

Recursive Stochastic Algorithms for Global Optimization in \mathbb{R}^d

by

Saul B. Gelfand¹ and Sanjoy K. Mitter²

Abstract

We consider an algorithm of the form $X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k$, 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. An algorithm of this type arises by adding slowly decreasing white Gaussian noise to a stochastic gradient algorithm. We show 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. Our 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.

¹ Computer Vision and Image Processing Laboratory, School of Electrical Engineering, Purdue University, West Lafayette, IN 47907

² Department of Electrical Engineering and Computer Science, and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. Research partially supported by the Air Force Office of Scientific Research under contract 89-0276 and by the Army Research Office under contract DAAL03-86-K-0171 (Center for Intelligent Control Systems).

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 descent algorithm

$$X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k \quad (1.1)$$

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

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

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 $\nabla U(\cdot)$ or $U(\cdot)$ (c.f. [1]). The asymptotic behavior of $\{Z_k\}$ has been much 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 $Z_k \rightarrow S$ as $k \rightarrow \infty$ w.p.1. However, in general $Z_k \not\rightarrow S^*$ (unless of course $S = S^*$). The idea behind the additional $b_k W_k$ 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 $U(\cdot)$. We shall in fact 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 which depends only on $U(\cdot)$), and $\{X_k\}$ is tight, then $X_k \rightarrow S^*$ as $k \rightarrow \infty$ in probability. We also give a condition for the tightness of $\{X_k\}$.

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 [2] include conditions on $\{a_k\}$, $\{b_k\}$ and $\{\xi_k\}$, and also the method of analysis; these are discussed further below.

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

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

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

$$dY(t) = -\nabla U(Y(t))dt + c(t)dW(t) \quad (1.4)$$

where $W(\cdot)$ is a standard d -dimensional Wiener process and $c(\cdot)$ is a positive function with $c(t) \rightarrow 0$ as $t \rightarrow \infty$. 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], [5] for global optimization problems that arise in image processing applications with continuous grey levels. Now the asymptotic behavior of $Y(t)$ as $t \rightarrow \infty$ has been studied intensively by a number of researchers. In [2], [5] convergence results were obtained by considering a version of (1.4) with a reflecting boundary; in [6], [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 $Y(t) \rightarrow S^*$ at $t \rightarrow \infty$ 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 noises.

We 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 $X_k \not\rightarrow S^*$ 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 actually 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. Also the analysis in [2] is different from what is done here, in that we obtain the asymptotic behavior of X_k as $k \rightarrow \infty$ from the corresponding behavior of $Y(t)$ as $t \rightarrow \infty$. However, the most significant difference between our work and what is done in [2] (and more generally in other work on global optimization such as [8]) is that we establish the convergence of an algorithm which finds a global minimum of a function

when it is not known apriori what bounded region contains such a point.

The paper is organized as follows. In Section 2 we state our assumptions and main result. In Sections 3 we take up the proof of this result. In Section 4 we prove a general tightness criterion, which is then used in Section 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

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

which is closely related to the continuous time algorithm

$$dY(t) = -\nabla U(Y(t))dt + c(t)dW(t), \quad t \geq 0. \quad (2.2)$$

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

$$t_k = \sum_{n=0}^{k-1} a_n, \quad k \geq 0,$$

and

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

In the sequel we shall assume some or all of the following conditions (α, 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

$$\min U(\mathbf{x}) = 0$$

$$U(\mathbf{x}) \rightarrow \infty \text{ as } |\mathbf{x}| \rightarrow \infty$$

$$|\nabla U(\mathbf{x})| \rightarrow \infty \text{ as } |\mathbf{x}| \rightarrow \infty$$

$$\lim_{|\mathbf{x}| \rightarrow \infty} (|\nabla U(\mathbf{x})|^2 - \Delta U(\mathbf{x})) > -\infty$$

(A2) For $\epsilon > 0$ let

$$d\pi^\epsilon(\mathbf{x}) = \frac{1}{Z^\epsilon} \exp\left(-\frac{2U(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}, \quad Z^\epsilon = \int \exp\left(-\frac{2U(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} < \infty$$

π^ϵ has a unique weak limit π as $\epsilon \rightarrow 0$

(A3) $\lim_{|\mathbf{x}| \rightarrow \infty} \left\langle \frac{\nabla U(\mathbf{x})}{|\nabla U(\mathbf{x})|}, \frac{\mathbf{x}}{|\mathbf{x}|} \right\rangle > L(d), \quad L(d) = \left(\frac{4d-4}{4d-3}\right)^{1/2}$

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

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

W_k is independent of \mathcal{F}_k .

