = \bigcup_{m=1}^{m_n} \{t_{\mu_m^-}<\sigma<t_{\nu_m^-},\ \sigma\leqq t_{k_n},\ \tau>t_{k_n}\},$$

then

$$P\{\sigma\leqq t_{k_n},\ \tau>t_{k_n}\}\leqq P\{D_n\}+P\{E_n^+\}+P\{E_n^-\}.$$

CLAIM 1. *$\lim_{n\to\infty} P\{D_n\}=0$ uniformly for $|x|\leqq r$ and all x_0.*

CLAIM 2. *$\lim_{n\to\infty} P\{E_n^{\pm}\}=0$ uniformly for $|x|\leqq r$ and all x_0.*

Assuming that Claims 1 and 2 hold, we have $P\{\sigma\leqq t_{k_n}, \tau>t_{k_n}\}\to 0$ as $n\to\infty$. And the convergence is uniform for $|x|\leqq r$ and all x_0. This proves (3.8) and hence Lemma 4 when $d=1$.

Proof of Claim 1. Using the Chebyshev inequality and a standard estimate for the tail probability of a Gaussian random variable we have

$$\begin{aligned}
P\{D_n\} &\leqq \sum_{k=n}^{k_n-1} P\{-R_4 \leqq X_k \leqq R_2,\, X_{k+1}-X_k > R_3-R_2\} \\
&\qquad \cup \{-R_2 \leqq X_k \leqq R_4,\, X_{k+1}-X_k < -(R_3-R_2)\} \\
&\leqq \sum_{k=n}^{k_n-1} P\{|X_k| \leqq R_4,\, |X_{k+1}-X_k| > R_3-R_2\} \\
&= \sum_{k=n}^{k_n-1} P\{|X_k| \leqq R_4,\, |-a_k(U'(X_k)+\xi_k)+b_k W_k| > R_3-R_2\} \\
&\leqq \sum_{k=n}^{k_n-1} \left(P\left\{ |X_k| \leqq R_4,\, a_k|\xi_k| > \frac{R_3-R_2}{3} \right\} + P\left\{ b_k |W_k| > \frac{R_3-R_2}{3} \right\} \right), \quad n \text{ large} \\
&\leqq c_1 \sum_{k=n}^{k_n-1} \left(a_k^2 E\{|\xi_k|^2,\, |X_k| \leqq R_4\} + \exp\left(-\frac{c_2}{b_k^2} \right) \right) \\
&\leqq c_3 \sum_{k=n}^{k_n-1} \left(a_k^{2+\alpha} + \exp\left(-\frac{c_2}{b_k^2} \right) \right) \\
&\leqq c_5 \sum_{k=n}^{\infty} \left(\frac{1}{k^{2+\alpha}} + \exp(-c_4 k) \right) \to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

since $\alpha > -1$. This completes the proof of Claim 1. □

Proof of Claim 2. Since the proofs for E_n^+ and E_n^- are symmetric, we only consider E_n^+. For convenience we suppress the + sign throughout, i.e., $E_n \triangleq E_n^+$, $\mu_m \triangleq \mu_m^+$, $\nu_m \triangleq \nu_m^+$.

For $1 \leqq m \leqq m_n$ let

$$E_{n,m} = \{t_{\mu_m} < \sigma < t_{\nu_m},\, \sigma \leqq t_{k_n},\, \tau > t_{k_n}\}.$$

We have

$$\begin{aligned}
P\{E_{n,m}\} &= P \bigcup_{k=n+2}^{k_n} \{t_{\mu_m} < t_k < t_{\nu_m},\, \sigma = t_k,\, \tau > t_{k_n}\} \\
&= P \bigcup_{k=n+2}^{k_n} \{X_k - Y_k > R_4 - R_1,\, t_{\mu_m} < t_k < t_{\nu_m},\, \sigma = t_k,\, \tau > t_{k_n}\} \\
&\leqq P \bigcup_{k=n+2}^{k_n} \{X_k - Y_k > R_4 - R_1,\, t_{\mu_m} < t_k \leqq t_{\nu_m} \wedge \sigma \wedge \tau\} \\
&= P\left\{ \max_{k:\, t_{\mu_m} < t_k \leqq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} [X_k - Y_k] > R_4 - R_1 \right\} \\
&= P\left\{ \max_{k:\, t_{\mu_m} < t_k \leqq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} \left[X_{\mu_m} - Y_{\mu_m} - \sum_{l=\mu_m}^{k-1} a_l(U'(X_l) - U'(Y_l)) \right.\right. \\
&\qquad \left.\left. - \sum_{l=\mu_m}^{k-1} a_l(\xi_l - \zeta_l) \right] > R_4 - R_1 \right\}.
\end{aligned}$$

Note that the $b_k W_k$ terms have been eliminated at this point; it is here we see the utility of comparing $X(t)$ and $Y(t)$. Now suppose $t_{\mu_m} < t_k \leqq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}$. Then $X_{\mu_m} \in (R_2, R_3]$, $Y_{\mu_m} \in (-R_1, R_1)$, which implies $X_{\mu_m} - Y_{\mu_m} \leqq R_3 + R_1 = (R_4 - R_1)/2$.

Also $X_l \in (R_2, R_4]$, $Y_l \in (-R_1, R_1)$ for all l such that $\mu_m \leqq l < k$, which implies $U'(X_l) - U'(Y_l) > 0$ for all l such that $\mu_m \leqq l < k$. Now let

$$\eta_k = (\xi_k - \zeta_k)\mathbf{1}_{\{|X_k| \leqq R_4\}}.$$

Note that by (A4′) and Corollary 1

$$E\{|\eta_k|^2|\mathscr{F}_k\} \leqq c_1 a_k^{\alpha \wedge 1}, \quad |E\{\eta_k|\mathscr{F}_k\}| \leqq c_1 a_k^{\beta \wedge (1/2)} \quad \text{w.p.1.}$$

