Hitting Time and Convergence Rate Bounds for Symmetric Langevin Diffusions

by

Gareth O. Roberts[1] and Jeffrey S. Rosenthal[2]

(November, 2016; revised May 2017)

Abstract. We provide quantitative bounds on the convergence to stationarity of real-valued Langevin diffusions with symmetric target densities.

Keywords. Langevin diffusion; computable bounds; coupling; stochastic monotonicity.

1 Introduction

Quantitative (computable) bounds on the convergence of Markov processes to stationarity are an important and widely studied topic, particularly in the context of Markov chain Monte Carlo (MCMC) algorithms (see e.g. Roberts and Tweedie, 1999; Rosenthal, 1995b, 1996, 2002; Jones and Hobert, 2001, 2004; Baxendale, 2005; and references therein). Most of this research has focused on discrete-time Markov chains. However, continuous-time Langevin diffusions also converge to stationary distributions, and they have applications in e.g. MCMC (Roberts and Tweedie, 1996) and in molecular dynamics (Pavliotis, 2014).

A study of the qualitative exponential convergence properties of Langevin diffusions was initiated in Roberts and Tweedie (1996). Quantitative bounds on their convergence are also of interest. A start in this direction was made for a few specific examples by Roberts and Rosenthal (1996) and Roberts and Tweedie (2000), but much remains to be done.

The current paper was inspired by a question from John Lafferty (personal communication), who asked about quantitative convergence bounds for Langevin diffusions on $\mathbf{R}$ with target densities proportional to $e^{-|x|^\beta}$ for some fixed $\beta > 1$. Below we provide quantitative convergence upper bounds for such diffusions, and more generally for any symmetric real Langevin diffusion satisfying certain conditions. Our bounds are conservative, but are still numerically modest. For example, we show that in the $e^{-|x|^\beta}$ case with $\beta = 2$, if we begin the

[1]Department of Statistics, University of Warwick, CV4 7AL, Coventry, U.K. Email: `g.o.roberts@lancaster.ac.uk`. Supported in part by EPSRC grants EP/20620/01 and EP/S61577/01.

[2]Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: `jeff@math.toronto.edu`. Web: `http://probability.ca/jeff/` Supported in part by NSERC of Canada.

diffusion at $y = 2$, then it converges to within 0.01 of stationarity in total variation distance by time 27. (Or, if $\beta = 1.1$ and $y = 10$, then it is within 0.01 of stationarity by time 68.)

Our proof requires bounds on hitting times, which are also developed below using probability generating functions. It uses the coupling inequality, and the stochastic monotonicity of the diffusions, following the general approach of Lund et al. (1996a, 1996b). Similar constructions have been used for Langevin diffusions in other contexts, see e.g. Silvestrov (1994, 1996) and Kartashov (1996).

2 Assumptions and Hitting Probabilities

Let $\pi : \mathbf{R} \to [0, \infty)$ be a target density on $\mathbf{R}$, satisfying the following:

(A1) (i) π is *symmetric*, i.e. $\pi(-x) = \pi(x)$ for all $x \in \mathbf{R}$;
(ii) π is *continuously differentiable*, i.e. is in C^1;
(iii) π is *unimodal*, i.e. $-\nabla \log \pi(x) \geq 0$ for all $x \geq 0$;
(iv) π has *light tails*, i.e. there is $b > 0$ such that $-\frac{1}{2}\nabla \log \pi(x) \geq b$ for all $x \geq 1$.

For example, (A1) is satisfied if $\pi(x) \propto e^{-|x|^\beta}$ for any fixed $\beta > 1$, in which case for $x \geq 0$ we have $-\frac{1}{2}\nabla \log \pi(x) = -\frac{1}{2}\nabla(-x^\beta) = \frac{1}{2}\beta x^{\beta-1}$, which is ≥ 0 for $x \geq 0$, and which is non-decreasing on $[1, \infty)$ so we can take $b = -\frac{1}{2}\nabla \log \pi(1) = \beta/2 > 0$.

On the other hand, if $0 < \beta < 1$ (or if the tails of π are polynomial), then $\lim_{x\to\infty} \nabla \log \pi(x) = 0$, so (A1) is not satisfied. Indeed, it follows from Roberts and Tweedie (1996, Theorem 2.4) that the convergence to stationarity is not even exponential in those cases. However, the convergence is still polynomial; this follows from results of Fort and Roberts (2005).

