Bernoulli **9**(4), 2003, 725–743

Geometric ergodicity of discrete-time approximations to multivariate diffusions

NIELS RICHARD HANSEN

Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark. E-mail: richard@stat.ku.dk

A discrete-time approximation scheme called local linearization of the Langevin diffusion on $\mathbb{R}^k$ is considered, with emphasis on the ergodic properties of the approximation considered as a discrete-time Markov chain. We will derive criteria for the scheme to be geometrically ergodic, and illustrate the use of these criteria by means of examples. Furthermore, we discuss the scheme in relation to other schemes and the use of such discretization schemes as proposals in a Metropolis–Hastings algorithm.

Keywords: geometric drift; geometric ergodicity; Langevin diffusions; Markov chain Monte Carlo; Markov chains; stochastic differential equations

1. Introduction

In this paper we consider discrete-time approximation schemes for the solution of a stochastic differential equation (SDE), and we will focus on the properties of such approximation schemes as discrete-time Markov chains. We will then be able to discuss aspects of the schemes related to their numerical stability and to discuss whether the schemes have the same qualitative behaviour as the diffusion itself.

We will restrict ourselves to SDEs of the form

$$\mathrm{d}X(t) = -\nabla H(X(t))\mathrm{d}t + \sigma \mathrm{d}W(t) \tag{1}$$

where $(W(t))_{t\geqslant 0}$ is a k-dimensional Brownian motion, $\sigma > 0$ and $H : \mathbb{R}^k \to \mathbb{R}$ is a potential. In the literature on approximations of the solution to SDEs (such as Kloeden and Platen 1992) the focus is on developing the solution of the SDE using stochastic Taylor expansions, usually leading to rough approximations of the term $\nabla H(X(t))\mathrm{d}t$ as a constant over small time intervals. We will consider a better approximation of the term $\nabla H(X(t))\mathrm{d}t$, which is called local linearization (Biskey *et al*. 1996; Ozaki 1992; Stramer and Tweedie 1999a). Another approach is to consider implicit schemes instead (Mattingly *et al*. 2002).

One of the reasons for looking at SDEs of the form (1) is that they appear as continuous-time analogues of the many discrete-time Markov chain Monte Carlo (MCMC) algorithms considered in, for instance, Roberts and Tweedie (1996a) and Stramer and Tweedie (1999a). Furthermore, they appear as simple models of Brownian particles in a force-field given by the potential H – at least on a sufficiently large time scale – as described in Section 6.4 in Gardiner (1983).

1350–7265

We will in this paper focus on the ergodic properties of the local linearization scheme as a discrete-time Markov chain, and in particular we will derive criteria for geometric ergodicity, our general results being Theorems 4.3 and 4.5 In Theorem 4.10 we use the general results to verify geometric ergodicity for a large class of potentials. We will also compare the results obtained with similar results related to the classical forward Euler discretization scheme and some implicit schemes such as backward Euler. Finally, we will discuss the use of these discretization schemes as proposal distributions in the Metropolis–Hastings algorithm.

2. Langevin diffusions and approximation schemes

In this section discrete-time approximation schemes for the solution of the kind of stochastic differential equations described in Section 1 are considered.

Definition 2.1. *Let $H : \mathbb{R}^k \to \mathbb{R}$ be a differentiable function. The Langevin diffusion with potential H is then the solution $(X(t))_{t\geqslant 0}$ to the Langevin equation*

$$\mathrm{d}X(t) = -\nabla H(X(t))\mathrm{d}t + \sigma\,\mathrm{d}W(t), \tag{2}$$

where $(W(t))_{t\geqslant 0}$ is a k-dimensional Brownian motion and $\sigma > 0$.

The name 'Langevin' is used in, for instance, Mattingly *et al*. (2002) and also in the physics literature for another SDE, but in the statistical literature such as Roberts and Tweedie (1996a) it seems that 'Langevin equation' and 'Langevin diffusion' have been reserved for the case given by (2). We use the word 'Langevin' in this last sense. One should especially note that Mattingly *et al*. (2002) call equations given by (2) gradient systems instead.

We are implicitly assuming that there is a solution to (2), and, as proved by Roberts and Tweedie (1996a, Theorem 2.1), that there exists a continuous, non-explosive solution to (2) if H is continuously differentiable, if $x \mapsto \exp(-H(x))$ is integrable and if there are constants α and β such that

$$-\langle \nabla H(x), x\rangle \leqslant \beta|x|^2 + \alpha. \tag{3}$$

Furthermore, the solution is μ^{Leb}-irreducible, aperiodic, the compact sets are small and the solution has an invariant initial distribution with a density π satisfying $\pi(x) \propto \exp(-(2/\sigma)H(x))$. Finally, if Q^t, for $t > 0$, denotes the transition probabilities for the solution, then

$$\|Q^t(x,\cdot) - \pi\|_{tv} \to 0$$

for $t \to \infty$.