We note that the measure π concentrates on S^* , the global minima of $U(\cdot)$. For example, if S^* consists of a finite number of points, then π exists and is uniformly distributed over S^* . The existence of π and a simple characterization in terms of the Hessian of $U(\cdot)$ is discussed in [9]. In [7] (A1), (A2) were needed for the analysis of $Y(t)$ as $t \rightarrow \infty$; here we also need (A3), (A4) for the analysis of X_k as $k \rightarrow \infty$. (A3) asserts that $\nabla U(\mathbf{x})$ has a sufficiently large radial component for $|\mathbf{x}|$ 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) which plays a critical role in the convergence of $Y(t)$ as $t \rightarrow \infty$. C_0 has an interpretation in terms of the action functional for the dynamical system $\dot{z}(t) = -\nabla U(z(t))$; see [7] for an explicit expression for C_0 and some examples. Here is Chiang et. al.'s theorem on the convergence of $Y(t)$ as $t \rightarrow \infty$.

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

$$\lim_{t \rightarrow \infty} E_{0, y_0} \{f(Y(t))\} = \pi(f) \quad (2.3)$$

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 shall say that $\{X_k^{x_0} : k \geq 0, x_0 \in K_1\}$ is tight if given $\epsilon > 0$ there exists a compact $K_2 \subset \mathbb{R}^d$ such that $P_{0, x_0} \{X_k \in K_2\} > 1 - \epsilon$ for all $k \geq 0$ and $x_0 \in K_1$. Here is our theorem on the convergence of X_k as $k \rightarrow \infty$.

Theorem 2: Assume (A1), (A2), (A3), (A4) hold with $\alpha > -1$ and $\beta > 0$. Also assume that $\{X_k^{x_0} : k \geq 0, x_0 \in K\}$ is tight for K a compact set. Then for $B/A > C_0$ and any bounded continuous function $f(\cdot)$ on \mathbb{R}^d

$$\lim_{k \rightarrow \infty} E_{0, x_0} \{f(X_k)\} = \pi(f) \quad (2.4)$$

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; an example is given in Section 5. In Section 5 we actually give a condition for tightness of a class of algorithms somewhat more general than (2.1), and then use it to prove

tightness and ultimately convergence for two example algorithms.

Since π concentrates on S^* , we have, of course, that (2.3) and (2.4) imply $Y(t) \rightarrow S^*$ as $t \rightarrow \infty$ and $X_k \rightarrow S^*$ as $k \rightarrow \infty$ in probability, respectively.

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

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

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

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

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

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

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

uniformly for x_0 in a compact set and all x .

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

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

$$\lim_{s \rightarrow \infty} E_{0,x_0} \{f(X(\beta(s)))\} = \pi(f) \tag{2.5}$$

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

$$\begin{aligned}
& |E_{0,x_0} \{f(X(\beta(s)))\} - \pi(f) | \\
& \leq \int P_{0,x_0} \{X(s) \in dx\} |E_{0,x_0;s,x} \{f(X(\beta(s)))\} - \pi(f) | \\
& \leq \int_{|x| \leq r} P_{0,x_0} \{X(s) \in dx\} |E_{0,x_0;s,x} \{f(X(\beta(s)))\} - \pi(f) | \\
& \quad + 2 \|f\| P_{0,x_0} \{ |X(s) | > r \}
\end{aligned} \tag{2.6}$$

Now by the tightness assumption

$$\sup_{s \geq 0} P_{0,x_0} \{ |X(s) | > r \} \rightarrow 0 \quad \text{as } r \rightarrow \infty . \tag{2.7}$$