Hence

$$\begin{aligned}
P\{E_{n,m}\} &\leqq P\left\{ \max_{k: t_{\mu_m} < t_k \leqq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} \sum_{l=\mu_m}^{k-1} a_l \eta_l > \frac{R_4 - R_1}{2} \right\} \\
&\leqq P\left\{ \max_{\mu_m < k \leqq \nu_m \wedge k_n} \sum_{l=\mu_m}^{k-1} a_l \eta_l > \frac{R_4 - R_1}{2} \right\} \\
&= P\left\{ \max_{n+1 \leqq k \leqq k_n - 1} \sum_{l=n+1}^{k} a_l \eta_l \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}} > \frac{R_4 - R_1}{2} \right\} \qquad (3.9) \\
&\leqq P\left\{ \max_{n+1 \leqq k \leqq k_n - 1} \sum_{l=n+1}^{k} a_l (\eta_l - E\{\eta_l|\mathscr{F}_l\}) \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}} \right. \\
&\qquad \left. + \max_{n+1 \leqq k \leqq k_n - 1} \sum_{l=n+1}^{k} a_l E\{\eta_l|\mathscr{F}_l\} \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}} > \frac{R_4 - R_1}{2} \right\}.
\end{aligned}$$

But

$$\begin{aligned}
&\max_{n+1 \leqq k \leqq k_n - 1} \left| \sum_{l=n+1}^{k} a_l E\{\eta_l|\mathscr{F}_l\} \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}} \right| \\
&\leqq \sum_{l=n+1}^{k_n - 1} a_l |E\{\eta_l|\mathscr{F}_l\}| \qquad (3.10) \\
&\leqq c_1 \sum_{l=n+1}^{k_n - 1} a_l^{(3/2) \wedge (1+\beta)} \\
&\leqq c_2 \sum_{l=n+1}^{\infty} \frac{1}{k^{(3/2) \wedge (1+\beta)}} \to 0 \quad \text{as } n \to \infty
\end{aligned}$$

since $\beta > 0$. Combining (3.9) and (3.10) gives for n large enough

$$P\{E_{n,m}\} \leqq P\left\{ \max_{n+1 \leqq k \leqq k_n - 1} \sum_{l=n+1}^{k} a_l (\eta_l - E\{\eta_l|\mathscr{F}_l\}) \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}} > \frac{R_4 - R_1}{4} \right\}. \qquad (3.11)$$

Let $\tilde{\eta}_k = \eta_k - E\{\eta_k|\mathscr{F}_k\}$ and

$$S_{m,k} = \sum_{l=n+1}^{k} a_l \tilde{\eta}_l \mathbf{1}_{\{\mu_m \leqq l < \nu_m\}}, \qquad k \geqq n+1.$$

Since $\tilde{\eta}_l$ is $\mathscr{F}_{l+1}$-measurable and $\{\mu_m \leqq l < \nu_m\} \in \mathscr{F}_l$, $\{S_{m,k}, \mathscr{F}_{k+1}\}_{k \geqq n+1}$ is a martingale. Hence applying Doob's inequality to (3.11) gives for n large enough

$$\begin{aligned}
P\{E_{n,m}\} &\leqq P\left\{ \max_{n+1 \leqq k \leqq k_n - 1} S_{m,k} > \frac{R_4 - R_1}{4} \right\} \\
&\leqq c_3 E\{S_{m,k_n-1}^2\} \\
&= c_3 \sum_{k=n+1}^{k_n - 1} a_k^2 E\{|\tilde{\eta}_k|^2 \mathbf{1}_{\{\mu_m \leqq k < \nu_m\}}\}.
\end{aligned}$$

Finally,

$$
\begin{aligned}
P\{E_n\} &\leqq \sum_{m=1}^{m_n} P\{E_{n,m}\} \\
&\leqq c_3 \sum_{k=n+1}^{k_n-1} a_k^2 E\left\{|\tilde{\eta}_k|^2 \sum_{m=1}^{m_n} \mathbf{1}_{\{\mu_m \leqq k < \nu_m\}}\right\} \\
&\leqq c_3 \sum_{k=n+1}^{k_n-1} a_k^2 E\{|\tilde{\eta}_k|^2\} \\
&\leqq c_3 \sum_{k=n+1}^{k_n-1} a_k^2 E\{|\eta_k|^2\} \\
&\leqq c_4 \sum_{k=n+1}^{k_n-1} a_k^{3\wedge(2+\alpha)} \\
&\leqq c_5 \sum_{k=n+1}^{\infty} \frac{1}{k^{3\wedge(2+\alpha)}} \to 0 \quad \text{as } n \to \infty
\end{aligned}
$$

since $\alpha > -1$. This completes the proof of Claim 2.

Proof for $d > 1$. We now show how the above proof for $d = 1$ can be extended to $d > 1$.

Let u^i denote the ith component of a vector u. Suppose for the moment that there exists $R_2 > R_1$ such that for $R_3 = R_2 + 1$ and $R_4 = 2R_3 + 3R_1$, we have

$$
\sup_{\substack{x^i \leqq -R_2 \\ |x^j| \leqq R_4 \,\forall j \neq i}} \frac{\partial U}{\partial x^i}(x) < \inf_{|x^j| \leqq R_1 \forall j} \frac{\partial U}{\partial x^i}(x), \tag{3.12}
$$

$$
\inf_{\substack{x^i \geqq R_2 \\ |x^j| \leqq R_4 \,\forall j \neq i}} \frac{\partial U}{\partial x^i}(x) > \sup_{|x^j| \leqq R_1 \,\forall j} \frac{\partial U}{\partial x^i}(x). \tag{3.13}
$$

For $s > 0$, $R_0 > 0$, and $i = 1, \cdots, d$ let

$$
\sigma_i(s, R_0) = \inf\{t \geqq s : |X^i(t)| > R_0\}.
$$

Then we can show that as $n \to \infty$

$$
\begin{aligned}
&P_{0,x_0;t_n,x}\{\sigma(t_n, \sqrt{d}\, R_4) \leqq t_{k_n}, \tau(t_n, R_1) > t_{k_n}\} \\
&\qquad \leqq \sum_{i=1}^{d} P_{0,x_0;t_n,x}\{\sigma_i(t_n, R_4) \leqq t_{k_n}, \sigma_i(t_n, R_4) \leqq \sigma_j(t_n, R_4)\ \forall j \neq i, \tau(t_n, R_1) > t_{k_n}\} \to 0
\end{aligned}
$$

similarly to the proof given above that

$$
P_{0,x_0;t_n,x}\{\sigma(t_n, R_4) \leqq t_{k_n}, \tau(t_n, R_1) > t_{k_n}\} \to 0
$$

in the scalar case $d = 1$. So (3.8) and hence Lemma 4 holds for $R = \sqrt{d}\, R_4$.