If π is a strictly positive, twice continuously differentiable density on $\mathbb{R}^k$, then if $H(x) = -\frac{1}{2}\log \pi(x)$ satisfies (3), the Langevin diffusion with potential H and $\sigma = 1$ has π as invariant initial distribution. Note that if π is only known up to a constant of proportionality, ∇H is completely known and hence the Langevin equation is completely known. Therefore the Langevin diffusion can be viewed as a continuous-time analogue of the many discrete-time MCMC algorithms.

A simple approximation scheme is the Euler scheme, where we assume that ∇H is almost constant in small time intervals of length $h > 0$ and this leads to an approximating Markov chain in discrete-time at the times $0, h, 2h, 3h, \ldots$ with transition probabilities

$$Q_0(x, \cdot) = N(x - h\nabla H(x), \sigma^2 hI).$$

Roberts and Tweedie (1996a) demonstrate that this kind of approximation is quite ill behaved in an MCMC set-up if the tails of π are decaying too rapidly, that is, if the potential is increasing too fast. Indeed, even if π is a Gaussian distribution, one has to choose h sufficiently small to avoid the approximation becoming transient – in contrast to the behaviour of the Langevin diffusion itself. Mattingly *et al*. (2002) also give examples of loss of ergodicity for the Euler scheme.

To avoid the problems with the Euler scheme we will consider the local linearization of the diffusion instead. The local linearization scheme can be derived based on the solution of the Ornstein–Uhlenbeck diffusion.

For $A \in \mathbb{R}^k$ and $B, C \in M(k)$ (the set of $k \times k$ matrices) the k-dimensional Ornstein–Uhlenbeck diffusion is the solution $(Y(t))_{t \geqslant 0}$ to the SDE

$$\mathrm{d}Y(t) = (A + BY(t))\mathrm{d}t + C\mathrm{d}W(t), \tag{4}$$

where $(W(t))_{t \geqslant 0}$ is a k-dimensional Brownian motion. The solution of this diffusion is well known and one can give an explicit expression in terms of stochastic integrals (Ikeda and Watanabe 1989, Section 4.8). The Gaussian transition probabilities of the solution can be found explicitly:

$$Q^t(x, \cdot) = N(\xi(x, t), \Sigma(t)), \tag{5}$$

where

$$\xi(x, t) = \int_0^t \exp(sB)A\,\mathrm{d}s + \exp(tB)x$$

$$\Sigma(t) = \int_0^t \exp(sB)CC^{\mathrm{T}} \exp(sB^{\mathrm{T}})\mathrm{d}s.$$

See, for instance, Jacobsen (1991; 1994, Theorem 2.2).

If H is twice continuously differentiable we can use the best affine approximation of ∇H instead of a constant approximation, that is, a first-order Taylor approximation. Thus with $\nabla^2 H(x)$ the matrix of second-order partial derivatives at x, approximate $X(t)$ by the process $\tilde{X}(t)$, which on $[kh, (k+1)h]$, given $\tilde{X}(kh) = x$, is the solution to

$$\mathrm{d}\tilde{X}(t) = (-\nabla H(x) - \nabla^2 H(x)(\tilde{X}(t) - x))\mathrm{d}t + \sigma\mathrm{d}W(t).$$

On the interval $[kh, (k+1)h]$ this is an Ornstein–Uhlenbeck diffusion, and the transition probabilities for the discrete-time chain $(\tilde{X}(kh))_{k \in \mathbb{N}_0}$ become

$$Q_1(x, \cdot) = N(\xi(x, h), \Sigma(x, h)),$$

where

$$\xi(x,\ h) = \int_0^h \exp(-s\nabla^2 H(x))\mathrm{d}s(\nabla^2 H(x)x - \nabla H(x)) + \exp(-h\nabla^2 H(x))x$$

$$= x - \int_0^h \exp(-s\nabla^2 H(x))\mathrm{d}s\nabla H(x),$$

$$\Sigma(x,\ h) = \sigma^2 \int_0^h \exp(-2s\nabla^2 H(x))\mathrm{d}s.$$

Definition 2.2. *A Markov chain with transition probabilities* Q_1 *is called a local linearization of the Langevin diffusion with time-step* h.