Also by Lemmas 1, 2 and assumption (A2)

$$\begin{aligned}
& \sup_{|x| \leq r} |E_{0,x_0;s,x} \{f(X(\beta(s)))\} - \pi(f) | \\
& \leq \sup_{|x| \leq r} |E_{0,x_0;s,x} \{f(X(\beta(s)))\} - E_{s,x} \{f(Y(\beta(s)))\} | \\
& \quad + \sup_{|x| \leq r} |E_{s,x} \{f(Y(\beta(s)))\} - \pi^{c(s)}(f) | \\
& \quad + |\pi^{c(s)}(f) - \pi(f) | \rightarrow 0 \quad \text{as } s \rightarrow \infty .
\end{aligned} \tag{2.8}$$

Combining (2.6) - (2.8) and letting $s \rightarrow \infty$ and then $r \rightarrow \infty$ gives (2.5) 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 $k = 0, 1, \dots$ let $\mathcal{F}_k = \sigma(X_0, Y_0, \xi_0, \dots, \xi_{k-1}, W(s), 0 \leq s \leq t_k)$. Let K be a compact subset of \mathbb{R}^d . There exists $L > 0$ such that

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

$W(t) - W(t_k)$ 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, \dots 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

$$\begin{aligned} \sigma(s, R) &= \inf\{t \geq s : |X(t)| > R\} \\ \tau(s, R) &= \inf\{t \geq s : |Y(t)| > R\} \end{aligned}$$

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

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

uniformly for $|x| \leq r$.

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

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

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

Lemma 5: Assume the conditions of Theorem 2. Then for $0 < r < R$

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

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

The proofs of Lemmas 4, 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 3,4. Fix $s > 0$ for the moment and let $\sigma = \sigma(s,R)$, $\tau = \tau(s,R)$. Henceforth assume all quantities are conditioned on $X(0) = x_0$, $X(s) = Y(s) = x$, $|x| \leq r$. We have

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

Now by Lemmas 3, 4

$$P\{\sigma \wedge \tau \leq \beta(s)\} \rightarrow 0 \quad \text{as } s \rightarrow \infty \tag{3.2}$$

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

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

Combining (3.1)-(3.3) and letting $s \rightarrow \infty$ and then $\epsilon \rightarrow 0$ gives the Lemma. □

The proofs of Lemmas 4, 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 $t \geq s \geq 0$. 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 $s \geq 0$ and all y .

Proof: In this proof we can and will assume that $\nabla U(\cdot)$ is bounded and Lipschitz (note that $U(\cdot)$ is C^2 and $|Y(u)| \leq R$ for $s \leq u \leq t \wedge \tau(s,R)$). Fix $s \geq 0$ and let $\tau = \tau(s,R)$. Henceforth assume all quantities are conditioned on $Y(s) = y$. Now for $t \geq s$ we can write

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

Let d_1, d_2 be Lipschitz constants for $\nabla U(\cdot), c(\cdot)$, respectively. Then

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

and

$$\begin{aligned} E \left\{ \left| \int_s^{t \wedge \tau} (c(u) - c(s)) dW(u) \right|^2 \right\} &\leq \int_s^t |c(u) - c(s)|^2 du \\ &\leq d_2^2 \int_s^t (u-s)^2 du = O((t-s)^3) \end{aligned} \quad (3.6)$$

as $t \downarrow s$, uniformly for $s \geq 0$ 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

$$E \{ |\zeta_k|^2 | \mathcal{F}_k \} \leq M a_k, \quad |E \{ \zeta_k | \mathcal{F}_k \}| \leq M a_k^{1/2} \quad \text{w.p.1.}$$

Proof: Observe that ζ_k is $\{Y(t_k), W(t) - W(t_k), t_k < t \leq t_{k+1}\}$ measurable. Since $Y(t_k)$ is \mathcal{F}_k measurable and $\{W(t) - W(t_k), t_k < t \leq t_{k+1}\}$ is independent of \mathcal{F}_k , we must have $P\{\zeta_k \in \cdot \mid \mathcal{F}_k\} = P\{\zeta_k \in \cdot \mid Y(t_k)\}$ 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 $b_k W_k$ term) between them. Then the escape time estimate for $X(t)$ from a bounded region essentially follows from that for $Y(t)$ in Lemma 3. It seems very difficult to work directly with $X(t)$.

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

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

uniformly for $|x| \leq r$ and all x_0 . The Lemma then follows by some minor details which are omitted.