It remains to establish (3.12) and (3.13). We only consider (3.13). Let $D(R_2) = \{x : x^i \geqq R_2, |x^j| \leqq R_4 \,\forall j \neq i\}$. Since R_1 is fixed here, there will exist R_2 such that (3.13) holds if we can show

$$
\lim_{R_2 \to \infty} \inf_{x \in D(R_2)} \frac{\partial U}{\partial x^i}(x) = \infty.
$$

We proceed by breaking $\nabla U(x)$ into radial and tangential components and comparing the projection of these components on e^i, the ith standard basis element in $\mathbb{R}^d$. So let

$$\hat{x}=\frac{x}{|x|}, \qquad |x|>0,$$

$$\hat{\theta}=\frac{\nabla U(x)-\langle\nabla U(x),\hat{x}\rangle\hat{x}}{|\nabla U(x)-\langle\nabla U(x),\hat{x}\rangle\hat{x}|}, \qquad |\nabla U(x)-\langle\nabla U(x),\hat{x}\rangle\hat{x}|>0$$

$$=0, \qquad |\nabla U(x)-\langle\nabla U(x),\hat{x}\rangle\hat{x}|=0$$

and

$$g(x)=\frac{\langle\nabla U(x),\hat{\theta}\rangle\langle\hat{\theta},e^i\rangle}{\langle\nabla U(x),\hat{x}\rangle\langle\hat{x},e^i\rangle}, \quad x^i \text{ large.}$$

Of course $\langle\hat{x},\hat{\theta}\rangle=0$. Then

$$\begin{aligned}\overline{\lim_{R_2\to\infty}}\sup_{x\in D(R_2)} g^2(x) &\leqq \overline{\lim_{R_2\to\infty}}\sup_{x\in D(R_2)} \frac{|\nabla U(x)|^2-\langle\nabla U(x),\hat{x}\rangle^2}{\langle\nabla U(x),\hat{x}\rangle^2}\cdot\frac{1-\langle\hat{x},e^i\rangle^2}{\langle\hat{x},e^i\rangle^2}\\ &=\overline{\lim_{R_2\to\infty}}\sup_{x\in D(R_2)}\left[\left\langle\frac{\nabla U(x)}{|\nabla U(x)|},\frac{x}{|x|}\right\rangle^{-2}-1\right]\cdot\frac{|x|^2-(x^i)^2}{(x^i)^2}\\ &<(L(d)^{-2}-1)4(d-1)=1,\end{aligned}$$

where the first inequality follows from Bessel's inequality (applied to $\nabla U(x)$ and to e^i), and the last inequality follows from (A3) and the fact that if $x\in D(R_2)$ then $|x|^2-(x^i)^2\leqq(d-1)R_4^2$ and $(x^i)^2\geqq R_2^2$ (and also $R_4\sim 2R_2$ as $R_2\to\infty$). Hence

$$\begin{aligned}\underline{\lim_{R_2\to\infty}}\inf_{x\in D(R_2)}\frac{\partial U}{\partial x^i}(x) &=\underline{\lim_{R_2\to\infty}}\inf_{x\in D(R_2)}[\langle\nabla U(x),\hat{x}\rangle\langle\hat{x},e^i\rangle+\langle\nabla U(x),\hat{\theta}\rangle\langle\hat{\theta},e^i\rangle]\\ &=\underline{\lim_{R_2\to\infty}}\inf_{x\in D(R_2)}\langle\nabla U(x),\hat{x}\rangle\langle\hat{x},e^i\rangle(1+g(x))\\ &=\underline{\lim_{R_2\to\infty}}\inf_{x\in D(R_2)}|\nabla U(x)|\left\langle\frac{\nabla U(x)}{|\nabla U(x)|},\frac{x}{|x|}\right\rangle\frac{x^i}{|x|}(1+g(x))=\infty.\end{aligned}$$

Hence (3.13) and similarly (3.12) follows. This completes the proof of Lemma 4.

3.2. Proof of Lemma 5. The idea behind this proof is that if $X(s)=Y(s)$ and $X(t)$ and $Y(t)$ remain in a fixed bounded set on large time intervals $t\in[s,\beta(s)]$ (and they do by Lemmas 3 and 4), then we can develop a recursion for estimating $E\{|X(\beta(s))-Y(\beta(s))|^2\}$, and from the recursion we can show that $E\{|X(\beta(s))-Y(\beta(s))|^2\}\to 0$ as $s\to\infty$. This is true even though the interval length $\beta(s)-s\to\infty$ as $s\to\infty$.

For each n let k_n be the integer that satisfies $\beta(t_n)\in[t_{k_n},t_{k_{n+1}})$. We show that

$$\lim_{n\to\infty}E_{0,x_0;t_n,x}\{|X(t_{k_n})-Y(t_{k_n})|^2,\sigma(t_n,R)\wedge\tau(t_n,R)>t_{k_n}\}=0. \tag{3.14}$$

The lemma then follows by some minor details, which are omitted.

In this proof we can and will assume that $\nabla U(\cdot)$ is bounded and Lipschitz function on $\mathbb{R}^d$, and ξ_k satisfies (A4′) with $K=\mathbb{R}^d$ (instead of K a compact subset of $\mathbb{R}^d$), i.e.,

$$E\{|\xi_k|^2|\mathcal{F}_k\}\leqq La_k^\alpha, \quad |E\{\xi_k|\mathcal{F}_k\}|\leqq La_k^\beta \quad \text{w.p.1} \tag{3.15}$$

(if $\sigma(t_n, R) \wedge \tau(t_n, R) > t_{k_n}$ then $|X(t)|$, $|Y(t)| \leqq R$ for $t_n \leqq t \leqq t_{k_n}$, and so $U(x)$ can be modified for $|x| > R$ and we can set $\xi_k = 0$ for $|X_k| > R$ without loss of generality).

