Geometric and Subgeometric Convergence of Diffusions with Given Stationary Distributions, and Their Discretizations

O. Stramer†and R.L. Tweedie‡

January 27, 1997

Abstract

We describe algorithms for estimating a given measure \( \pi \) known up to a constant of proportionality, based on a large class of diffusions (extending the Langevin model) for which \( \pi \) is invariant. We show that under weak conditions one can choose from this class in such a way that the diffusions converge at exponential rate to \( \pi \), and one can even ensure that convergence is independent of the starting point of the algorithm. When convergence is less than exponential we show that it is often polynomial at known rates. We then consider methods of discretizing the diffusion in time, and find methods which inherit the convergence rates of the continuous time process. These contrast with the behaviour of the naive or Euler discretization, which can behave badly even in simple cases.

KEYWORDS: Markov chain Monte Carlo, diffusions, Langevin models, posterior distributions, irreducible Markov processes, exponential ergodicity, uniform ergodicity, Euler schemes
AMS Subject Classifications: 60J05, 60J10, 60K25

1 Introduction

The use of diffusions has recently been suggested [5, 14, 13] as a continuous-time method of approach to problem of simulating a probability density \( \pi(x) \) which is only known up to a constant factor. This is especially relevant for Markov chain Monte Carlo methods, where for example one seeks to simulate a Bayesian posterior distribution, but many other contexts also use this approach: see [1, 2, 8, 14, 16, 21, 15].

One such class of algorithms, given by so-called Langevin diffusions, has been recently studied in Roberts and Tweedie [14]. They found that these continuous time processes perform well on many target densities, producing exponential rates of convergence, but if the target density has tails heavier than exponential, then convergence might be sub-exponential.

We develop here a wider class of diffusion algorithms when \( \pi \) is one-dimensional, and show that they can in general be chosen to converge geometrically fast: indeed, for

*Work supported in part by NSF Grants DMS 9504798 and DMS 9504561
†Postal Address: Department of Statistics and Actuarial Science, University of Iowa, Iowa City IA 52242, USA
‡Postal Address: Department of Statistics, Colorado State University, Fort Collins CO 80523, USA
almost all π the diffusion can actually be constructed to ensure that we have uniform convergence from all starting points. We address the multi-dimensional problem in a sequel [19].

For a discussion of some of the stability properties enjoyed by geometrically ergodic chains in simulation, we refer the reader to [14]: we note for example that such chains have Central Limit Theorems and the like available, which makes it much easier to work with the outcomes of the algorithms.

In a related paper [18], we also address the question of employing the discrete approximations to these diffusions as candidate distributions in the Hastings-Metropolis algorithm, and show that in this context also, convergence can be made both rapid and (in surprising generality) uniform in the starting point of the algorithm.

2 Diffusions with given stationary distributions

We will consider diffusions on \( \mathbb{R} \) which have a given stationary distribution \( \pi \), where we assume that we know \( \pi \) only up to a constant of proportionality: that is, when only \( \pi(x)/\pi(y) \), or equivalently \( \nabla \log \pi \), is known, where \( \nabla \) is the usual differential operator \( \nabla f(x) = df/dx \). We restrict ourselves to \( \mathbb{R} \) for convenience, and the results hold with little change for diffusions on a compact interval \( I \subset \mathbb{R} \). In higher dimensions, however, one needs more care and this will be addressed in [18].

One example of such a process is the Langevin diffusion for the density \( \pi \), studied in [14]. This is the diffusion process \( L(t) \), which satisfies the stochastic differential equation

\[
dL(t) = dW(t) + \frac{1}{2} \nabla \log \pi(L(t))dt,
\]

here \( W(t) \) is standard one-dimensional Brownian motion. It is then well-known that \( L(t) \) is a time-homogeneous reversible Markov processes evolving on \( (\mathbb{R}, \mathcal{B}) \), with transition probability law \( P^t_L(x,A) = P(L(t) \in A | L(0) = x), x \in \mathbb{R}, A \in \mathcal{B} \); here \( \mathcal{B} \) denotes the Borel \( \sigma \)-field on \( \mathbb{R} \). With this construction, \( \pi \) can be shown to be the stationary or invariant measure for \( L \): that is,

\[
\pi(A) = \int \pi(dy)P^t_L(y,A), \quad A \in \mathcal{B},
\]

for every \( t \geq 0 \).

It then follows that under some weak regularity conditions (see section 2.1 of [14]) the transition probabilities \( P^t_L(x,A) \) converge to \( \pi \) in the total variation norm: that is, for \( \pi \)-a.e \( x \)

\[
\|P^t_L(x,\cdot) - \pi\| = \frac{1}{2} \sup_{A \in \mathcal{B}} |P^t_L(x,A) - \pi(A)| \to 0.
\]

The Langevin diffusion is only one of a class of diffusions that have these properties, and a main aim of this paper is to describe the speed of convergence of other diffusion models which satisfy (2). We will find that the Langevin diffusion is in fact a rather
inefficient choice in this respect: this is similar to the experience of [6] who consider related problems from a rather different perspective. Note also that in [14], it is pointed out (p. 343) that non-reversible algorithms (which means in this case non-Langevin diffusions) may converge more rapidly than reversible algorithms: this is certainly born out by our results.

We will let $X(t)$ be a solution to the stochastic differential equation

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \quad X(0) = x \in \mathbb{R},$$  \hspace{1cm} (3)

where $W$ is a one dimensional Brownian motion; and we write $P^X_t(x,A) = \mathbb{P}(X(t) \in A|X(0) = x)$ for all $x \in \mathbb{R}, A \in \mathcal{B}$. In order for $X(t)$ to have the given stationary distribution $\pi$ we choose functions $b: \mathbb{R} \rightarrow \mathbb{R}$, $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ such that

$$b(x) = \left[\frac{1}{2} \nabla \log \pi(x)\right] \sigma^2(x) + \sigma(x) \sigma'(x).$$  \hspace{1cm} (4)

Note that if $\sigma(x) \equiv 1$, then $X(t)$ is the Langevin diffusion $L(t)$ defined as in (1).

Throughout the paper we will assume that the functions $b, \sigma$ can be chosen such that

(i) $b$ and $\sigma$ are bounded on compact subintervals of $\mathbb{R}$;

(ii) For any compact subinterval $J$ of $\mathbb{R}$ there exists some $d > 0$ such that $\sigma^2(x) > d$ for all $x \in J$.

(iii) $p(-\infty+) = -\infty$, $p(\infty-) = \infty$ where $p'(x) = \frac{1}{\pi(x)\sigma^2(x)}$.

Weaker conditions are possible but these serve to ensure that the diffusion corresponding to (3) is well defined: obviously, however, this imposes some weak constraints on $\pi$ through (4).