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

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

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

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

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

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

$$\sup_{x \leq -R_2} U'(x) < \inf_{|x| \leq R_1} U'(x),$$

$$\inf_{x \geq R_2} U'(x) > \sup_{|x| \leq R_1} U'(x).$$

Let $R_3 = R_2 + 1$ and $R_4 = 2R_3 + 3R_1$. We shall 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

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

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

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

- at some time k , $n \leq k < k_n$, 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 , $n \leq k < k_n$, 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 , $k < \ell \leq k_n$
- at some time k , $n \leq k < k_n$, 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 , $k < \ell \leq k_n$

Now define \mathcal{F}_k -stopping times

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

$$\nu_1^+ = \inf\{k > \mu_1^+ : X_k \leq R_2\}$$

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

$$\nu_2^+ = \inf\{k > \mu_2^+ : X_k \leq R_2\}$$

⋮

and

$$\begin{aligned}
 \mu_1^- &= \inf\{k > n : X_{k-1} \geq -R_2, -R_3 \leq X_k < -R_2\} \\
 \nu_1^- &= \inf\{k > \mu_1^- : X_k \geq -R_2\} \\
 \mu_2^- &= \inf\{k > \nu_1^- : X_{k-1} \geq -R_2, -R_3 \leq X_k < -R_2\} \\
 \nu_2^- &= \inf\{k > \mu_2^- : X_k \geq -R_2\} \\
 &\vdots
 \end{aligned}$$

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

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

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

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

then

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

Claim 1: $\lim_{n \rightarrow \infty} P\{D_n\} = 0$ uniformly for $|x| \leq r$ and all x_0 .

Claim 2: $\lim_{n \rightarrow \infty} P\{E_n^\pm\} = 0$ uniformly for $|x| \leq r$ and all x_0 .

Assuming that Claims 1, 2 hold, we have $P\{\sigma \leq t_{k_n}, \tau > t_{k_n}\} \rightarrow 0$ as $n \rightarrow \infty$. And the convergence is uniform for $|x| \leq r$ 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

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

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 shall 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 $1 \leq m \leq m_n$ let

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

We have

$$\begin{aligned}
P\{E_{n,m}\} &= P \bigcup_{k=n+2}^{k_n} \{t_{\mu_m} < t_k < t_{\nu_m}, \sigma = t_k, \tau > t_{k_n}\} \\
&= P \bigcup_{k=n+2}^{k_n} \{X_k - Y_k > R_4 - R_1, t_{\mu_m} < t_k < t_{\nu_m}, \sigma = t_k, \tau > t_{k_n}\} \\
&\leq P \bigcup_{k=n+2}^{k_n} \{X_k - Y_k > R_4 - R_1, t_{\mu_m} < t_k \leq t_{\nu_m} \wedge \sigma \wedge \tau\} \\
&= P \left\{ \max_{k: t_{\mu_m} < t_k \leq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} [X_k - Y_k] > R_4 - R_1 \right\} \\
&= P \left\{ \max_{k: t_{\mu_m} < t_k \leq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} [X_{\mu_m} - Y_{\mu_m} - \sum_{\ell=\mu_m}^{k-1} a_\ell (U'(X_\ell) - U'(Y_\ell)) \right. \\
&\quad \left. - \sum_{\ell=\mu_m}^{k-1} a_\ell (\xi_\ell - \zeta_\ell)] > R_4 - R_1 \right\}
\end{aligned}$$

Suppose $t_{\mu_m} < t_k \leq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}$. Then $X_{\mu_m} \in (R_2, R_3]$, $Y_{\mu_m} \in (-R_1, R_1)$, which implies $X_{\mu_m} - Y_{\mu_m} \leq R_3 + R_1 = (R_4 - R_1)/2$. Also $X_\ell \in (R_2, R_4]$, $Y_\ell \in (-R_1, R_1)$ for all ℓ such that $\mu_m \leq \ell < k$, which implies $U'(X_\ell) - U'(Y_\ell) > 0$ for all ℓ such that $\mu_m \leq \ell < k$. Now let