Fix n for the moment and let $\sigma = \sigma(t_n, R)$, $\tau = \tau(t_n, R)$. Let

$$\zeta_k = \zeta(t_k, t_{k+1} \wedge \tau(t_k, R))$$

and

$$Y_{k+1} = Y_k - a_k(\nabla U(Y_k) + \zeta_k) + b_k W_k.$$

Note that if $Y(t_n) = Y_n$ and $\tau > t_{k_n}$, then $Y(t_{k_n}) = Y_{k_n}$. Henceforth assume all quantities are conditioned on $X(0) = X_0 = x_0$, $X(t_n) = X_n = Y(t_n) = Y_n = x$, $|x| \leqq r$. Then

$$\begin{aligned} E\{|X(t_{k_n}) - Y(t_{k_n})|^2, \sigma \wedge \tau > t_{k_n}\} &= E\{|X_{k_n} - Y_{k_n}|^2, \sigma \wedge \tau > t_{k_n}\} \\ &\leqq E\{|X_{k_n} - Y_{k_n}|^2\}. \end{aligned} \tag{3.16}$$

We proceed to show that the right side of (3.16) tends to zero as $n \to \infty$. Let

$$\Delta_k = X_k - Y_k, \qquad \eta_k = \xi_k - \zeta_k.$$

Note that by (3.15) and Corollary 1

$$E\{|\eta_k|^2|\mathcal{F}_k\} \leqq c_1 a_k^{\alpha \wedge 1}, \quad |E\{\eta_k|\mathcal{F}_k\}| \leqq c_1 a_k^{\beta \wedge (1/2)} \quad \text{w.p.1.}$$

Now using Holder's inequality and the fact that X_k, Y_k, and hence Δ_k are $\mathcal{F}_k$ measurable we have

$$\begin{aligned} E\{|\Delta_{k+1}|^2\} &= E\{|\Delta_k - a_k(\nabla U(X_k + \Delta_k) - \nabla U(X_k) + \eta_k)|^2\} \\ &= E\{|\Delta_k|^2\} - 2a_k E\{\langle \Delta_k, \nabla U(X_k + \Delta_k) - \nabla U(X_k)\rangle\} \\ &\quad -2a_k E\{\langle \Delta_k, \eta_k\rangle\} + a_k^2 E\{|\nabla U(X_k + \Delta_k) - \nabla U(X_k)|^2\} \\ &\quad + 2a_k^2 E\{\langle \nabla U(X_k + \Delta_k) - \nabla U(X_k), \eta_k\rangle\} + a_k^2 E\{|\eta_k^2|\} \\ &\leqq E\{|\Delta_k|^2\} + 2d_1 a_k E\{|\Delta_k|^2\} \\ &\quad + 2a_k E\{|\Delta_k|^2\}^{1/2} E\{|E\{\eta_k|\mathcal{F}_k\}|^2\}^{1/2} + 2d_1^2 a_k^2 E\{|\Delta_k|^2\} \\ &\quad + 2d_1 a_k^2 E\{|\Delta_k|^2\}^{1/2} E\{|E\{\eta_k|\mathcal{F}_k\}|^2\}^{1/2} + a_k^2 E\{E\{|\eta_k|^2|\mathcal{F}_k\}\} \\ &\leqq (1 + c_2 a_k) E\{|\Delta_k|^2\} + c_2 a_k^\delta, \end{aligned}$$

where d_1 is a Lipschitz constant for $\nabla U(\cdot)$ and $\delta = \min[\frac{3}{2}, 2+\alpha, 1+\beta]$. Using the assumptions that $\alpha > -1$ and $\beta > 0$ we have $\delta > 1$. Now for each n

$$\begin{gathered} E\{|\Delta_{k+1}|^2\} \leqq (1 + c_2 a_k) E\{|\Delta_k|^2\} + c_2 a_k^\delta, \qquad k \geqq n, \\ E\{|\Delta_n|^2\} = 0, \end{gathered}$$

and if we replace the inequality with equality, the resulting difference equation is unstable as $k \to \infty$ (recall that $a_k = A/k$, k large). Nonetheless, we make the following claim.

CLAIM 3. *There exists $\gamma > 1$ such that*

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

Assume the claim holds. Since $t_{k_n} \leqq \beta(t_n) \leqq t_n + 2t_n^{2/3} < \gamma t_n$ for n large, it follows that

$$\lim_{n \to \infty} E\{|\Delta_{k_n}|^2\} = 0.$$

This proves (3.14) and hence Lemma 5. It remains to prove the claim.

Proof of Claim 3. For each n let $\{u_{n,k}\}_{k\geqq n}$ be a sequence of nonnegative numbers such that

$$u_{n,k+1}\leqq(1+a_k)u_{n,k}+a_k^\delta,\qquad k\geqq n,$$
$$u_{n,n}=0,$$

where $\delta>1$. Now

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

since $1+x\leqq e^x$. Also $\sum_n^{k-1}a_m\leqq A(\log(k/n)+1/n)$ and $\sum_n^{k-1}a_m^\delta\leqq A(1/(\delta-1)n^{\delta-1}+1/n^\delta)$, and if $t_k\leqq\gamma t_n$ then $k\leqq c_1n^\gamma$. Choose γ such that $1<\gamma<1+(\delta-1)/A$. It follows that

$$\sup_{k:t_n\leqq t_k\leqq\gamma t_n}u_{n,k}\leqq c_2n^{(\gamma-1)A-(\delta-1)}\to0\quad\text{as }n\to\infty.$$

The claim follows by setting $u_{n,k}=E\{|\Delta_k|^2\}$.

Remark. The proof of Claim 3 does *not* work if $a_k=A/k^\eta$ for any $\eta<1$.

4. General tightness criterion. In this section we consider the tightness of an algorithm of the form

$$X_{k+1}=X_k-a_k(\psi_k(X_k)+\xi_k)+b_kW_k,\qquad k\geqq0 \tag{4.1}$$

where $\{a_k\}$, $\{b_k\}$, $\{\xi_k\}$, and $\{W_k\}$ are defined as in § 2, and $\{\psi_k(x): x\in\mathbb{R}^d\}$ is an $\mathbb{R}^d$-valued random vector field for $k=0,1,\cdots$. We will deal with the following conditions in this section (α, β, γ_1, and γ_2 are constants whose values will be specified later).