To continue, let $\{X_t\}$ be a Langevin diffusion for π, so $dX_t = \frac{1}{2}\nabla \log \pi(X_t)\, dt + dB_t$ (where $\{B_t\}$ is standard Brownian motion). Let H_y be this diffusion's first hitting time of 0, conditional on starting at $X_0 = y$. We wish to bound the tail probabilities of H_y.

Our key computation is the following bound on the probability generating function of H_y, i.e. of $M_y(s) := \mathbf{E}(s^{H_y})$. We require that:

(A2) The value $s \in \mathbf{R}$ satisfies that $s > 1$, and $s < \exp(b^2/2)$, and

$$1 < \frac{\exp\left(b - \sqrt{b^2 - 2\log s}\right)}{\cos(\sqrt{2\log s})} < 2\,.$$

Remark. It is computed directly that

$$\frac{d}{ds}\log\left(\frac{\exp\left(b-\sqrt{b^2-2\log s}\right)}{\cos(\sqrt{2\log s})}\right) \;=\; \frac{1}{b} \;>\; 0\,,$$

from which it follows that for any fixed $b > 0$, (A2) holds for all sufficiently small $s > 1$. More specifically, let $b_0 \doteq 0.54$ be the first positive solution to $e^b/\cos(b) = 2$. If $0 < b \leq b_0$, then the valid interval for s is $1 < s < \exp(b^2/2)$. If $b > b_0$, then the valid interval for s is from 1 to smallest $s_0 > 1$ for which

$$\frac{\exp\left(b-\sqrt{b^2-2\log s_0}\right)}{\cos(\sqrt{2\log s_0})} = 2\,.$$

Lemma 1. *If π satisfies (A1), then for any $y \in \mathbf{R}$, and any s satisfying (A2), the probability generating function $M_y(s) := \mathbf{E}(s^{H_y})$ of the Langevin diffusion for π satisfies $M_y(s) \leq B(\max(1,|y|), s, b)$, where*

$$B(y,s,b) \;=\; \frac{\exp\left((y-1)\left[b-\sqrt{b^2-2\log s}\right]\right)}{\cos(\sqrt{2\log s})} \;\Big/\; \left[2-\frac{\exp\left(b-\sqrt{b^2-2\log s}\right)}{\cos(\sqrt{2\log s})}\right].$$

Lemma 1 is proved in Section 5 below.

Assuming Lemma 1, we immediately obtain a bound on the tail probabilities of the hitting time of 0, as follows.

Proposition 2. *If π satisfies (A1), then for any $y \in \mathbf{R}$ and $t > 0$, and any s satisfying (A2), the hitting time H_y of the Langevin diffusion for π satisfies*

$$\mathbf{P}(H_y \geq t) \;\leq\; s^{-t}\,B\Big(\max(1,|y|),\, s,\, b\Big)\,.$$

Proof. It follows by Markov's inequality that

$$\mathbf{P}(H_y \geq t) \;=\; \mathbf{P}(s^{H_y} \geq s^t) \;\leq\; s^{-t}\,\mathbf{E}(s^{H_y}) \;=\; s^{-t}\,M_y(s) \;\leq\; s^{-t}\,B(\max(1,|y|),s,b)\,. \quad \blacksquare$$

Numerical Example. Suppose $\pi(x) \propto e^{-|x|^\beta}$, so $b = \beta/2$ as above. Then if, say, $y = \beta = 2$, then $b = \beta/2 = 1$, and choosing $s = 1.3$, we compute numerically from Proposition 2 that $\mathbf{P}(H_y \geq t) < 0.01$ whenever $t \geq 27$. This indicates that this process has probability over 99% of hitting 0 by time 27. By contrast, if $\beta = 1.1$ and $y = 10$, then taking $s = 1.14$ gives that $\mathbf{P}(H_y \geq t) < 0.01$ whenever $t \geq 68$.

3 Convergence Bounds: Reflected Case

Our main interest is in bounds on the convergence time to stationarity. We achieve this via the coupling inequality, using *stochastic monotonicity* as in Lund et al. (1996a, 1996b).