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

Note that by (A4') and Corollary 1

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

Hence

$$\begin{aligned}
P\{E_{n,m}\} &\leq P\left\{\max_{k:t_{\mu_m} < t_k \leq t_{\nu_m} \wedge \sigma \wedge \tau \wedge t_{k_n}} \sum_{\ell=\mu_m}^{k-1} a_\ell \eta_\ell > (R_4 - R_1)/2\right\} \\
&\leq P\left\{\max_{\mu_m < k \leq \nu_m \wedge k_n} \sum_{\ell=\mu_m}^{k-1} a_\ell \eta_\ell > (R_4 - R_1)/2\right\} \\
&= P\left\{\max_{n+1 \leq k \leq k_n-1} \sum_{\ell=n+1}^k a_\ell \eta_\ell 1_{\{\mu_m \leq \ell < \nu_m\}} > (R_4 - R_1)/2\right\} \\
&\leq P\left\{\max_{n+1 \leq k \leq k_n-1} \sum_{\ell=n+1}^k a_\ell (\eta_\ell - E\{\eta_\ell | \mathcal{F}_\ell\}) 1_{\{\mu_m \leq \ell < \nu_m\}}\right. \\
&\quad \left. + \max_{n+1 \leq k \leq k_n-1} \sum_{\ell=n+1}^k a_\ell E\{\eta_\ell | \mathcal{F}_\ell\} 1_{\{\mu_m \leq \ell < \nu_m\}} > (R_4 - R_1)/2\right\} \quad (3.9)
\end{aligned}$$

But

$$\begin{aligned}
&\max_{n+1 \leq k \leq k_n-1} \left| \sum_{\ell=n+1}^k a_\ell E\{\eta_\ell | \mathcal{F}_\ell\} 1_{\{\mu_m \leq \ell < \nu_m\}} \right| \\
&\leq \sum_{\ell=n+1}^{k_n-1} a_\ell |E\{\eta_\ell | \mathcal{F}_\ell\}| \\
&\leq c_1 \sum_{\ell=n+1}^{k_n-1} a_\ell^{(3/2) \wedge (1+\beta)} \\
&\leq c_2 \sum_{\ell=n+1}^{\infty} \frac{1}{k^{(3/2) \wedge (1+\beta)}} \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.10)
\end{aligned}$$

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

$$P\{E_{n,m}\} \leq P\left\{\max_{n+1 \leq k \leq k_n-1} \sum_{\ell=n+1}^k a_\ell (\eta_\ell - E\{\eta_\ell | \mathcal{F}_\ell\}) 1_{\{\mu_m \leq \ell < \nu_m\}} > (R_4 - R_1)/4\right\} \quad (3.11)$$

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

$$S_{m,k} = \sum_{\ell=n+1}^k a_\ell \tilde{\eta}_\ell 1_{\{\mu_m \leq \ell < \nu_m\}}, \quad k \geq n+1.$$

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

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

Finally

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

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 i^{th} 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

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

$$\inf_{\substack{x^i \geq R_2 \\ |x^j| \leq R_4 \forall j \neq i}} \frac{\partial U}{\partial x^i}(x) > \sup_{|x^j| \leq R_1 \forall j} \frac{\partial U}{\partial x^i}(x). \quad (3.13)$$

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

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

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

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

similarly to the proof given above that

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

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

$$\begin{aligned} \hat{x} &= \frac{x}{|x|}, \quad |x| > 0 \\ \hat{\theta} &= \frac{\nabla U(x) - \langle \nabla U(x), \hat{x} \rangle \hat{x}}{|\nabla U(x) - \langle \nabla U(x), \hat{x} \rangle \hat{x}|}, \quad |\nabla U(x) - \langle \nabla U(x), \hat{x} \rangle \hat{x}| > 0 \\ &= 0, \quad |\nabla U(x) - \langle \nabla U(x), \hat{x} \rangle \hat{x}| = 0 \end{aligned}$$

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

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

where e^i is the i^{th} standard basis element in \mathbb{R}^d . Let $D(R_2) = \{x : x^i \geq R_2, |x^j| \leq R_4 \forall j \neq i\}$. Then using (A3)

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

and the same inequality holds with $g^2(x)$ replace by $g(x)$. Hence

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

Since $\frac{\partial U}{\partial x^i}(\cdot)$ is bounded on a compact set (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 $E\{|X(\beta(s)) - Y(\beta(s))|^2\} \rightarrow 0$ as $s \rightarrow \infty$. This is true even though the interval length $\beta(s) - s \rightarrow \infty$ as $s \rightarrow \infty$.