We note that (4) is equivalent to choosing $b$ and $\sigma$ to satisfy

$$\pi(x) \propto \frac{1}{\sigma^2(x)p'(x)}, \quad x \in \mathbb{R},$$  \hspace{1cm} (5)

where for some fixed $c \in \mathbb{R}$

$$p'(x) = \exp \left[-2 \int_{\infty}^{x} \frac{b(\zeta)d\zeta}{\sigma^2(\zeta)}\right], \quad x \in \mathbb{R}. $$  \hspace{1cm} (6)

Alternatively, one can fix $b$ and solve the equation (4) for $\sigma$: it is straightforward to note that this gives

$$\sigma^2(x) = -\int_{x}^{\infty} 2[p(y)/\pi(x)]b(y)dy$$  \hspace{1cm} (7)

as the correct relationship.

We now have, using the notation of [10]

**Theorem 2.1** Suppose $X(t)$ satisfies (3) with $b, \sigma$ satisfying (5). Then:
(i) \( X(t) \) has the Feller property; that is, \( P_X^t(x, \cdot) \) is weakly continuous as a function of \( x \);

(ii) some skeleton chain of \( X(t) \) is \( \mu^{Leb} \)-irreducible;

(iii) \( \pi \) is invariant for \( X(t) \); and

(iv) for all \( x \in \mathbb{R} \)

\[
\| P_X^t(x, \cdot) - \pi \| \to 0. \tag{8}
\]

Proof We have from [7] that the process \( p(X(t)) \) with \( p' \) as defined in (6) is \( B(A_t) \), where \( \{ B(t), \mathcal{F}(t) \} \) is a standard Brownian motion and \( A_t \) is a stopping time for \( \mathcal{F}(t) \). Hence some skeleton chain is \( \mu^{Leb} \)-irreducible. The Feller property follows from Exercise 7.3.3 and Lemma 11.1.1 of [20], and so (8) follows directly from Theorem 6.1 of [10]. 

There is no unique way to choose a continuous time process that converges to \( \pi \). We illustrate this in the following three examples.

Example 1 The exponential class \( \mathcal{E} \)

We will say that \( \pi \in \mathcal{E} \), as introduced in [14], if for some \( x_0 \), and some constants \( \gamma > 0 \) and \( 0 < \beta < \infty \), \( \pi \) takes the form

\[
\pi(x) \propto e^{-\gamma |x|^{\beta}}, \quad |x| \geq x_0, \tag{9}
\]

so that for \( |x| \geq x_0 \)

\[
\nabla \log \pi(x) = -\gamma \beta (\text{sgn}(x)) |x|^{\beta-1} \tag{10}
\]

If, instead of the Langevin model with \( \sigma^2 \equiv 1 \), we choose the polynomial form \( \sigma^2(x) = \alpha_s x^{\gamma_s} \) for some \( \alpha_s > 0 \) and \( \gamma_s > 0 \), then from (4) we must choose

\[
b(x) = -\left(\frac{1}{2} \gamma \beta x^{\beta-1}\right) \alpha_s x^{\gamma_s} + \frac{1}{2} \alpha_s \gamma_s x^{\gamma_s-1}. \tag{11}
\]

for \( x > x_0 \), and symmetrically for \( x < -x_0 \). Clearly such a \( b \) satisfies our continuity requirements provided \( \pi(x) \) is smooth enough for \( |x| \) small.

Thus for large positive \( x \) the drift function \( b \) satisfies

\[
b(x) \sim \alpha_s x^{\gamma_b} \tag{12}
\]

with

\[
\gamma_b = \gamma_s + \beta - 1. \tag{13}
\]

It follows, for example, that if \( \pi \) has simple exponential tails \( (\beta = 1) \), then we require the same polynomial order for the choice of \( b, \sigma^2 \); for Gaussian tails we need to ensure the drift is an order higher than the dispersion; and so on.

It was shown in [14] that for the Langevin diffusion, exponential convergence occurs if and only if \( \beta \geq 1 \) for \( \pi \in \mathcal{E} \). We will see below that one can choose \( b, \sigma \) such that
exponential convergence occurs for more general diffusions no matter what the value of the exponent $\beta$; and moreover that such convergence can even be made uniform in the starting point.

□

Example 2 Distributions with polynomial tails

We will also consider distributions with polynomial tails. As a concrete example of this we will take $\pi$ as $T_n$, the t distribution with $n \geq 1$ degrees of freedom: that is, $\pi(x) \propto (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$, $x \in \mathbb{R}$.

Then we can choose many combinations of $b, \sigma$, including:

(i) $\sigma(x) \equiv 1$ and $b(x) = \frac{(n+1)x}{2(n+x^2)}$ (i.e. the Langevin choice); or
(ii) $\sigma^2(x) = n + x^2$ and $b(x) = \frac{n+1}{2} x$; or
(iii) $\sigma^2(x) = (n + x^2)^{1.5}$ and $b(x) = \frac{n-2}{2} (n + x^2)^{0.5}$

In the first case we have

$$ b(x) \sim \frac{(n+1)}{2} x^{-1}, \quad \sigma^2(x) \sim x^0; \quad (14) $$

in the second case we have the polynomial tails given by

$$ b(x) \sim \frac{n-2}{2} x^1, \quad \sigma^2(x) \sim x^2; \quad (15) $$

and in the third case we have the polynomial tails given by

$$ b(x) \sim \frac{n+1}{2} x^2, \quad \sigma^2(x) \sim x^3. \quad (16) $$

We shall see that in the first case the chain is not exponentially ergodic (as we might conjecture from the behaviour of models in $E$); in the second case we do achieve an exponential rate of convergence; and in the third case we even achieve uniform ergodicity.

□

Example 3 Distributions with mixed tails We can obviously handle more general distributions. For example, suppose that

$$ \pi(x) \propto |x|^\alpha e^{-\gamma |x|^\beta}, \quad |x| \geq x_0, \quad \alpha \geq 0, \beta > 0. \quad (17) $$

for some $\alpha \geq 0$ and $\beta > 0$. Then we can choose either of the following:

(i) $\sigma(x) \equiv 1$ and $b(x) = \frac{\alpha}{2x} - \frac{\gamma |x|^\beta - 1 \text{sgn}(x)}{2}$ (again, the Langevin choice).
(ii) \( \sigma^2(x) = x^2 \) and \( b(x) = x^2 \left( \frac{\alpha}{2x^2} - \frac{\gamma}{2x} + \frac{1}{2} \text{sgn}(x) \right) + x \).

We shall see that the behaviour of the diffusion will be determined by \( \beta \) and not by \( \alpha \) in this context.

\[ \Box \]

The practical problem in which we are interested is to choose diffusion processes from the many satisfying (5) that will give us a good rates of convergence to \( \pi \), and we turn to this now.