We first consider the version of the process which is "reflected" or "folded" at zero, which is equivalent to considering just the *absolute value* of the process. That is, we consider the process $R_t := |X_t|$. This process behaves just like $\{X_t\}$ on the positive half-line. But when it hits 0, it reflects back to the positive part rather than go negative.

By symmetry, this process has a stationary density $\overline{\pi}$ equal to π restricted to the positive half-line, i.e. $\overline{\pi}(x) = 2\,\pi(x)\mathbf{1}_{x\geq 0}$. We seek specific quantitative computable bounds on the total variation distance to stationarity of this process after time t, i.e. on

$$\|\mathcal{L}_y(R_t) - \overline{\pi}\| \;:=\; \sup_{A\subseteq\mathbf{R}} |\mathbf{P}_y(R_t\in A) - \overline{\pi}(A)|\,,$$

where the supremum is taken over all measurable subsets $A \subseteq \mathbf{R}$, and the subscript y indicates that the process was started at the state y, and $\overline{\pi}(A) := \int_A \overline{\pi}(x)\,dx$. We shall prove the following.

Proposition 3. *If π satisfies (A1), then for the reflected Langevin diffusion for π, started at state $y \geq 1$, for any s satisfying (A2), the total variation distance to stationarity at time $t > 0$ satisfies*

$$\|\mathcal{L}_y(R_t) - \overline{\pi}\| \;\leq\; 2\,\pi[0,y]\,s^{-t}\,B(y,s,b) \;+ 2\int_y^\infty \pi(z)\,s^{-t}\,B(z,s,b)\,dz\,,$$

with $B(y,s,b)$ as in Lemma 1.

Proof. Let $y \geq 1$ and $z > 0$, and let $\{X_t\}$ and $\{\widetilde{X}_t\}$ be two separate copies of the diffusion process, started at $X_0 = y$ and $\widetilde{X}_0 = z$. Let $R_t = |X_t|$ and $\widetilde{R}_t = |\widetilde{X}_t|$. We wish to couple the processes so that $\{R_t\}$ and $\{\widetilde{R}_t\}$ (started at y and at z, respectively) are both driven by the exact same Brownian motion $\{B_t\}$. This can be accomplished by designing $\{X_t\}$ and $\{\widetilde{X}_t\}$ so that when $X_t > 0$, $dX_t = \frac{1}{2}\nabla\log\pi(X_t)dt + dB_t$, while when $X_t < 0$, $dX_t = \frac{1}{2}\nabla\log\pi(X_t)dt - dB_t$, and similarly when $\widetilde{X}_t > 0$, $d\widetilde{X}_t = \frac{1}{2}\nabla\log\pi(\widetilde{X}_t)dt + dB_t$, while when $\widetilde{X}_t < 0$, $d\widetilde{X}_t = \frac{1}{2}\nabla\log\pi(\widetilde{X}_t)dt - dB_t$. With this choice, it follows from (A1) that whenever $R_t > 0$ and $\widetilde{R}_t > 0$, then $dR_t = \frac{1}{2}\nabla\log\pi(R_t)dt + dB_t$ and $dR_t = \frac{1}{2}\nabla\log\pi(R_t)dt + dB_t$, with the *same* Brownian motion $\{B_t\}$. Hence, by continuity, the relative ordering of R_t and $\widetilde{R}_t$ is preserved, i.e. if $y \leq z$ then $R_t \leq \widetilde{R}_t$ for all t, while if $y \geq z$ then $R_t \geq \widetilde{R}_t$ for all t.

With this coupling construction, it follows that if the *larger* of the two processes hits 0, then the other process must also equal 0 at that same time. Hence, the two processes must couple (i.e., become equal) by the time the larger of the two processes hits 0. But the larger of the two processes started at $\max(y, z)$. Therefore, if the coupling time is U, then U is stochastically bounded above by the hitting time $H_{\max(y,z)}$ of 0 from the state $\max(y, z)$, i.e. $\mathbf{P}(U \geq t) \leq \mathbf{P}(H_{\max(y,z)} \geq t)$.