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

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

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, and ξ_k satisfies (A4') with $K = \mathbb{R}^d$ i.e.

$$E\{ |\xi_k|^2 | \mathcal{F}_k \} \leq L a_k^\alpha, \quad |E\{ \xi_k | \mathcal{F}_k \}| \leq L a_k^\beta \quad \text{w.p.1} \quad (3.15)$$

(note that $U(\cdot)$ is C^2 and if $\sigma(t_n, R) \wedge \tau(t_n, R) > t_{k_n}$ then $|X(t)|, |Y(t)| \leq R$ for $t_n \leq t \leq t_{k_n}$).

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 $X(0) = X_0 = x_0$, $X(t_n) = X_n = Y(t_n) = Y_n = x$, $|x| \leq r$. Then

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

We proceed to show that the r.s. of (3.16) tends to 0 as $n \rightarrow \infty$. Let

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

Note that by (3.15) and Corollary 1

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

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

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

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

$$\begin{aligned}
\mathbb{E}\{|\Delta_{k+1}|^2\} &\leq (1 + c_2 a_k) \mathbb{E}\{|\Delta_k|^2\} + c_2 a_k^\delta, \quad k \geq n \\
\mathbb{E}\{|\Delta_n|^2\} &= 0
\end{aligned}$$

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

Claim 3: There exists $\gamma > 1$ such that

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

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

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

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

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

$$\begin{aligned} u_{n,k+1} &\leq (1 + a_k)u_{n,k} + a_k^\delta, & k \geq n, \\ u_{n,n} &= 0, \end{aligned}$$

where $\delta > 1$. Now

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

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

$$\sup_{k: t_n \leq t_k \leq \gamma t_n} u_{n,k} \leq c_2 n^{(\gamma-1)A - (\delta-1)} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

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

$$X_{k+1} = X_k - a_k(\psi_k(X_k) + \xi_k) + b_k W_k, \quad k \geq 0 \tag{4.1}$$

where $\{a_k\}$, $\{b_k\}$, $\{\xi_k\}$, $\{W_k\}$ are defined as in Section 2 and $\psi_k(\cdot)$ is specified below. We will deal with the following conditions in this Section ($\alpha, \beta, \gamma_1, \gamma_2$ are constants whose values will be specified later).

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

$$E\{|\xi_k|^2 | \mathcal{F}_k\} \leq L_1 a_k^\alpha, \quad |E\{\xi_k | \mathcal{F}_k\}| \leq L_1 a_k^\beta \quad \text{w.p.1}$$

(B2) For $k = 0, 1, \dots$, let $\{\psi_k(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d\}$ be an \mathbb{R}^d -valued random vector field

(B3) Let K be a compact subset of \mathbb{R}^d . There exists $L_2 > 0$ such that

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

(B4) There exists $L_3, R > 0$ such that

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

(B5) There exists $L_4, R > 0$ such that

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

(B6) There exists $L_5, R > 0$ such that

$$E\{\langle \psi_k(\mathbf{x}), \mathbf{x} \rangle | \mathcal{F}_k\} \geq L_5 E\{|\psi_k(\mathbf{x})| |\mathbf{x}| | \mathcal{F}_k\} \quad \forall |\mathbf{x}| > R, \quad \text{w.p.1}$$

Theorem 3: Assume that (B1)-(B6) hold with $\alpha > -1$, $\beta > 0$ and $0 \leq \gamma_1 \leq \gamma_2 < 1$. Let $\{X_k\}$ be given by (4.1) and K be a compact subset of \mathbb{R}^d . Then $\{X_k^{x_0} : k \geq 0, x_0 \in K\}$ is a tight sequence of random variables.

The proof of Theorem 3 will require the following Lemmas.

Lemma 6: Assume the conditions of Theorem 3. Then there exists an integer k_0 and an $M_1 > 0$ such that

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

for $k \geq k_0$ and all x_0 .