(B1) For $k=0,1,\cdots$, let $\mathcal{F}_k=\sigma(X_0,W_0,\cdots,W_{k-1},\xi_0,\cdots,\xi_{k-1})$. There exists $L_1>0$ such that

$$E\{|\xi_k|^2|\mathcal{F}_k\}\leqq L_1a_k^\alpha,\quad|E\{\xi_k|\mathcal{F}_k\}|\leqq L_1a_k^\beta\quad\text{w.p.1}$$

W_k is independent of $\mathcal{F}_k$.

(B2) Let K be a compact subset of $\mathbb{R}^d$. There exists $L_2>0$ such that

$$E\{|\psi_k(x)|^2|\mathcal{F}_k\}\leqq L_2\quad\forall x\in K,\quad\text{w.p.1.}$$

(B3) There exists $L_3,R>0$ such that

$$E\{|\psi_k(x)||\mathcal{F}_k\}^2\geqq L_3\frac{|x|^2}{a_k^{\gamma_1}}\quad\forall|x|>R,\quad\text{w.p.1.}$$

(B4) There exists $L_4,R>0$ such that

$$E\{|\psi_k(x)|^2|\mathcal{F}_k\}\leqq L_4\frac{|x|^2}{a_k^{\gamma_2}}\quad\forall|x|>R,\quad\text{w.p.1.}$$

(B5) There exists $L_5,R>0$ such that

$$E\{\langle\psi_k(x),x\rangle|\mathcal{F}_k\}\geqq L_5E\{|\psi_k(x)||x||\mathcal{F}_k\}\quad\forall|x|>R,\quad\text{w.p.1.}$$

THEOREM 3. *Assume that* (B1)-(B5) *hold with* $\alpha>-1$, $\beta>0$, *and* $0\leqq\gamma_1\leqq\gamma_2<1$. *Let* $\{X_k\}$ *be given by* (4.1) *and* K *be a compact subset of* $\mathbb{R}^d$. *Then* $\{X_k^{x_0}: k\geqq0, x_0\in K\}$ *is a tight family of random variables.*

The proof of Theorem 3 will require the following lemmas.

LEMMA 6. *Assume the conditions of Theorem* 3. *Then there exist an integer* k_0 *and an* $M_1>0$ *such that*

$$E_{0,x_0}\{|X_{k+1}|^2\}-E_{0,x_0}\{|X_k|^2\}\leqq0\quad\text{if }E_{0,x_0}\{|X_k|^2\}\geqq M_1,$$

for $k\geqq k_0$ *and all* x_0.

Proof. Assume all quantities are conditioned on $X_0 = x_0$. Now it follows from (B2)–(B5) and the fact that X_k is $\mathscr{F}_k$-measurable that

$$E\{|\psi_k(X_k)|^2, |X_k| \leqq R\} \leqq L_2,$$

$$E\{|\psi_k(X_k)|^2, |X_k| > R\} \geqq L_3 a_k^{-\gamma_1} E\{|X_k|^2, |X_k| > R\},$$

$$E\{|\psi_k(X_k)|^2, |X_k| > R\} \leqq L_4 a_k^{-\gamma_2} E\{|X_k|^2, |X_k| > R\},$$

$$E\{\langle \psi_k(X_k), X_k\rangle, |X_k| > R\} \geqq L_5 L_3^{1/2} a_k^{-\gamma_1/2} E\{|X_k|^2, |X_k| > R\}.$$

Let $D \in \mathscr{F}_k$. Then using Holder's inequality and the fact that X_k is $\mathscr{F}_k$-measurable and W_k is independent of $\mathscr{F}_k$ we have

$$\begin{aligned}
&E\{|X_{k+1}|^2, D\} - E\{|X_k|^2, D\} \\
&= E\{|X_k - a_k(\psi_k(X_k) + \xi_k) + b_k W_k|^2, D\} - E\{|X_k|^2, D\} \\
&= -2a_k E\{\langle X_k, \psi_k(X_k)\rangle, D\} - 2a_k E\{\langle X_k, \xi_k\rangle, D\} \\
&\quad + 2b_k E\{\langle X_k, W_k\rangle, D\} + a_k^2 E\{|\psi_k(X_k)|^2, D\} \\
&\quad + 2a_k^2 E\{\langle \psi_k(X_k), \xi_k\rangle, D\} - 2a_k b_k E\{\langle \psi_k(X_k), W_k\rangle, D\} \\
&\quad + a_k^2 E\{|\xi_k|^2, D\} - 2a_k b_k E\{\langle \xi_k, W_k\rangle, D\} + b_k^2 E\{|W_k|^2, D\} \\
&\leqq -2a_k E\{\langle X_k, \psi_k(X_k)\rangle, D\} \\
&\quad + 2a_k E\{|X_k|^2, D\}^{1/2} E\{|E\{\xi_k|\mathscr{F}_k\}|^2\}^{1/2} \\
&\quad + 2b_k E\{\langle X_k, E\{W_k\}\rangle, D\} + a_k^2 E\{|\psi_k(X_k)|^2, D\} \\
&\quad + 2a_k^2 E\{|\psi_k(X_k)|^2, D\}^{1/2} E\{E\{|\xi_k|^2|\mathscr{F}_k\}\}^{1/2} \\
&\quad + 2a_k b_k E\{|\psi_k(X_k)|^2, D\}^{1/2} E\{|W_k|^2\}^{1/2} \\
&\quad + a_k^2 E\{E\{|\xi_k|^2|\mathscr{F}_k\}\} \\
&\quad + 2a_k b_k E\{E\{|\xi_k|^2|\mathscr{F}_k\}\}^{1/2} E\{|W_k|^2\}^{1/2} + b_k^2 E\{|W_k|^2\}.
\end{aligned} \tag{4.2}$$