Hence, by the usual coupling inequality (e.g. Roberts and Rosenthal, 2004, Section 4.1), the total variation distance of the processes after time t satisfies

$$\begin{aligned} \|\mathcal{L}_y(R_t) - \mathcal{L}_z(\widetilde{R}_t)\| \; &:= \; \sup_{A \subseteq \mathbf{R}} |\mathbf{P}(R_t \in A) - \mathbf{P}(\widetilde{R}_t \in A)| \\ &\leq \; \mathbf{P}(U \geq t) \; \leq \; \mathbf{P}(H_{\max(y,z)} \geq t)\,. \end{aligned}$$

Therefore, by Proposition 2, for any $s \geq 1$, $\|\mathcal{L}_y(R_t) - \mathcal{L}_z(\widetilde{R}_t)\| \; \leq \; s^{-t}\, B(\max(1, y, z), s, b)$.

Suppose now that we start the $\{R_t\}$ process at $y \geq 1$, and start the $\{\widetilde{R}_t\}$ process at a state $Z \sim \overline{\pi}$ chosen randomly from the stationary distribution. Then

$$\begin{aligned} \|\mathcal{L}_y(R_t) - \overline{\pi}\| \; &\equiv \; \left\| \mathcal{L}_y(R_t) - \left(\mathbf{E}_{Z\sim\overline{\pi}} \mathcal{L}_Z(\widetilde{R}_t) \right) \right\| \\ &\leq \; \mathbf{E}_{Z\sim\overline{\pi}} \|\mathcal{L}_y(R_t) - \mathcal{L}_Z(\widetilde{R}_t)\| \\ &\leq \; \mathbf{E}_{Z\sim\overline{\pi}} [\mathbf{P}(H_{\max(y,Z)} \geq t)] \\ &\leq \; \mathbf{E}_{Z\sim\overline{\pi}} [s^{-t}\, B(\max(1, y, Z), s, b)] \\ &= \; \int_0^\infty \overline{\pi}(z)\, s^{-t}\, B(\max(1, y, z), s, b)\, dz \\ &= \; \overline{\pi}[0, y]\, s^{-t}\, B(y, s, b) \; + \; \int_y^\infty \overline{\pi}(z)\, s^{-t}\, B(z, s, b)\, dz \\ &= \; 2\, \pi[0, y]\, s^{-t}\, B(y, s, b) \; + 2 \int_y^\infty \pi(z)\, s^{-t}\, B(z, s, b)\, dz\,. \end{aligned}$$

∎

Numerical Example. Again let $\pi(x) \propto e^{-|x|^\beta}$. Then from the above,

$$\|\mathcal{L}_y(R_t) - \overline{\pi}\| \; \leq \; 2\, \pi[0, y]\, s^{-t}\, B(y, s, b) \; + 2 \int_y^\infty \pi(z)\, s^{-t}\, B(z, s, b)\, dz\,.$$

Suppose again that $y = \beta = 2$ with $b = \beta/2 = 1$. Then choosing $s = 1.3$, Proposition 3 shows that $\|\mathcal{L}_y(R_t) - \overline{\pi}\| < 0.01$ whenever $t \geq 27$, i.e. the process also converges to within 99% of its stationarity distribution by time 27. By contrast, with $y = 10$ and $\beta = 1.1$, we find choosing $s = 1.14$ that $\|\mathcal{L}_y(R_t) - \overline{\pi}\| < 0.01$ whenever $t \geq 68$.

4 Convergence Bounds: Unreflected Case

Finally, we consider convergence bounds on the full, unreflected diffusion $\{X_t\}$. Here we cannot use stochastic monotonicity directly, because the full diffusion has no lowest state (or even lower bound) on which to force two copies of the diffusion to couple.

Nevertheless, using the symmetry condition A1(i), we are able to prove that the same convergence time bounds hold for the unreflected case as for the reflected case:

Proposition 4. *If π satisfies (A1), then for the full unreflected Langevin diffusion for π, started at state $y \geq 1$, for any s satisfying (A2), the total variation distance to stationarity at time $t > 0$ satisfies*

$$\|\mathcal{L}_y(X_t) - \pi\| \;\leq\; 2\,\pi[0,y]\, s^{-t}\, B(y,s,b) \;+ 2\int_y^\infty \pi(z)\, s^{-t}\, B(z,s,b)\, dz\,,$$

with $B(y,s,b)$ as in Lemma 1.