Proof: Assume all quantities are conditioned on $X_0 = x_0$. Now it follows from (B2)-(B6) and the fact that X_k is \mathcal{F}_k -measurable that

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

$$E\{|\psi_k(\mathbf{X}_k)|^2, |\mathbf{X}_k| > R\} \geq L_3 a_k^{-\gamma_1} E\{|\mathbf{X}_k|^2, |\mathbf{X}_k| > R\}$$

$$E\{|\psi_k(\mathbf{X}_k)|^2, |\mathbf{X}_k| > R\} \leq L_4 a_k^{-\gamma_2} E\{|\mathbf{X}_k|^2, |\mathbf{X}_k| > R\}$$

and

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

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

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

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

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

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 $\alpha > -1$, $\beta > 0$, $0 \leq \gamma_1 \leq \gamma_2 < 1$ we have $\delta_1 > 1$, $\delta_2 > 1$ and $\delta_3 > 1/2$, and since $b_k = o(a_k^{1/2})$ we get

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

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

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

$$\begin{aligned}
& E\{ |X_{k+1}|^2, |X_k| \leq R \} - E\{ |X_k|^2, |X_k| \leq R \} \\
& \leq c_5(a_k^{\delta_4} + a_k^{\delta_5} b_k + b_k^2)
\end{aligned}$$

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

$$\begin{aligned}
& E\{ |X_{k+1}|^2, |X_k| \leq R \} - E\{ |X_k|^2, |X_k| \leq R \} \\
& \leq c_6 a_k \leq c_6 a_k^{1-\gamma_1/2} \tag{4.4}
\end{aligned}$$

for all $k \geq 0$.

Finally, let $M_1 = c_6/c_4 + R + 1$. Then combining (4.3), (4.4) gives the Lemma. \square

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

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

for $k \geq 0$ and all x_0 .

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

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

and

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

Combining these equations gives the Lemma. □

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

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

and by Lemmas 6 and 7 we also have

$$E_{0,x_0} \{ |X_{k+1}|^2 \} - E_{0,x_0} \{ |X_k|^2 \} \leq 0 \quad \text{if} \quad E_{0,x_0} \{ |X_k|^2 \} \geq c_1$$

and

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

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

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

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

5. TIGHTNESS AND CONVERGENCE FOR TWO EXAMPLE ALGORITHMS

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

$$(A5) \quad \liminf_{|x| \rightarrow \infty} \frac{|\nabla U(x)|}{|x|} > 0$$

$$(A6) \quad \overline{\lim}_{|x| \rightarrow \infty} \frac{|\nabla U(x)|}{|x|} < \infty$$

Example 1:

Here we consider the following algorithm:

$$X_{k+1} = X_k - a_k(\nabla U(X_k) + \xi_k) + b_k W_k, \quad k \geq 0 \quad (5.1)$$

Theorem 4: Assume (A1)-(A3), (B1), (A5), (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

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

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), (A6) hold, i.e. there exists M_1, M_2 such that

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

Intuitively, the upper bound on $|\nabla U(x)|$ is needed to prevent potentially unbounded oscillations of $\{X_k\}$ around the origin. It would be desirable to modify (5.1) in such a way that 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:

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

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

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

uniformly for x_0 in a compact set.

Proof: Let

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

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

Let

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

Let

$$\mathbf{X}_{k+1} = \mathbf{X}_k - a_k(\psi_k(\mathbf{X}_k) + \xi_k'') + b_k W_k, \quad (5.5)$$

(this defines ξ_k'') and $\mathcal{F}_k'' = \sigma(\mathbf{X}_0, \xi_0'', \dots, \xi_{k-1}'', W_0, \dots, W_{k-1})$. We shall show that $(\xi_k'', W_k, \mathcal{F}_k'')$ satisfies (B1) and $(\psi_k(\mathbf{x}), \mathcal{F}_k'')$ satisfies (B2)-(B6) with $\gamma_1 = 0$, $\gamma_2 = 2\gamma$. Hence by Theorem 3 $\{\mathbf{X}_k^{x_0} : k \geq 0, x_0 \in K\}$ 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 can not 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

$$E\{|\xi_k'|^2 | \mathcal{F}_k'\} \leq M_1 a_k^\alpha, \quad \forall \mathbf{X}_k \in K, \quad \text{w.p.1.}$$

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

Proof: Clearly

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

Hence for $\mathbf{X}_k \in K$ and k large enough

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

where we have used the Chebyshev inequality. It is easy to see that the inequality actually holds for all $k \geq 0$. 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 $M_1, M_2, M_3, M_4, M_5, R > 0$ such that