3  Exponential rates of convergence for diffusions

As shown in [9, 4, 14, 22], an appropriate norm by which to measure rates of convergence is the \( V \)-norm \( \| \mu \|_V \), defined for a measurable function \( V \geq 1 \) on \( \mathbb{R} \) and a signed measure \( \mu \) as \( \sup_{|g| \leq V} |\mu(g)| \). Although formally stronger than the total variation norm, which corresponds to \( V \equiv 1 \), it can be shown that they are in general equivalent, in the sense that geometric or polynomial convergence in total variation actually implies similar convergence in \( V \)-norm for some appropriate \( V \).

For any \( V \geq 1 \), we formally define \( V \)-uniform ergodicity by requiring that for all \( x \in \mathbb{R} \)

\[ \| P^t_{\mathcal{X}}(x, \cdot) - \pi \|_V \leq V(x) R^t \rho, \quad t \geq 0, \tag{18} \]

for some \( R < \infty, \rho < 1 \). We call \( X(t) \) exponentially ergodic if it is \( V \)-uniformly ergodic for some such \( V \), and if \( X(t) \) is \( V \)-uniformly ergodic for \( V \equiv 1 \), which ensures that the convergence is independent of the starting point, then we call \( X(t) \) uniformly ergodic. It is known [9] that if convergence is indeed independent of the starting point then it must be exponential in the sense of (18) with \( V \equiv 1 \), and that if there is exponential convergence in any sense, then there exists a \( V \) such that we have \( V \)-exponential ergodicity; so we lose no generality by using these seemingly strong forms of convergence.

We will apply drift criteria to verify the exponential convergence of \( X(t) \), as developed in [9]. Our results will be based on the Itô formula, in the following form. Let \( V : \mathbb{R} \rightarrow \mathbb{R} \) be a non-negative twice continuously differentiable function. Then \( V(X(t)) \) satisfies the stochastic differential equation

\[ dV(X(t)) = \mathcal{L}_V(X(t))dt + \dot{V}(X(t))\sigma(X(t))dW(t), \tag{19} \]

where \( \mathcal{L}_V(x) := \dot{V}(x) b(x) + \frac{1}{2} \sigma^2(x) \frac{dV(x)}{dx} \) is the mean velocity of \( V(X(t)) \) at \( X(t) = x \). It is easily seen that \( \mathcal{L}_V \) is the extended generator as defined in [11].

We then have from Theorem 2.1 and Theorem 6.1 of [9] that \( X(t) \) is \( V \)-uniformly ergodic for any \( V \geq 1 \) that is twice continuously differentiable such that

\[ \mathcal{L}_V \leq -cV + b \| \sigma \|_C \tag{20} \]
for some constants $b, c > 0$, and some compact non-empty set $C$. If $V$ is bounded then we conclude from Theorem 16.0.2 of [9] and Itô’s formula that $X(t)$ is uniformly ergodic. We now have

**Theorem 3.1** Let $b, \sigma^2$ and $\pi$ be as in (4). Assume that there exist constants $\alpha_b, \alpha_s > 0$ and $\gamma_b, \gamma_s$ satisfying
\[
\gamma_b - \gamma_s + 1 \geq 0
\] (21)
such that as $|x| \to \infty$
\[
\frac{|b(x)|}{|x|^{\gamma_b}} \to \alpha_b
\] (22)
and
\[
\frac{\sigma^2(x)}{|x|^{\gamma_s}} \to \alpha_s
\] (23)

Then

(i) $X(t)$ is exponentially ergodic if and only if either $\gamma_b \geq 1$ or $\gamma_b \geq \frac{\gamma_s}{2}$;

(ii) $X(t)$ is uniformly ergodic if $\gamma_b > 1$.

**Proof** We first note that for the Langevin case $\gamma_s = 0$, (i) follows from [14]. Assume now that $\gamma_b \geq \gamma_s > 0$ and define $V(x) = \exp(s|x|^{\gamma_b-\gamma_s+1}), x \in \mathbb{R}$ for some $0 < s < 2\alpha_b/[\alpha_s(\gamma_b - \gamma_s + 1)]$. We now verify that (20) holds for large $|x|$.

In the case that $\gamma_b \geq 1$ we have that if $\gamma_b - \gamma_s + 1 > 0$, then (20) holds for the test function $V(x) = |x|^{2r}, x \in \mathbb{R}, r > 0$ and if $\gamma_b - \gamma_s + 1 = 0$ then (20) holds for $V(x) = |x|^{2r}, x \in \mathbb{R}, 0 < 2r < 1 + \frac{2\alpha_b}{\alpha_s}$.

Otherwise $\gamma_b < 1$ and $\gamma_s < 2$. Now (20) holds if $2\gamma_b - \gamma_s > 0$ for the test function $V(x) = \exp(|x|^\gamma)$, where $1 - \gamma_b \leq \alpha \leq 1 - \frac{\gamma_s}{2}$. On the other hand, if $2\gamma_b - \gamma_s < 0$, then let us define $f(x) := |x|^{1 - \frac{\gamma_s}{2}}/[1 - \frac{\gamma_s}{2}]$. Now the Itô formula shows that for large values of $f(X(t))$,
\[
df(X(t)) \approx b(f(X(t)))dt + dW_t,
\]
where $b(f(X(t))) \approx -k|f(x)|^{\frac{2\gamma_b - \gamma_s}{\gamma_b - \gamma_s}}$ for some $k > 0$. It follows from [14] that $f(X(t))$ is not exponentially ergodic if $2\gamma_b - \gamma_s < 0$.

Finally, assume that $\gamma_b > 1$ and define $V(x)$ to be a smooth function on $\mathbb{R}$ such that for large $M$ and small $\delta$,
\[
V(x) = \begin{cases} 
1 & \text{if } |x| < M - \delta \\
2 - |x|^{-\alpha} & \text{if } |x| > M 
\end{cases}
\]
where $\alpha > 0$ and $\alpha > \gamma_s - 2$. It is now easy to check that (20) holds for the test function $V$ and hence $X(t)$ is uniform ergodic. $\square$
For a given distribution \( \pi(x) \), we now illustrate that different choices of \( b \) and \( \sigma \) will give us different rates of convergence: that is, convergence is not just a function of the shape of the tails of the target distribution.

**Example 1 (ctd.)** Suppose \( \pi \in \mathcal{E} \), with \( 0 < \beta < 1 \). If we choose the Langevin diffusion then \( \gamma_\alpha = 0 \) and \( \gamma_\beta = \beta - 1 < 0 \), and as in [14] we do not have exponential rate of convergence.

If however we choose \( 1 - \beta < \gamma_\beta \leq 1 \) then \( \gamma_\alpha = \gamma_\beta + 1 - \beta \leq 2 \gamma_\beta \) and we have exponential ergodicity, but not necessarily independent of the starting point.

Finally, if we choose \( \gamma_\beta > 1 \) then we find \( \gamma_\alpha = \gamma_\beta + 1 - \beta < \gamma_\beta + 1 \) and we have uniform ergodicity.