Let $D = \{X_k > R\}$. Then using (4.2) we have

$$\begin{aligned}
&E\{|X_{k+1}|^2, |X_k| > R\} - E\{|X_k|^2, |X_k| > R\} \\
&\leqq -c_1 a_k^{1-\gamma_1/2} E\{|X_k|^2, |X_k| > R\} \\
&\quad + c_2((a_k^{\delta_1} + a_k^{1-\gamma_2/2} b_k) E\{|X_k|^2, |X_k| > R\} + a_k^{\delta_2} + a_k^{\delta_3} b_k + b_k^2),
\end{aligned}$$

where $\delta_1 = \min[1+\beta, 2-\gamma_2, 2+(\alpha-\gamma_2)/2]$, $\delta_2 = \min[1+\beta, 2+(\alpha-\gamma_2)/2, 2+\alpha]$, and $\delta_3 = \min[1-\gamma_2/2, 1+\alpha/2]$. Using the assumptions that $\alpha > -1$, $\beta > 0$, and $0 \leqq \gamma_1 \leqq \gamma_2 < 1$, we have $\delta_1 > 1$, $\delta_2 > 1$, and $\delta_3 > \frac{1}{2}$, and since $b_k = o(a_k^{1/2})$ we get

$$\begin{aligned}
&E\{|X_{k+1}|^2, |X_k| > R\} - E\{|X_k|^2, |X_k| > R\} \\
&\leqq (-c_3 a_k^{1-\gamma_1/2} + o(a_k^{1-\gamma_1/2})) E\{|X_k|^2, |X_k| > R\} + o(a_k^{1-\gamma_1/2}) \\
&\leqq -c_4 a_k^{1-\gamma_1/2}(E\{|X_k|^2\} - R - 1)
\end{aligned} \tag{4.3}$$

for all $k \geqq k_0$, if we choose k_0 large enough.

Let $D = \{X_k \leqq R\}$. Then using (4.2) we have

$$E\{|X_{k+1}|^2, |X_k| \leqq R\} - E\{|X_k|^2, |X_k| \leqq R\} \leqq c_5(a_k^{\delta_4} + a_k^{\delta_5} b_k + b_k^2),$$

where $\delta_4 = \min[1, 1+\beta, 2+\alpha/2, 2+\alpha]$ and $\delta_5 = \min(1, 1+\alpha/2]$. Using the assumptions that $\alpha > -1$ and $\beta > 0$ we have $\delta_4 = 1$ and $\delta_5 > \frac{1}{2}$, and since $b_k = o(a_k^{1/2})$ we get

$$E\{|X_{k+1}|^2, |X_k| \leqq R\} - E\{|X_k|^2, |X_k| \leqq R\} \leqq c_6 a_k \leqq c_6 a_k^{1-\gamma_1/2} \tag{4.4}$$

for all $k \geqq 0$.

Finally, let $M_1 = c_6/c_4 + R + 1$. Then combining (4.3) and (4.4) gives the lemma. □

LEMMA 7. *Assume the conditions of Theorem* 3. *Then there exists an* $M_2 > 0$ *such that*

$$E_{0,x_0}\{|X_{k+1}|^2\} - E_{0,x_0}\{|X_k|^2\} \leqq M_2(E_{0,x_0}\{|X_k|^2\} + 1)$$

for $k \geqq 0$ *and all* x_0.

Proof. Similarly to the proof of Lemma 6 we can show that conditioned on $X_0 = x_0$

$$E\{|X_{k+1}|^2, |X_k| > R\} - E\{|X_k|^2, |X_k| > R\} \leqq c_1 a_k^{1/2}(E\{|X_k|^2\} + 1)$$

and

$$E\{|X_{k+1}|^2, |X_k| \leqq R\} - E\{|X_k|^2, |X_k| \leqq R\} \leqq c_1 a_k.$$

Combining these equations gives the lemma. □

Proof of Theorem 3. Let M_1, M_2, and k_0 be as in Lemmas 6 and 7. By Lemma 7 there exists $c_1 \geqq M_1$ such that

$$E_{0,x_0}\{|X_k|^2\} \leqq c_1, \quad \forall k \leqq k_0, \quad x_0 \in K,$$

and by Lemmas 6 and 7 we also have

$$E_{0,x_0}\{|X_{k+1}|^2\} - E_{0,x_0}\{|X_k|^2\} \leqq 0 \quad \text{if } E_{0,x_0}\{|X_k|^2\} \geqq M_1$$

and

$$E_{0,x_0}\{|X_{k+1}|^2\} - E_{0,x_0}\{|X_k|^2\} \leqq M_2(E_{0,x_0}\{|X_k|^2\} + 1)$$

for $k \geqq k_0$ and all x_0. Let $c_2 = c_1 + M_2(M_1 + 1)$. Then by induction we get

$$E_{0,x_0}\{|X_k|^2\} \leqq c_2 \quad \forall k \geqq 0, \quad x_0 \in K,$$

and the tightness of $\{X_k^{x_0}: k \geqq 0, x_0 \in K\}$ follows from this. □

5. Tightness and convergence for two example algorithms. In this section we apply Theorems 2 and 3 to establish the tightness and ultimately the convergence of two example algorithms. Define $U(\cdot)$, $\{a_k\}$, $\{b_k\}$, $\{\xi_k\}$, and $\{W_k\}$ as in § 2. We will need to consider one or both of the following conditions:

(A5) $\varliminf_{|x|\to\infty} |\nabla U(x)|/|x| > 0$.

(A6) $\varlimsup_{|x|\to\infty} |\nabla U(x)|/|x| < \infty$.

Example 1. Here we consider the following algorithm:

$$X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k, \qquad k \geqq 0. \tag{5.1}$$

THEOREM 4. *Assume* (A1)–(A3), (B1), (A5), and (A6) *hold with* $\alpha > -1, \beta > 0$. *Let* $\{X_k\}$ *be given by* (5.1). *Then for* $B/A > C_0$ *and any bounded continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

$$\lim_{k\to\infty} E_{0,x_0}\{f(X_k)\} = \pi(f)$$

uniformly for x_0 *in a compact set.*

Proof. The assumptions of Theorem 2 and Theorem 3 (with $\psi_k(x) = \nabla U(x)$ and $\gamma_1 = \gamma_2 = 0$) are satisfied. □

Observe that the proof of tightness of $\{X_k^{x_0}\}$ using Theorem 3 requires that (A5) and (A6) hold, i.e., there exists M_1 and M_2 such that

$$M_1|x| \leqq |\nabla U(x)| \leqq M_2|x|, \quad |x| \text{ large.}$$

Intuitively, the upper bound on $|\nabla U(x)|$ is needed to prevent potentially unbounded oscillations of $\{X_k\}$ around the origin. It is possible to modify (5.1) in such a way that only the lower bound on $|\nabla U(x)|$ (i.e., (A5)) but not the upper bound on $|\nabla U(x)|$ (i.e., (A6)) is needed. Since we still want convergence to a global minimum of $U(\cdot)$, which is not known to lie in a specified bounded domain, standard multiplier and projection methods [1] are precluded. The next example gives a modification of (5.1), which has the desired properties.