Proof. We again proceed, similarly to Proposition 3, by defining an appropriate coupling to preserve the ordering of the absolute values of the processes. We do this in stages. First, we jointly define two copies $\{X_t\}$ and $\{\widetilde{X}_t\}$ of the Langevin diffusion for π, by $X_0 = y$, $\widetilde{X}_0 \sim \pi$, $dX_t = \frac{1}{2}\nabla \log \pi(X_t)\, dt + dB_t$, and $d\widetilde{X}_t = \frac{1}{2}\nabla \log \pi(\widetilde{X}_t)\, dt - dB_t$, where $\{B_t\}$ is the same standard Brownian motion in both cases. In particular, $\{X_t\}$ and $\{\widetilde{X}_t\}$ are *anti-coupled*, i.e. are driven by the same Brownian motion but with opposite signs.

For this joint process, let

$$\tau \;=\; \inf\{t \geq 0 : X_t = \widetilde{X}_t\}$$

be the first time they meet. Thus, τ is a stopping time for the joint process. Finally, define another process $\{\widehat{X}_t\}$ by $\widehat{X}_0 = \widetilde{X}_0$, and

$$d\widehat{X}_t \;=\; \begin{cases} \frac{1}{2}\nabla \log \pi(\widetilde{X}_t)\, dt - dB_t\,, & t \leq \tau \\ \frac{1}{2}\nabla \log \pi(\widetilde{X}_t)\, dt + dB_t\,, & t > \tau \end{cases}$$

That is, $\{\widehat{X}_t\}$ is the same as $\{\widetilde{X}_t\}$ up to the meeting time τ, after which $\{\widehat{X}_t\}$ is the same as $\{X_t\}$. (This construction is valid since τ is a joint stopping time.)

We claim that this joint process preserves the absolute-value ordering of the processes $\{X_t\}$ and $\{\widehat{X}_t\}$, i.e. if $|X_0| \geq |\widehat{X}_0|$ then $|X_t| \geq |\widehat{X}_t|$ for all $t \geq 0$, while if $|X_0| \leq |\widehat{X}_0|$ then $|X_t| \leq |\widehat{X}_t|$ for all $t \leq 0$. Indeed, when X_t and $\widehat{X}_t$ have the same sign, the ordering is preserved simply because of the continuous sample paths. The only remaining case is if X_t

and $\widehat{X}_t$ are of opposite sign. But if that is true, then $X_t \neq \widehat{X}_t$, so we must have $t < \tau$, so that $\{X_t\}$ and $\{\widehat{X}_t\}$ are driven by Brownian motions of opposite sign. By the symmetry condition A1(i), just as in Proposition 3, this means that their absolute values are driven by the same Brownian motion, with the same drift function, and hence again cannot cross because of the continuous sample paths. So, in any case, the absolute-value ordering is preserved.

The rest of the argument is identical to that of Proposition 3. Indeed, when the larger (in absolute value) of the two processes reaches zero, the smaller one must also reach zero, so they must have coupled by that time. That is, conditional on $X_0 = y$ and $\widetilde{X}_0 = z$, we must have $\tau \leq H_{\max(y,z)}$. Hence, if we start the $\{X_t\}$ process at a state $y \geq 1$, and start the $\{\widetilde{X}_t\}$ process at a state $Z \sim \pi$ chosen randomly from the stationary distribution, then applying Proposition 2, we must again have

$$\begin{aligned}
\|\mathcal{L}_y(X_t) - \pi\| \;&\equiv\; \left\| \mathcal{L}_y(X_t) - \left(\mathbf{E}_{Z\sim\pi} \mathcal{L}_Z(\widetilde{X}_t) \right) \right\| \\
&\leq\; \mathbf{E}_{Z\sim\pi} \|\mathcal{L}_y(X_t) - \mathcal{L}_Z(\widetilde{X}_t)\| \\
&\leq\; \mathbf{E}_{Z\sim\pi} [\mathbf{P}(H_{\max(|y|,|Z|)} \geq t)] \\
&=\; \mathbf{E}_{Z\sim\overline{\pi}} [\mathbf{P}(H_{\max(|y|,Z)} \geq t)] \\
&\leq\; 2\,\pi[0,y]\, s^{-t}\, B(y,s,b) \;+ 2 \int_y^\infty \pi(z)\, s^{-t}\, B(z,s,b)\, dz\,,
\end{aligned}$$