**Example 2 (ctd.)** Now suppose that \( \pi \sim \mathcal{T}_n \). If we choose the Langevin diffusion then we have \( \alpha_\beta = \frac{n+1}{2} \), \( \alpha_\alpha = 1 \), \( \gamma_\beta = -1 \) and \( \gamma_\alpha = 0 \), and we do not have an exponential rate of convergence.

However, if we choose \( \sigma^2(x) = n + x^2 \) and \( b(x) = \frac{-(n-1)x}{2} \), then we have \( \alpha_\beta = \frac{n-1}{2} \), \( \alpha_\alpha = 1 \), \( \gamma_\beta = 1 \) and \( \gamma_\alpha = 2 \), and we have \( V \)-uniform convergence for any \( V(x) = |x|^{2r} \), \( 0 < 2r < n \). As we might expect, for larger \( n \) we thus can ensure exponential convergence of higher order moments.

Finally, even for this heavy-tailed distribution, if we choose \( \sigma^2(x) = (n + x^2)^\theta \) for some \( \theta > 1 \), and \( b(x) = \frac{-(n-1)x(n + x^2)^{\theta-1}}{2} \), then we have \( \gamma_\beta = 2\theta - 1 > 0 \), \( \gamma_\alpha = 2\theta > 2 \), and we have uniform ergodicity.

These examples illustrate general principles. We now prove, for virtually any \( \pi \), that one can always construct some diffusion that converges uniformly fast to \( \pi \) as its stationary distribution.

**Theorem 3.2** Let \( \pi \) be an arbitrary probability density on \( \mathbb{R} \) which is positive and twice differentiable almost everywhere.

(i) Suppose that \( \pi \) has asymptotically exponential tails such that

\[
-\frac{\log \pi(x)}{|x|^\beta} \to \alpha, \quad x \to \infty,
\]

for some \( \alpha, \beta > 0 \). Then there exist diffusions defined as in (4) which are uniformly ergodic.

(ii) Suppose that \( \pi \) has asymptotically polynomial tails such that

\[
\pi(x)|x|^\beta \to \alpha, \quad x \to \infty,
\]

for some \( \alpha > 0, \beta > 2 \). Then there exist diffusions defined as in (4) which are uniformly ergodic.
Proof Assume that (24) holds. Then for $|x| \geq x_0$, $\pi(x) \sim e^{-\alpha|(|x|^{\beta})}$, so that as in Example 1 we can choose for $|x| > x_0$, $b(x) \sim \alpha_4 x^{3\beta}$ and $\sigma^2(x) = \alpha_4 |x|^{\gamma}$ where $\alpha_4, \alpha_5 > 0, \gamma_5 > 1$ and $\gamma_5 = \gamma_5 + 1 - \beta < \gamma_5 + 1$. Now (i) follows directly from Theorem 3.1 (ii).

Now assume that (25) holds for some $\beta > 2$. Then for $|x| > x_0$, $\nabla \log \pi(x) \sim \frac{-\beta \text{sgn}(x)}{|x|}$. Thus if we choose $\sigma^2(x) = \alpha_4 |x|^{\gamma}$ with $2 < \gamma_5 < \beta$ then $b(x) \sim -\alpha_4 \left(\frac{\beta}{2} - \frac{\gamma_5}{2}\right) |x|^{\gamma - 1} (\text{sgn}(x))$, and (ii) follows from Theorem 3.1 (ii). \hfill \Box

4 Subgeometric rates of convergence

Although one can find exponentially or even uniformly converging diffusions for virtually all distributions $\pi$ as in Theorem 3.2, these may require some computation and it may be convenient to use, for example, the simple Langevin diffusion if its convergence properties are not too poor, especially if we can start from the “center” of $\pi$ in some sense.

We have seen that some diffusions may not be exponentially ergodic. We therefore consider conditions under which we can show that there is at least a subgeometric rate of convergence to $\pi$. Guided by the results in [22], our goal will be to find conditions under which

$$r(t)\|P^t\pi(x, \cdot) - \pi\|_f \to 0$$

as $t \to \infty$, where for some $k_1, k_2 \in \mathbb{Z}_+$, $f(x) = |x|^{k_1} \vee 1$, $x \in \mathbb{R}$ and $r(t) = t^{k_2} \vee 1$, $t \geq 0$. A more general subgeometric rate of convergence was studied in [22], but for our purposes this combination is more than sufficient, and it seems plausible that such polynomial convergence of polynomial moments is what we might expect in this case.

**Theorem 4.1** Suppose that

$$b^2(x) + \sigma^2(x) \leq K(1 + x^2), \quad x \in \mathbb{R},$$

where $K$ is some positive constant. Let $V(x) = |x|^{k_1} \vee 1$, $x \in \mathbb{R}$ and assume that for all $r \leq k \leq L$, $k, L \in \mathbb{Z}_+$,

$$L_V(x) \leq -c|x|^{k-r} + b \mathbb{1}_C,$$

where $k - r \in \mathbb{Z}_+$, $b, c > 0$ and $C$ is some compact non-empty set. Then (26) holds with

$$f(x) = |x|^{k-r} \vee 1, \quad r(t) = t^{L-k} \vee 1, \quad r \leq k \leq L.$$

Proof We first show that the transition law $P^h$ of the $h$-skeleton chain $X_h$ satisfies, for any $h > 0$,

$$n^{L-k}\|P^h(x, \cdot) - \pi\|_f \to 0$$

(29)
as \( n \to \infty \). Dynkin's formula (see [11], Section 1.3) shows that for any \( h > 0 \)
\[
E_x[|X_h|^k] = |x|^k + E_x \int_0^h \mathcal{L}_V(X(t)) \, dt \\
\leq |x|^k - cE_x \int_0^h |X(t)|^{k-r} \, dt + bh.
\] (30)

Using Problem 5.3.15 of [7] and (27) we conclude that for any \( h > 0 \) there exists \( c^*, \rho > 0 \) such that
\[
E_x[|X_h|^k] \leq |x|^k - c^*|x|^{k-r}, \quad |x| > \rho.
\] (31)
The same argument as in the proof of Proposition 5.2 of [22] now gives (29).

We next show that subgeometric convergence of the \( h \)-skeletons guarantees the same behaviour for the process \( X(t) \). Similarly to the proof of Theorem 6.1 of [11], we write for arbitrary \( t \geq 0, t = nh + s \) with \( 0 \leq s < h \)
\[
t^{L-k}||P^t_X\pi(x,\cdot) - \pi||_f = (nh + s)^{L-k} \sup_{|g| \leq f} |P^{nh+s}_X g - \int g \, d\pi|
\leq ((n+1)h)^{L-k} \int |P^{nh}_X (x, dw) - \pi(dw)| \int P^y_X (w, dy) f(y).\]

Using (30), for any \( s \in [0, h] \) we have that \( \int P^{h}_X (w, dy) f(y) \leq bf(w) \). It now follows from (29) that \( t^{L-k}||P^t_X\pi(x,\cdot) - \pi||_f \to 0 \) as \( t \to \infty \). \( \square \)

As in [22], such results stated in terms of polynomial moments and rates of convergence can be extended to fractional indices, and for simplicity we will assume that \( \alpha_b, \alpha_s, \gamma_b \) and \( \gamma_s \) are fractional indices. The following is then a direct consequence of Theorem 4.1.