Example 2. Here we consider the following algorithm:

$$(5.2)\qquad \begin{aligned} X_{k+1} &= X_k - a_k(\nabla U(X_k)+\xi_k) + b_k W_k \quad &&\text{if} \quad |\nabla U(X_k)+\xi_k| \leqq \frac{|X_k| \vee 1}{a_k^\gamma} \\ &= X_k - a_k^{1-\gamma} X_k + b_k W_k &&\text{if} \quad |\nabla U(X_k)+\xi_k| > \frac{|X_k| \vee 1}{a_k^\gamma}, \end{aligned}$$

where $\gamma > 0$. Intuitively, note that if K is a fixed compact set, $X_k \in K$, ξ_k is not too large, and k is very large, then X_k is updated to X_{k+1} as in (5.1). Also note that in (5.2) (like (5.1)), $\nabla U(X_k)$ and ξ_k only appear as the sum $\nabla U(X_k)+\xi_k$. This means that we can use noisy or imprecise measurements of $\nabla U(\cdot)$ in (5.2) in exactly the same way as in (5.1).

THEOREM 5. *Assume* (A1)-(A3), (B1), *and* (A5) (*but not necessarily* (A6)) *hold with* $\alpha > 0$. *Let* $\{X_k\}$ *be given by* (5.2) *with* $0 < \gamma < \frac{1}{2}$. *Then for* $B/A > C_0$ *and any bounded continuous function* $f(\cdot)$ *on* $\mathbb{R}^d$

$$(5.3)\qquad \lim_{k\to\infty} E_{0,x_0}\{f(X_k)\} = \pi(f)$$

uniformly for x_0 *in a compact set.*

Proof. Let

$$(5.4)\qquad X_{k+1} = X_k - a_k(\nabla U(X_k)+\xi'_k) + b_k W_k$$

(this defines ξ'_k) and $\mathcal{F}'_k = \sigma(X_0, \xi'_0, \cdots, \xi'_{k-1}, W_0, \cdots, W_{k-1})$. We show that $(\xi'_k, W_k, \mathcal{F}'_k)$ satisfies (A4). Hence by Theorem 2 if $\{X_k^{x_0}: k \geqq 0, x_0 \in K\}$ is tight for K compact then (5.3) holds.

Let

$$\begin{aligned} \psi_k(x) &= \nabla U(x) \quad &&\text{if} \quad |\nabla U(x)+\xi_k| \leqq \frac{|x| \vee 1}{a_k^\gamma} \\ &= \frac{x}{a_k^\gamma} &&\text{if} \quad |\nabla U(x)+\xi_k| > \frac{|x| \vee 1}{a_k^\gamma}. \end{aligned}$$

Let

$$(5.5)\qquad X_{k+1} = X_k - a_k(\psi_k(X_k)+\xi''_k) + b_k W_k$$

(this defines ξ''_k) and $\mathcal{F}''_k = \sigma(X_0, \xi''_0, \cdots, \xi''_{k-1}, W_0, \cdots, W_{k-1})$. We show that $(\xi''_k, W_k, \mathcal{F}''_k)$ satisfies (B1) and $(\psi_k(x), \mathcal{F}''_k)$ satisfies (B2)-(B5) with $\gamma_1 = 0$, $\gamma_2 = 2\gamma$. Hence by Theorem 3 $\{X_k^{x_0}: k \geqq 0, x_0 \in K\}$ is tight for K compact and (5.3) does hold. These assertions are proved in Claims 4 and 5 below.

Remark. The proof shows the importance of separating the tightness and convergence issues. Even though we can write algorithm (5.2) in the form of algorithm (5.4), we cannot apply Theorem 3 to (5.4) to prove tightness because $U(\cdot)$ may not satisfy (A6), and ξ'_k may not satisfy (B1) even though ξ_k satisfies (B1).

CLAIM 4. *Let K be a compact subset of $\mathbb{R}^d$. Then there exists $M_1 > 0$ such that*

$$E\{|\xi'_k|^2|\mathcal{F}'_k\} \leqq M_1 a_k^\alpha \quad \forall X_k \in K, \quad \textit{w.p.1.}$$

Also, W_k is independent of $\mathcal{F}'_k$.

Proof. Clearly,

$$\begin{aligned} \xi'_k &= \xi_k && \text{if} \quad |\nabla U(X_k) + \xi_k| \leqq \frac{|X_k| \vee 1}{a_k^\gamma} \\ &= \frac{X_k}{a_k^\gamma} - \nabla U(X_k) && \text{if} \quad |\nabla U(X_k) + \xi_k| > \frac{|X_k| \vee 1}{a_k^\gamma}. \end{aligned}$$

Hence for $X_k \in K$ and k large enough

$$\begin{aligned} E\{|\xi'_k|^2|\mathcal{F}_k\} &\leqq E\{|\xi_k|^2|\mathcal{F}_k\} + E\left\{\left|\frac{X_k}{a_k^\gamma} - \nabla U(X_k)\right|^2, |\nabla U(X_k) + \xi_k| > \frac{|X_k| \vee 1}{a_k^\gamma} \,\middle|\, \mathcal{F}_k\right\} \\ &\leqq L_1 a_k^\alpha + \frac{c_1}{a_k^{2\gamma}} \Pr\left\{|\nabla U(X_k) + \xi_k| > \frac{|X_k| \vee 1}{a_k^\gamma} \,\middle|\, \mathcal{F}_k\right\} \\ &\leqq L_1 a_k^\alpha + \frac{c_1}{a_k^{2\gamma}} \Pr\left\{|\xi_k| > \frac{c_2}{a_k^\gamma} \,\middle|\, \mathcal{F}_k\right\} \\ &\leqq L_1 a_k^\alpha + c_3 E\{|\xi_k|^2|\mathcal{F}_k\} \leqq M_1 a_k^\alpha \quad \text{w.p.1,} \end{aligned}$$

where we have used the assumption that $\gamma > 0$ and the Chebyshev inequality. It is easy to see that the inequality actually holds for all $k \geqq 0$. Since $\mathcal{F}'_k \subset \mathcal{F}_k$, the claim follows.