- (i) $\mathbb{E}\{|\xi_k''|^2 | \mathcal{F}_k''\} \leq M_1 a_k^\alpha$ w.p.1. Also W_k is independent of \mathcal{F}_k''
- (ii) $\mathbb{E}\{|\psi_k(x)|^2 | \mathcal{F}_k''\} \leq M_2 \quad \forall x \in K, \quad \text{w.p.1}$
- (iii) $\mathbb{E}\{|\psi_k(x)| | \mathcal{F}_k''\}^2 \geq M_3 |x|^2 \quad \forall |x| > R, \quad \text{w.p.1}$
- (iv) $\mathbb{E}\{|\psi_k(x)|^2 | \mathcal{F}_k''\} \leq M_4 \frac{|x|^2}{a_k^{2\gamma}} \quad \forall |x| > R, \quad \text{w.p.1}$
- (v) $\mathbb{E}\{\langle \psi_k(x), x \rangle | \mathcal{F}_k''\} \geq M_5 \mathbb{E}\{|\psi_k(x)| |x| | \mathcal{F}_k''\} \quad \forall |x| > R, \quad \text{w.p.1.}$

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

(i): Clearly

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

Hence

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

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

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

$$\begin{aligned} & E\{|\psi_k(x)|^2 | \mathcal{F}_k\} \\ & \leq |\nabla U(x)|^2 + \frac{|x|^2}{a_k^{2\gamma}} \Pr\{|\nabla U(x) + \xi_k| > \frac{|x| \vee 1}{a_k^\gamma} | \mathcal{F}_k\} \\ & \leq c_1 + \frac{c_1}{a_k^{2\gamma}} \Pr\{|\xi_k| > \frac{c_2}{a_k^\gamma} | \mathcal{F}_k\} \\ & \leq c_1 + c_3 E\{|\xi_k|^2 | \mathcal{F}_k\} \leq M_2 \quad \text{w.p.1,} \end{aligned}$$

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

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

$$\begin{aligned} & E\{|\psi_k(x)|^2 | \mathcal{F}_k\} \\ & \leq \frac{|x|^2}{a_k^{2\gamma}} + |\nabla U(x)|^2 P\{|\nabla U(x) + \xi_k| \leq \frac{|x| \vee 1}{a_k^\gamma} | \mathcal{F}_k\} \\ & \leq \frac{|x|^2}{a_k^{2\gamma}} + E\{|\nabla U(x)|^2, |\nabla U(x)| \leq \frac{|x|}{a_k^\gamma} + |\xi_k| | \mathcal{F}_k\} \\ & \leq 4 \frac{|x|^2}{a_k^{2\gamma}} + 3 E\{|\xi_k|^2 | \mathcal{F}_k\} \leq M_4 \frac{|x|^2}{a_k^{2\gamma}} \quad \text{w.p.1.} \end{aligned}$$

Since $\mathcal{F}_k'' \subset \mathcal{F}_k$ (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.

6. REFERENCES

- [1] Kushner, H. and Clark, D., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Applied Math. Science Series 26, Springer, Berlin, 1978.
- [2] Kushner, H.J., "Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: Global minimization via Monte Carlo," *SIAM Journal Applied Mathematics*, 47, pp. 169-185, 1987.
- [3] Ljung, L., "Analysis of recursive stochastic algorithms," *IEEE Transactions on Automatic Control*, Vol. AC-22, pp. 551-575, 1977.
- [4] Grenender, U., *Tutorial in Pattern Theory*, Div. Applied Mathematics, Brown Univ., Providence, RI, 1984.
- [5] Geman, S. and Hwang, C.R., "Diffusions for global optimization," *SIAM Journal Control and Optimization*, 24, pp. 1031-1043, 1986.
- [6] Gidas, B., "Global optimization via the Langevin equation," *Proc. IEEE Conference on Decision and Control*, 1985.
- [7] Chiang, T.S., Hwang, C.R. and Sheu, S.J., "Diffusion for global optimization in \mathbb{R}^n ," *SIAM Journal Control and Optimization*, 25, pp. 737-752, 1987.
- [8] Dixon, L. C. W. and G. P. Szegö, *Towards Global Optimization*, North Holland, 1978.
- [9] Hwang, C.-R., "Laplaces method revisited: weak convergence of probability measures," *Annals of Probability*, 8, pp. 1177-1182, 1980.