**Theorem 4.2** Suppose that the conditions of Theorem 3.1 hold and \( X(t) \) is not exponentially ergodic. Then (26) holds with
\[
f(x) = |x|^{k-r} \vee 1, \quad x \in \mathbb{R}; \quad r(t) = t^{L-k} \vee 1, \quad t \geq 0,
\]
where \( r \leq k \leq L \) is a positive fractional index, \( r = 1 - \gamma_b \) and

(a) \( L \) is any positive fractional index if \( \gamma_b - \gamma_s + 1 > 0 \);
(b) \( L = \frac{2\alpha_b}{\alpha_s} \) if \( \gamma_b - \gamma_s + 1 = 0 \).

\( \square \)

We illustrate our result in the following examples.

**Example 4** Suppose that \( \pi \sim \mathcal{T}_n \) and choose \( \sigma(x) \equiv 1 \) and \( b(x) = \frac{-(n+1)x}{2(n+x^2)} \), so that \( \alpha_b = \frac{n+1}{2} \) and \( \alpha_s = 1 \). Let \( V(x) = |x|^k, x \in \mathbb{R} \). Then
\[
\mathcal{L}_V(x) = k|x|^{k-1} \text{sgn}(x) \frac{-(n+1)x}{2(n+x^2)} + \frac{b(k-1)x^{k-2}}{2}
\]
and the conditions of Theorem 4.1 hold with \( L = n + 1 \) and \( r = 2 \). As one might expect, we can guarantee a better rate of polynomial convergence when \( n \) is large.
Example 5 Consider now $\pi \in \mathcal{E}$ with $0 < \beta < 1$. The Langevin diffusion is, as we have seen, not geometrically ergodic. Recall that $\gamma_\alpha = 0$ and $\gamma_\beta = \beta - 1 < 0$, so if again we take $V(x) = |x|^k, x \in \mathbb{R}$, then there exists $K > 0$ such that for all $|x| > x_0$

$$\mathcal{L}_V(x) \leq -K|x|^{k-1-\gamma_\beta}.$$ 

Thus the conditions of Theorem 4.1 hold with $L = \infty$ and $r = 1 - \gamma_\beta = 2 - \beta$.

Thus although the convergence is not geometric, the total variation (for example) does converge faster than any particular polynomial rate.

5 Approaches to Discretization

One of the most striking results in [14] is that naive discretizations of the Langevin diffusion do not converge in the same manner as the diffusion itself: in particular it is possible that such discretizations may even be transient when the diffusion is exponentially ergodic.

In the second part of this paper we consider rather more delicate approaches to discretization. We will see that the behaviour of the diffusions above can indeed be emulated, leading to an efficient class of practical algorithms.

The $h$-step Euler scheme studied by Roberts and Tweedie [14] simply uses the chain

$$U_h(1) = U_h(0) + \frac{1}{2}h \nabla \log \pi(U_h(0)) + \sqrt{h}Z_n,$$

where $h > 0$ and the random variables $Z_n$ are distributed as independent standard normal distributions; that is,

$$P_U(x, \cdot) \sim N(x + \frac{h}{2} \nabla \log \pi(x), \sqrt{h}).$$

This Euler discretization is known to be rather inappropriate in many circumstances, as is born out in the work in [14]. In particular, Ozaki [12] notes that with the Euler method:

(i) The trajectory of the discrete approximation $U_h$ does not coincide with that of $X(t)$ at the discrete times $t = 0, h, 2h, \ldots$, even for linear drift $b$ and constant $\sigma$;

(ii) If the drift $b$ is a polynomial of degree $> 1$, then the Euler approximation explodes to infinity if it starts from some large initial value, even though the diffusion itself tends to move quickly to some compact interval.

11
We now define a process which will lead to a more appropriate discretization. The suggested approximation does not follow the trajectories of $X(t)$ exactly, of course: but we will show that it has robust convergence properties, especially in comparison to the Euler method.

Consider the discretized chain with transition law $Q_h$ defined by taking, for some $h > 0$,

$$ Q_h(x, \cdot) = N(\mu_{x,h}, \sigma_{x,h}^2) $$

where the parameters starting from $x$ are given by

$$ \mu_{x,h} = x \exp(b(x)h/x) $$

$$ \sigma_{x,h}^2 = \frac{\sigma^2(x)}{2b(x)}[1 - \exp(2b(x)h/x)] $$

where we assume that $b, \sigma^2$ satisfy (4).

The choice of $Q_h$ is based on a time-inhomogeneous diffusion approximation which essentially linearizes the drift over small time intervals, as we now describe. For fixed $h > 0$, we define (in a manner similar to that proposed by Ozaki [12]) the diffusion $D_h(t)$ as a solution to the stochastic differential equation

$$ dD_h(t) = b(t, D_h(t))dt + \sigma(t, D_h(t))dW(t), \quad D_h(0) = x, \quad x \in \mathbb{R} $$

where $W$ is a one dimensional Brownian motion, and we write, for $kh \leq t < (k+1)h$, $k = 0, 1, \ldots$

$$ b(t, D_h(t)) = b(D_h(kh)) \frac{D_h(t)}{D_h(kh)} $$

$$ \sigma(t, D_h(t)) = \sigma(D_h(kh)). $$

The result we then use is

**Proposition 5.1** For $kh \leq t < (k+1)h$, the process $D_h(t)$ given $D_h(kh) = x$, $x \in \mathbb{R}$, is the Ornstein-Uhlenbeck process, and hence for $t = 0, h, 2h, \ldots$ we have that, provided $b(x) \neq 0$, $D_h(t + h)$ given $D_h(t) = x$, $x \in \mathbb{R}$, has Normal distribution with mean and variance given by $\mu_{x,h}, \sigma_{x,h}^2$ as in (35).

**Proof** This is standard: see for instance Example 5.6.8 of [7].

It is clear heuristically that (36) should give a better approximation to $X(t)$ than the Euler scheme: the Euler scheme at $t = 0, h, 2h, \ldots$ is obtained as a solution to (36) with $D_h(t)$ replaced by $D_h(kh)$ for $kh \leq t < (k+1)h$, and thus the approximation $D_h$ assumes linearity of the drift $b$ in a small interval, while the Euler method assumes that the drift is constant in a small interval. Other approaches to this problem will be considered in [17].