CLAIM 5. *Let K be a compact subset of $\mathbb{R}^d$. Then there exists M_1, M_2, M_3, M_4, and M_5, $R > 0$ such that*

(i) $E\{|\xi''_k|^2|\mathcal{F}''_k\} \leqq M_1 a_k^\alpha$ *w.p.1. Also W_k is independent of $\mathcal{F}''_k$,*

(ii) $E\{|\psi_k(x)|^2|\mathcal{F}''_k\} \leqq M_2$ *for all $x \in K$, w.p.1,*

(iii) $E\{|\psi_k(x)||\mathcal{F}''_k\}^2 \geqq M_3|x|^2$ *for all $|x| > R$, w.p.1,*

(iv) $E\{|\psi_k(x)|^2|\mathcal{F}''_k\} \leqq M_4(|x|^2/a_k^{2\gamma})$ *for all $|x| > R$, w.p.1,*

(v) $E\{\langle\psi_k(x), x\rangle|\mathcal{F}''_k\} \geqq M_5 E\{|\psi_k(x)||x||\mathcal{F}''_k\}$ *for all $|x| > R$, w.p.1.*

Proof. First observe that (iii) and (v) follow immediately from (A3) and (A5).

(i) Clearly,

$$\begin{aligned} \xi''_k &= \xi_k \quad \text{if} \quad |\nabla U(X_k) + \xi_k| \leqq \frac{|X_k| \vee 1}{a_k^\gamma} \\ &= 0 \quad \text{if} \quad |\nabla U(X_k) + \xi_k| > \frac{|X_k| \vee 1}{a_k^\gamma}. \end{aligned}$$

Hence

$$E\{|\xi''_k|^2|\mathcal{F}_k\} \leqq E\{|\xi_k|^2|\mathcal{F}_k\} \leqq M_1 a_k^\alpha \quad \text{w.p.1.}$$

Since $\mathcal{F}''_k \subset \mathcal{F}_k$, (i) must hold.

(ii) For $x \in K$ and k large enough

$$E\{|\psi_k(x)|^2|\mathscr{F}_k\} \leqq |\nabla U(x)|^2 + \frac{|x|^2}{a_k^{2\gamma}} \Pr\left\{|\nabla U(x)+\xi_k| > \frac{|x| \vee 1}{a_k^\gamma} \middle| \mathscr{F}_k\right\}$$

$$\leqq c_1 + \frac{c_1}{a_k^{2\gamma}} \Pr\left\{|\xi_k| > \frac{c_2}{a_k^\gamma} \middle| \mathscr{F}_k\right\}$$

$$\leqq c_1 + c_3 E\{|\xi_k|^2|\mathscr{F}_k\} \leqq M_2 \quad \text{w.p.1},$$

where we have used the assumption that $\gamma > 0$ and the Chebyshev inequality. It is easy to see that the inequality actually holds for $x \in K$ and all $k \geqq 0$. Since $\mathscr{F}_k'' \subset \mathscr{F}_k$, (ii) must hold.

(iv) For $|x|$ large enough and $k \geqq 0$

$$E\{|\psi_k(x)|^2|\mathscr{F}_k\} \leqq \frac{|x|^2}{a_k^{2\gamma}} + |\nabla U(x)|^2 P\left\{|\nabla U(x)+\xi_k| \leqq \frac{|x| \vee 1}{a_k^\gamma} \middle| \mathscr{F}_k\right\}$$

$$\leqq \frac{|x|^2}{a_k^{2\gamma}} + E\left\{|\nabla U(x)|^2, |\nabla U(x)| \leqq \frac{|x|}{a_k^\gamma} + |\xi_k| \middle| \mathscr{F}_k\right\}$$

$$\leqq 4\frac{|x|^2}{a_k^{2\gamma}} + 3E\{|\xi_k|^2|\mathscr{F}_k\} \leqq M_4 \frac{|x|^2}{a_k^{2\gamma}} \quad \text{w.p.1}.$$

Since $\mathscr{F}_k'' \subset \mathscr{F}_k$, (iv) must hold. This completes the proof of the claim and hence the theorem. □

As a final note observe that the algorithm (5.1) does require (A6), and also (B1) with $\alpha > -1$, $\beta > 0$. On the other hand, the algorithm (5.2) does not require (A6), but does require (B1) with $\alpha > 0$ (and hence $\beta > 0$ by Holder's inequality). It may be possible to allow $\{\xi_k\}$ with unbounded variance in (5.2) but this would require some additional assumptions on $\{\xi_k\}$ and we do not pursue this.

REFERENCES

[1] H. J. KUSHNER AND D. CLARK, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Applied Mathematical Science Series 26, Springer-Verlag, Berlin, 1978.

[2] H. J. KUSHNER, *Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo*, SIAM J. Appl. Math., 47 (1987), pp. 169–185.

[3] L. LJUNG, *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Control, 22 (1977), pp. 551–575.

[4] U. GRENENDER, *Tutorial in pattern theory*, Division of Applied Mathematics, Brown University, Providence, RI, 1984.

[5] S. GEMAN AND C. R. HWANG, *Diffusions for global optimization*, SIAM J. Control Optim., 24 (1986), pp. 1031–1043.

[6] B. GIDAS, *Global optimization via the Langevin equation*, in Proc. IEEE Conference on Decision and Control, Fort Lauderdale, FL, 1985, pp. 774–778.

[7] T. S. CHIANG, C. R. HWANG, AND S. J. SHEU, *Diffusion for global optimization in* $\mathbb{R}^n$, SIAM J. Control Optim., 25 (1987), pp. 737–752.

[8] L. C. W. DIXON AND G. P. SZEGO, *Towards Global Optimization*, North-Holland, Amsterdam, 1978.

[9] C.-R. HWANG, *Laplace's method revisited: weak convergence of probability measures*, Ann. Probab., 8 (1980), pp. 1177–1182.

[10] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin, New York, 1984.

[11] J. L. DOOB, *Stochastic Processes*, John Wiley, New York, 1952.