exactly as in the proof of Proposition 3. ∎

Numerical Example. Since the bounds in Proposition 4 are the same as those in Proposition 3, the same bounds still apply. For example, if $y = \beta = 2$, then choosing $s = 1.3$, Proposition 4 again shows that $\|\mathcal{L}_y(X_t) - \overline{\pi}\| < 0.01$ whenever $t \geq 27$, i.e. the full unreflected process also converges to within 99% of its stationarity distribution by time 27. Or, with $y = 10$ and $\beta = b = 1.1$, Proposition 4 with $s = 1.14$ yields that $\|\mathcal{L}_y(X_t) - \overline{\pi}\| < 0.01$ whenever $t \geq 68$.

Remark. It is perhaps surprising that we obtain identical convergence bounds in both the reflected and the unreflected (original) case. Indeed, in general the reflected case should converge faster. However, since our upper bounds are derived using the hitting time of the larger process (in absolute value) to reach zero, we obtain the same bound in both cases.

5 Proof of Lemma 1

Finally, we prove Lemma 1. We assume $y \geq 1$; the case $y \leq -1$ then follows by symmetry, and the case $|y| < 1$ then follows since, by monotonicity, the hitting time of 0 from such y is stochastically bounded above by the hitting time of 0 from 1.

We first introduce an indicator process $\{I_t\}_{t\geq 0}$. Intuitively, I_t indicates whether at time t we are "waiting to hit 1", or are "waiting to hit 0 or 2 from 1". Specifically, we begin with $I_0 = 0$ if $y > 1$, or $I_0 = 1$ if $y = 1$. Then, each time X_t hits 1, we set $I_t = 1$. And, each time X_t hits 2 or 0, we set $I_t = 0$.

Next, we let $\{\widetilde{X}_t\}$ be a slight modification of $\{X_t\}$, as follows. $\{\widetilde{X}_t\}$ mostly follows the same dynamics as $\{X_t\}$. However, whenever $I_t = 1$, the drift of $\{\widetilde{X}_t\}$ is instead 0, i.e. we replace the Langevin diffusion dynamics by standard Brownian motion. Also, when $I_t = 0$, the drift of $\{\widetilde{X}_t\}$ is instead the constant value $-b$.

Now, because of the assumptions (A1), this new process is *stochastically larger* than the original process up to time H_y, i.e. it can only take *longer* to hit 0. So, writing $\widetilde{H}_y$ for the first time the modified process $\{\widetilde{X}_t\}$ hits 0, and $\widetilde{M}_y(s) := \mathbf{E}(s^{\widetilde{H}_y})$ for its probability generating function, we must have $M_y(s) \leq \widetilde{M}_y(s)$ for all $y \geq 0$ and $s \geq 1$. That is, we can (and will) use the hitting time of 0 for the modified process, as an upper bound on the hitting time of 0 for the original process. So, it suffices to show that $\widetilde{M}_y(s) = B(y, s, b)$, which we now do.

For the modified process $\{\widetilde{X}_t\}$, we break up the journey from $y \geq 1$ to 0 into steps:

1. Reach the state 1.

2. From there, reach either state 0 or state 2.

3(a). If it reached state 0, then we're done.

3(b). If instead it reached state 2, then return to step 1.

Now, since $\{\widetilde{X}_t\}$ has zero drift on $[0, 2]$, it follows that after reaching state 1, the process has equal probability 1/2 of reaching either 0 or 2. Therefore, the number of times the process will return to step 1 before finally hitting 0 is a geometric random variable $G \sim$ Geometric(1/2) with $\mathbf{P}[G = k] = 2^{-k-1}$ for $k = 0, 1, 2, \ldots$. It then follows that the time for $\{\widetilde{X}_t\}$ to reach state 0 from y can be written as

$$\widetilde{H}_y \;=\; T_{y1} + \sum_{i=1}^{G} (\widetilde{T}^{(i)}_{12|2} + T^{(i)}_{21}) + \widetilde{T}_{10|0}\,.$$

Here T_{y1} is the random time for $\{\widetilde{X}_t\}$ to reach 1 from y with constant drift $-b$, and each $T^{(i)}_{21}$ is an *independent* random time for $\{X_t\}$ to reach 1 from 2 with drift $-b$, and each $\widetilde{T}^{(i)}_{12|2}$ is an *independent* random time for $\{X_t\}$ with drift 0 to reach 2 from 1 with drift 0 but conditional on reaching 2 before 0, and $\widetilde{T}_{10|0}$ is an *independent* random time for standard Brownian motion to reach 0 from 1 with drift 0 but conditional on reaching 0 before 2.