Since the chain with transition law $Q_h$ is just the embedded chain of this diffusion approximation, we might expect that it should mimic the behaviour of the diffusion, at least in the situation where $b$ is a low order polynomial.
6 Geometric Convergence of the Discretized Chain

We now consider convergence properties of the chain $D_h$ with transition law $Q_h$, and show that in general it does mimic the behaviour of the diffusion $X(t)$ itself. We do this in a number of steps, since the proofs vary under different conditions on $\pi, b, \sigma^2$. We will usually assume these satisfy (4), although since we are now dealing solely with a discrete time chain, the functions $b, \sigma$ do not need to satisfy the technical conditions needed to ensure that they are appropriate coefficients in a stochastic differential equation.

Note that we typically produce conditions under which there exists $\rho < 1$ such that the chain is geometrically ergodic in the sense that as $n \to \infty$, for all $x$

$$||Q_h^n(x, \cdot) - \pi_h|| \leq M_x \rho^n,$$  \hspace{1cm} (37)

where $\pi_h$ is the stationary measure for $Q_h$. Unless $h$ is small, $\pi_h$ need not be close to $\pi$, of course, although for sufficiently fine discretizations we would hope it would be similar. In [18] we develop ways of using Metropolis algorithms to adjust for this: here we consider the chains $D_h$ without such adjustment.

**Theorem 6.1** Suppose $b, \sigma$ are continuous, and that the conditions (22) and (23) hold. Then the discrete approximation $D_h$ defined as in (34) is geometrically ergodic if $\gamma_b \geq 1$ and either

(i) $\gamma_b - \gamma_s + 1 > 0$; or

(ii) $\gamma_b - \gamma_s + 1 = 0$ and $\alpha_s < 2\alpha_b$;

or if $1 > \gamma_b \geq \frac{\alpha_b}{2} \geq 0$.

**Proof** It is easy to check that $D_h$ is $\mu^\text{Leb}$-irreducible and from Proposition 6.1.2 of [9] it is weak Feller. Hence, from Theorem 15.0.1 of [9] it is suffices for geometric ergodicity to find a test function $V \geq 1$ such that for some compact set $C$ and some $\lambda < 1, b < \infty$

$$\int Q_h(x, dy)V(y) \leq \lambda V(x) + b\|I_h(x).$$ \hspace{1cm} (38)

Under either (i) or (ii), we verify (38) for the test function $V(x) = x^2$, using

$$E([D_h(h)]^2 | D_h(0) = x) = E(\mu_{x,h} + \sigma_{x,h}Z)^2$$

$$= x^2 \exp(2b(x)h/x) + \frac{x^2 \sigma^2(x)}{2b(x)}(1 - \exp(2b(x)h/x),$$

where $Z$ has a standard Normal distribution.

Now if $\gamma_b - \gamma_s + 1 > 0$, so that $\sigma^2(x)/2xb(x) \to 0$ as $x \to \infty$, we have $E([D_h(h)]^2 | D_h(0) = x) \leq \lambda x^2$ for large $|x|$, where $\lambda = \exp(-2\alpha_b h) < 1$. 

If, on the other hand, \( \gamma_s - \gamma_b + 1 = 0 \) and \( \alpha_s < 2\alpha_b \) then \( E([D_h(h)]^2 | D_h(0) = x) \leq \lambda x^2 \) for large \( |x| \), where in this case
\[
\lambda = (1 - \frac{\alpha_s}{2\alpha_b}) \exp(-2\alpha_b h) + \frac{\alpha_s}{2\alpha_b} < 1.
\]
Thus (38) holds in either case and the chain is geometrically ergodic. Finally if \( 1 > \gamma_b \geq \frac{\gamma_s}{2} \geq 0 \) then it can be shown that (38) holds for the test function \( V(x) = \exp(s|x|) \) for some \( s > 0 \); we omit the details. \( \square \)

The conditions of continuity on \( b, \sigma \) can no doubt be weakened. Similarly to the TAR models (see [9]) we can at least generalize Theorem 6.1 to the case that \( b \) and \( \sigma \) are continuous except possibly on a finite number of points, but we do not pursue this here.

We turn instead to conditions that guarantee uniform ergodicity of \( D_h \).

**Theorem 6.2** Assume that the conditions of Theorem 6.1 are satisfied with \( \gamma_b > 1 \) and \( \gamma_b \geq 1 + \gamma_s \). Then the discrete approximation \( D_h \) is uniformly ergodic.

**Proof** It follows from [9, Chapter 16] that \( D_h \) is uniformly ergodic if for some compact set \( C \),
\[
\sup_{x \in C} E_x[\tau_C] < \infty,
\]
where \( \tau_C \) is the first return time to \( C \). Let \( \epsilon > 0 \) be arbitrary and choose \( M > 0 \) (as we can under these conditions) such that

(i) \( P(|Z| < \frac{M-1}{\ell}) \geq 1 - \epsilon \), where \( Z \sim N(0,1) \), and \( \ell = \frac{\alpha_s}{2\alpha_b} \).

(ii) for all \( |x| > M \), \( |\mu_{x,h}| < 1 \) and \( \sigma_{x,h}^2 \leq \ell \).

Let \( C = [-M, M] \) and let \( \sigma P^n(x, B) = P_x(D_h(n) \in B | \tau_C \geq n) \): we now show by induction that \( \sigma P^n(x, C^c) \leq \epsilon^n \) for all \( n \geq 1 \) and all \( |x| > M \). If we write
\[
q(x, y) = \frac{1}{(2\pi \sigma_{x,h}^2)^{1/2}} \exp \left( \frac{\|y - \mu_{x,h}\|^2}{2\sigma_{x,h}^2} \right),
\]
then for all \( |x| > M \)
\[
\int_{y \in C^c} q(x, y) dy = P(|\mu_{x,h} + \sigma_{x,h} Z| > M)
\]
\[
= P(Z > \frac{M - \mu_{x,h}}{\sigma_{x,h}}) + P(Z < \frac{-M - \mu_{x,h}}{\sigma_{x,h}})
\]
\[
\leq P(Z > \frac{M-1}{\ell}) + P(Z < \frac{-M+1}{\ell})
\]
\[
= P(|Z| > \frac{M-1}{\ell}) \leq \epsilon,
\]

14
which proves $cP(x, C^c) \leq \varepsilon$. Suppose now that $cP^n(x, C^c) \leq \varepsilon^n$ for some $n \geq 1$. Then

$$cP^{n+1}(x, C^c) = \int_{y \in C^c} q(x, y)cP^n(y, C^c)dy \leq \varepsilon^n \int_{C^c} q(x, y)dy \leq \varepsilon^{n+1}.$$ 

Thus for all $n \geq 2$ and all $x$,

$$P_x(\tau_C > n) = \int_{y \in C^c} q(x, y)cP^n(y, C^c)dy \leq \varepsilon^n$$

as required; and hence $\sup_{x \in C} E(\tau_C) < \infty$. \hfill \Box

**Remark** Note that for the diffusion we need $\gamma_b > 1$ to guarantee uniform ergodicity but for the discrete approximations $D_h$ we need the extra condition $\gamma_b \geq 1 + \gamma_s$.

**Example: Changing the Discretization**

In Example 1 with $\pi(x) \propto \exp(-\gamma|x|^\beta), |x| \geq x_0$, the behaviour of the discretization used depends largely on $\beta$.

We consider first the Langevin diffusion with $\sigma = 1$, and compare the behaviour of the Euler scheme of discretization used in [14] and the discretization $D_h$ of Theorem 6.1 or Theorem 6.2.

(i) For $0 < \beta < 1$, it follows from [14] that the Euler approximation is transient, and it is not hard to see that $D_h$ is also transient in this case.

(ii) For $1 \leq \beta \leq 2$, $D_h$ is geometrically ergodic for all $h > 0$, while for the Euler approximation this is true except at $\beta = 2$ where it holds only for $\gamma h < 2$.

(ii) For $\beta > 2$, when the tails are light, $D_h$ is uniformly ergodic, while the Euler approximation is transient.

Thus $D_h$ is clearly more effective, even when the Langevin model is used. \hfill \Box

**Example: Changing the Diffusion Model**

We next illustrate the effect of changing the underlying diffusion for $\pi \in \mathcal{E}$ with $0 < \beta < 1$. As noted above, if we choose $\sigma \equiv 1$ then the Euler approximation is transient. If however we choose

$$b(x) = -\frac{\gamma \beta}{2}|x|^{\gamma_s} \text{sgn}(x) + \frac{\gamma_s}{2}|x|^{\gamma_s-1}$$

and $\sigma(x) = |x|^{\gamma_b}$, where $\gamma_b > 1$ and $\gamma_s = \gamma_b + 1 - \beta$, then from Theorem 3.1 the diffusion process $X(t)$ is exponentially ergodic; and moreover, from Theorem 6.1 $D_h$ is also geometrically ergodic.

Similarly, suppose we wish to estimate the $t$-distribution $T_n$ for some $n \geq 3$. If we choose

$$\sigma^2(x) = n + x^2, \quad b(x) = \frac{-(n+1)x}{2} \quad (40)$$

15
then \( \gamma_b = 1, \gamma_b - \gamma_s + 1 = 0 \) and \( \frac{\sigma^2}{2\sigma_b} = \frac{1}{n+1} < 1 \); and so \( D_h \) is geometrically ergodic for all \( h > 0 \).

Thus even when the tails of \( \pi \) are heavy we can ensure that convergence of \( D_h \) is exponential by appropriate choice of the model. \( \square \)

We now carry out these computations in a little more detail in the specific case when \( \pi(x) \propto e^{-x^2}, -\infty < x < \infty \): that is, when \( \pi \) has normal density with zero mean and variance 0.5. We will consider two cases:

Case (i) Here we will implement the Langevin model: choose \( b(x) = -x \) and \( \sigma \equiv 1 \), so that \( \gamma_b = 1 \) and \( \gamma_s = 0 \). From Theorem 6.1 the discrete approximation \( D_h \) is exponentially ergodic.

Case (ii) Secondly, we implement a model with stronger drift from infinity: we choose

\[
 b(x) = \begin{cases} 
-|x|^{1.5} \text{sgn}(x) + 0.25|x|^{-0.5} & \text{if } |x| \geq c \\
-x & \text{if } |x| < c 
\end{cases}
\]

and

\[
 \sigma^2(x) = \begin{cases} 
|x|^{0.5} & \text{if } |x| \geq c \\
1 & \text{if } |x| < c 
\end{cases}
\]

where we take \( c = 4 \) in our computations. Then \( \gamma_b = 1.5, \gamma_s = 0.5 \) and from Theorem 6.2 the discrete approximation \( D_h \) is uniformly (exponentially) ergodic.

We will use the approximating sequence \( D_h \) with \( h = 0.1 \) to show the effect of these different diffusion models. To assess convergence we will estimate \( m(x, t) = E(D_h(t)|D_h(0) = x) \) as a function of the starting point and the number of steps in the chain.

The estimation of the functions \( m(x, t) \) for \( 2 \leq t \leq 12 \) are shown in Figure 1. Note that these correspond to \( 10t \) steps in each case since we have \( h = 0.1 \). We have taken 100,000 iterations in each case to remove sampling error.

As expected the conditional mean of \( D_h(t) \) in Case (ii) converges to its stationary mean much more quickly as a function of \( x \) than in Case (i). Note here that all the initial values are “big”: but Case (ii) also appears to converge much faster as a function of time even when the initial values are small, or when they are even larger than shown here.

7 Subgeometric Convergence of Discretizations

We next consider the situation when we may not have geometric ergodicity of the discretizations, and in view of the results of the previous section, we will consider situations where \( \gamma_b < 1 \) and \( \gamma_b < \frac{\alpha_b}{2} \) in (22).
Figure 1: Conditional mean $m(x,t)$ for $4 \leq t \leq 12$ for Case (i) and Case (ii) with initials points $x=2$, $20$ and $100$.

To discretize in this situation, we can use the Euler scheme, and in this case we show that we get a specific subgeometric rate of convergence. The same argument as in the proof of Proposition 5.2 of [22] gives

$$r(n)||P^n_u(x, \cdot) - \pi||_f \to 0$$

where $P^n_u$ is the transition law of $U_n$ as defined in (32), and

$$f(x) = |x|^{L-r} \vee 1 \quad x \in \mathbb{R},$$

$$r(n) = n^{L-k} \vee 1 \quad n \geq 1;$$

here $L$ is any positive fractional index if $\gamma_b - \gamma_s + 1 > 0$ and $L = \frac{2\alpha_b}{\alpha_s}$ if $\gamma_b - \gamma_s + 1 = 0$.

This shows that if we choose $\gamma_b$ and $\gamma_s$ such that $\gamma_b - \gamma_s + 1 > 0$ then we get convergence in total variation at a rate that is certainly faster than any fixed polynomial: only in the case $\gamma_b - \gamma_s + 1 = 0$ do we get fixed polynomial convergence.

Example 6 (t distribution with 5 degrees of freedom) Let us take $\pi$ as $T_n$ with $n = 5$. For the models below we choose $h = 0.1$, and 100,000 replications, and consider the convergence in time $t$ of the estimates of the conditional variances $v(x, t) = V(U_n(t)|U_n(0) = x)$ for various initial values $x$.

We first choose $\sigma(x) \equiv 1$ and $b(x) = \frac{\nu(x+1)x}{2(\nu+1)}$ (i.e. $\alpha_b = \frac{n+1}{2}, \alpha_s = 1, \gamma_b = -1$ and $\gamma_s = 0$); then we do not have exponential rate of convergence for the Euler scheme, and in Figure 2 we show this behaviour for initial points $x = 2, 6$. As expected the conditional variances of $U_n(t)$ converge rather slowly to the stationary variance of the
process $X(t)$ when $x$ is far from the “center” and much more rapidly when $x$ is close to the “center”.