Hence, with corresponding notation, the probability generating function $\widetilde{M}_y(s) := \mathbf{E}(s^{\widetilde{H}_y})$ is given by

$$\widetilde{M}_y(s) \;=\; M_{y1}(s) \times M_G\Big(\widetilde{M}_{12|2}(s) \times M_{21}(s)\Big) \times \widetilde{M}_{10|0}(s)\,. \qquad (1)$$

Now, these various formulas are known. First, by symmetry, $\widetilde{M}_{12|2}(s) = \widetilde{M}_{10|0}(s) = M_0^*(s)$ where $M_x^*(s)$ is the probability generating function for the time taken by standard Brownian motion to reach ± 1 when started at x (where $-1 < x < 1$), and this is known to be given for $s > 1$ by

$$M_x^*(s) \;=\; \frac{\cos(x\sqrt{2\log s})}{\cos(\sqrt{2\log s})}\,,$$

so that $M_0^*(s) = 1/\cos(\sqrt{2\log s})$. Also, G has known probability generating function given by $M_G(r) := \mathbf{E}(r^G) = (1-1/2)/(1-r/2) = 1/(2-r)$ for $1 < r < 2$.

Finally, it is known (see e.g. equation 2.0.1 on p. 301 of Borodin and Salminen, 2002; Proposition 3.3.5 of Etheridge, 2002) that if $\{W_t\}$ is standard Brownian motion, and $T_{a,b} = \inf\{t \geq 0 : W_t = a + bt\}$, then for $\alpha > 0$ and $a > 0$ and $b \geq 0$,

$$\mathbf{E}\left[\exp(-\alpha T_{a,b})\right] \;=\; \exp\left(-a\left[b + \sqrt{b^2 + 2\alpha}\right]\right)\,.$$

Hence, with the identification $\alpha := -\log s$ and $a := -(y-1)$, this indicates that

$$M_{y1}(s) \;=\; \exp\left((y-1)\left[b - \sqrt{b^2 - 2\log s)}\right]\right)\,,$$

and $M_{21}(s)$ then follows by setting $y = 2$.

Plugging these various formulae into (1), it follows by direct algebra (verified using the *Mathematica* symbolic algebra software, Wolfram 1988) that $\widetilde{M}_y(s) = B(y,s,b)$ with $B(y,s,b)$ as stated. Then, by monotonicity, $M_y(s) \leq \widetilde{M}_y(s) = B(y,s,b)$, giving the result. ■

Remark. An examination of the proof of Lemma 1 indicates that assumption (A1) is not strictly necessary, and quantitative bounds could be obtained by similar methods for Langevin diffusions for other target densities π as well.

6 Conclusion

This paper has provided quantitative bounds on the convergence to stationarity of Langevin diffusions for one-dimensional symmetric target densities π. The resulting time bounds are fairly modest, giving values like 27 and 68 for different examples. This is in contrast to most of the previous related work, which considered *qualitative* convergence properties such as exponential convergence, but which did not provide any quantitative bounds.

One might wonder if our results can be extended to multidimensional diffusions. In general, the answer is essentially no. Specifically, our main result relies crucially on monotonicity arguments. As such, generalisation to the multidimensional case is difficult without strong symmetry assumptions. For example, in Roberts and Tweedie (2000), it is remarked that the stochastic monotonicity can be retained for spherically symmetric target densities. But even in that situation, the drift of $|\mathbf{x}|$ would go to infinity as $\mathbf{x} \to 0$, so that the proof herein could not be applied in that case.