Similar but more excessive behaviour occurs for heavier tailed distributions: for the $t$-distribution with 3 d.f. convergence is slow even from central starting points.

We next consider the approximating sequence $D_h$ based on the diffusion $\sigma^2(x) = n + x^2$, $b(x) = -\frac{1}{2}(n + 1)x$ as in (40). The estimation of the function $v(x,t)$ for $x = 2, 6, 20$ is shown in Figure 3. As expected the conditional variance of $D_h(t)$ converges radially, and virtually to the stationary variance $\frac{n}{n-2} = 1.66$ of the process $X(t)$: that is, in this case the fact that the stationary distribution of $D_h$ is different from the stationary distribution of the process itself does not appear to be a major problem, at least for variance estimation.

We finally consider the approximating sequence $D_h$ based $\sigma^2(x) = (n + x^2)^{1.5}$ and $b(x) = -1.5x(n + x^2)^{0.5}$, and consider changing the step size $h$. The estimation of the function $v(x,t)$ for $x = 2, 6, 20$ and $h = 0.1, 0.02, 0.01$ is shown in Figure 4. In this case the conditional variance of $D_h(t)$ converges, at exponential rate, to its stationary variance. However, if $h$ is not “small enough” then the stationary variance of the discrete approximation is different from the stationary variance of the diffusion even though the discretization itself converges well. This behaviour is almost identical for starting points much higher than 2.

![Graph](image)

Figure 2: Conditional variance $v(x,t)$ $0 \leq t \leq 40$ for Example 6 with $x = 2$ and $20$ and $h = 0.1$. 

18
Figure 3: *Conditional variance* $v(x,t) \ 0 \leq t \leq 6$ for Example 6 with $x = 2, 6$ and 20 and $h = 0.1$.

Figure 4: *Conditional variance* $v(x,t) \ 0 \leq t \leq 14$ for Example 6 with $x = 2, 6$ and 20.
8 Choosing the Discretization Procedure

Under the conditions of Theorem 3.1, the convergence properties of the approximations $D_h$ to $X(t)$ mimic those of the diffusion. In this section, we justify the asymptotic use of these approximations, since in principle they only approximate the diffusion process on some finite interval of time $[0,T]$.

We propose the following scheme. Let $G_h(t)$ denote the approximation to $X(t)$: in some cases this will be the Euler scheme $U_h$ and in others the discretization $D_h$. Then

1. Fix $t$ and choose $h$ small so that for some constants $L_1$ and $L_2$, we have $G_{h(t)}(t)$ and $G_{h(t)}(t)$ close in some sense for all $h' = [h(t)^{-1} + k]^{-1}$, where $L_1 \leq k \leq L_2$;

2. Increase $t$ and repeat step 1 until $G_{h(t)}(t)$ is “close” to the stationary distribution and $G_{h(t)}(t)$, $G_{h(t)}(t + k)$ are close in some sense for all $h' = [h(t)^{-1} + k]^{-1}$, where $L_1 \leq k \leq L_2$, and all $K_1 \leq k \leq K_2$ for some constants $K_1$ and $K_2$.

We now have:

**Theorem 8.1** Assume that

(i) the conditions of Theorem 3.1 hold;

(ii) $\lim_{t \to \infty} \|P_{X}^{t}(x, \cdot) - \pi\|_V = 0$ for $V \geq 1$ such that $V$ is continuous and $|V(x)| \leq K(1 + |x|^\kappa)$ for some $K > 0$, $\kappa \geq 0$ and all $x \in \mathbb{R}$.

Let $G_h$ be the discrete approximation to the diffusion process ($U_h$ if $\gamma_b < 1$ and $D_h$ if $\gamma_b \geq 1$). Then for all $x \in \mathbb{R}$,

$$\lim_{t \to \infty} \lim_{h \to 0} \|P_{G_h}^{t}(x, \cdot) - \pi\|_V = 0,$$

(41)

if $V$ is bounded or $\gamma_b \leq 2$.

**Proof** We first show that $E(V(G_h(t))|G_h(0) = x)$ converges as $h \to 0$ to $E(V(X(t))|X(0) = x)$ for any $t > 0$. If $V$ is bounded or $\gamma_b \leq 1$ then by Problem 5.3.15 of [7] $\sup_h E[|G_h(t)|^{j+\varepsilon}] < \infty$, for any positive integer $j$ and any $\varepsilon > 0$ and the proof follows directly from Theorem 25.12 of [3]. If $\gamma_b > 1$ then for every $\delta > 0$ and $h > 0$ there exists $M > 0$ such that for all $|x| > M$, we have $\mu_{x,h} < \delta$. Let

$$b^*(x) = \begin{cases} \frac{b(x)}{\beta(x)} & \text{if } |x| \leq M \\ \frac{b(x)}{\beta(x)} & \text{if } |x| > M \end{cases}$$

and $G_h^*$ be the discrete approximation for $X(t)$ with coefficients $b^*$ and $\sigma$. In this case $\gamma_b = 1$ and hence for any $t > 0$, $\varepsilon > 0$ and any positive integer $j$, $\sup_h E[|G_h^*(t)|^{j+\varepsilon}] < \infty$. It is easy to check that if $G_h$ is the discrete approximation for $X(t)$ with coefficients
$b$ and $\sigma$ then for every $h > 0$ and $t > 0$ $P(|G_h(t)| \leq M) \geq P(|G_h^* (t)| \leq M)$ and hence $\sup_h E[|G_h^*|^+ ] < \infty$. Once again the proof follows directly from Theorem 25.12 of [3]. Hence for every $t > 0$ and $x \in \mathbb{R}$,

$$\lim_{h \to 0} |E_x[V(G_h(t))] - E_x[V(X(t))]| = 0,$$

and hence for each $x \in \mathbb{R}$,

$$\lim_{t \to \infty} \lim_{h \to 0} \| P^t_{G_h}(x, \cdot ) - P^t_X((x, \cdot )) \| V = 0. \quad (42)$$

We now conclude from (ii) and (42) that

$$\lim_{t \to \infty} \lim_{h \to 0} \| P^t_{D_h}(x, \cdot ) - \pi \| V = 0.$$

\[\square\]

**Remark** In the case that $V$ is not bounded it is not clear if (41) holds for $\gamma_s > 2$. The value of $h$ needed to obtain a good approximation to the stationary distribution might be very small (see Figure 4), and hence one needs to be careful in using this scheme. In general it makes clear sense to "Metropolise" the discrete chain as in [14], and we address this in [18].

**Acknowledgement**

Related results have recently been developed independently by Roberts and Rosenthal [13]. In particular they consider use of the diffusion satisfying (4) in the special case where $\pi$ is in $\mathcal{E}$ and $b(x)$ is asymptotically linear, and they establish convergence results similar to those in Theorem 3.1 for Metropolis-adjusted discretizations in this case: such Metropolis-adjusted discretizations are studied in detail in [18].

**References**