Finally, we note that our convergence time estimates are of course *upper bounds*, and it is reasonable to ask how tight they are. In the special case that $\pi(x) \propto e^{-|x|^\beta}$ with $\beta = 2$, the corresponding diffusion is an Ornstein-Uhlenbeck process. In that case, it can be computed exactly that if $X_0 = y$, then $X_t \sim N(ye^{-t}, (1 - e^{-2t})/2)$. Hence, $\|\mathcal{L}_y(X_t) - \pi\|$ is simply the total variation distance between $N(ye^{-t}, (1 - e^{-2t})/2)$ and the stationary distribution $N(0, 1/2)$. We compute directly that this is < 0.01 whenever $t \geq 4.73$. So, in this one special case, we can say that the *true* convergence time is 4.73, while our upper bound is 27. This gives a ratio of $27/4.37 \doteq 6.2$, i.e. our upper bound is about 6.2 times as large as the true answer, so that the theory is not too far from the actual convergence time in this case.

Acknowledgements. We thank John Lafferty for asking this question, and the anonymous referees and editors for very helpful reports which greatly improved the manuscript.

References

P.H. Baxendale (2005), Renewal theory and computable convergence rates for geometrically ergodic Markov chains. Ann. Appl. Prob. **15**, 700–738.

A.N. Borodin and P. Salminen (2002), Handbook of Brownian motion : facts and formulae, 2nd ed. Birkhäuser Verlag, Basel / Boston.

A. Etheridge (2002), A course in Financial calculus. Cambridge University Press.

G. Fort and G.O. Roberts (2005), Subgeometric ergodicity of strong Markov processes. Ann. Appl. Prob. **15(2)**, 1565–1589.

G.L. Jones and J.P. Hobert (2001), Honest exploration of intractable probability distributions via Markov chain Monte Carlo. Statistical Science **16**, 312–334.

G.L. Jones and J.P. Hobert (2004), Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. Ann. Stat. **32**, 784–817.

M.V. Kartashov (1996), Computation and estimation of the exponential ergodicity exponent for general Markov processes and chains with recurrent kernels. (Ukrainian.) Teor. Imovir. Mat. Stat. **54**, 47–57. English translation in Theory Prob. Math. Stat. **54** (1997), 49–60.

R.B. Lund and R.L. Tweedie (1996a), Geometric convergence rates of stochastically ordered Markov chains. Math. Oper. Research **21**, 182–194.

R.B. Lund, S.P. Meyn, and R.L. Tweedie (1996b), Computable exponential convergence rates for stochastically ordered Markov processes. Ann. Appl. Prob. **6**, 218-237.

G.A. Pavliotis (2014), Stochastic processes and applications. Texts in Applied Mathematics. Springer, Berlin.

G.O. Roberts and J.S. Rosenthal (1996), Quantitative bounds for convergence rates of continuous time Markov processes. Elec. J. Prob. **1(9)**, 1–21.

G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. Prob. Surv. **1**, 20–71.

G.O. Roberts and R.L. Tweedie (1996), Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli **2(4)**, 341–363.

G.O. Roberts and R.L. Tweedie (1999), Bounds on regeneration times and convergence rates for Markov chains. Stoch. Proc. Appl. **80**, 211–229. See also the corrigendum, Stoch. Proc. Appl. **91** (2001), 337–338.

G.O. Roberts and R.L. Tweedie (2000), Rates of convergence of stochastically monotone and continuous time Markov models. J. Appl. Prob. **37**, 359–373.

J.S. Rosenthal (1995b), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558–566.

J.S. Rosenthal (1996), Convergence of Gibbs sampler for a model related to James-Stein estimators. Stat. and Comput. **6**, 269–275.

J.S. Rosenthal (2002), Quantitative convergence rates of Markov chains: A simple account. Elec. Comm. Prob. **7**, No. 13, 123–128.

D.S. Silvestrov (1994), Coupling for Markov renewal processes and the rate of convergence in ergodic theorems for processes with semi-Markov switchings. (English summary.) Acta Appl. Math. **34**, 109–124.

D.S. Silvestrov (1996), Recurrence relations for generalized hitting times for semi-Markov processes. (English summary.) Ann. Appl. Prob. **6(2)**, 617–649.

S. Wolfram (1988), Mathematica: A system for doing mathematics by computer. Addison-Wesley, New York.