The local linearization presented here and other similar schemes have been considered by Ozaki (1992) and by Biskey *et al*. (1996). Furthermore, Stramer and Tweedie (1999a) consider the scheme in an MCMC set-up primarily in the one-dimensional case.

3. Geometric ergodicity of Gaussian transition probabilities

Markov chains on $\mathbb{R}^k$ with Gaussian transition probabilities like those resulting from both the Euler scheme and the local linearization scheme can be treated quite generally. Thus we consider Markov kernels

$$Q(x,\ \cdot) = N(\xi(x),\ \Sigma(x)) \tag{6}$$

with $\xi : \mathbb{R}^k \to \mathbb{R}^k$ and $\Sigma : \mathbb{R}^k \to \mathrm{PD}(k)$ being measurable maps – here $\mathrm{PD}(k)$ denotes the set of positive definite $k \times k$ matrices. That is, $Q(x,\ \cdot)$ is a regular Gaussian distribution with mean value $\xi(x)$ and variance $\Sigma(x)$, and the two maps ξ and Σ will be called the *mean value map* and the *variance map*, respectively. Assume in the following that ξ as well as Σ *map compact sets into compact sets*, which is the case if, for instance, ξ and Σ are continuous.

Theorem 3.1. *The Markov kernel* Q *given by* (6) *is* μ^{Leb}*-irreducible and aperiodic, with* μ^{Leb} *being the Lebesgue measure, and the compact sets are small.*

Proof. The Markov kernel has density

$$q(x,\ y) = \det(\Sigma(x)^{-1/2}) f(\Sigma(x)^{-1/2}(y - \xi(x))) \tag{7}$$

with respect to μ^{Leb}, where f is the density for the normalized Gaussian distribution. Therefore $q(x,\ y) > 0$ for all $(x,\ y) \in \mathbb{R}^k \times \mathbb{R}^k$, in which case it is well known that Q is irreducible and aperiodic. Moreover, q can in fact be bounded below by a strictly positive constant on a compact set, which proves that the compact sets are small. □

We will give a general condition for geometric ergodicity based on the drift function $1+|x|^2$ – see the Appendix – so put

$$V_2(x) = 1 + |x|^2$$

from here onwards. The following lemma and corollary hold for more general Markov kernels on $\mathbb{R}^k$, but we state the result for the Gaussian ones only.

Lemma 3.2. *A Markov kernel Q given by* (6) *has V_2-geometric drift towards a compact set if and only if*

$$\limsup_{|x|\to\infty} \frac{|\xi(x)|^2 + \mathrm{tr}(\Sigma(x))}{|x|^2} < 1. \tag{8}$$

Proof. Calculation of the conditionally expected drift leads to

$$QV_2(x) = 1 + |\xi(x)|^2 + \mathrm{tr}(\Sigma(x)).$$

Assume that

$$\alpha := \limsup_{|x|\to\infty} \frac{|\xi(x)|^2 + \mathrm{tr}(\Sigma(x))}{|x|^2} < 1,$$

and choose $\beta < 1$ with $\alpha < \beta$. For some suitable compact set C we obtain

$$\begin{aligned} QV_2(x) &= 1 - \frac{|\xi(x)|^2 + \mathrm{tr}(\Sigma(x))}{|x|^2} + \frac{|\xi(x)|^2 + \mathrm{tr}(\Sigma(x))}{|x|^2} V_2(x) \\ &\leqslant \beta V_2(x) + b 1_C(x), \end{aligned}$$

with $b = \sup_{x\in C} 1 + |\xi(x)|^2 + \mathrm{tr}(\Sigma(x)) < \infty$. If, on the other hand,

$$QV_2(x) = 1 + |\xi(x)|^2 + \mathrm{tr}(\Sigma(x)) \leqslant \beta V_2(x),$$

for $|x| \geqslant R$ and some $\beta < 1$, then it follows immediately that

$$\limsup_{|x|\to\infty} \frac{|\xi(x)|^2 + \mathrm{tr}(\Sigma(x))}{|x|^2} \leqslant \beta < 1.$$

□

As a consequence of Lemma 3.2, Theorem 3.1 and Theorem A.3, we have the following corollary.

Corollary 3.3. *If* (8) *holds, then Q has a unique invariant probability measure and is V_2-geometrically ergodic.*

It should also be noticed that if Σ is a bounded map or if just $\mathrm{tr}(\Sigma(x)) = o(|x|^2)$, then (8) is equivalent to

$$\limsup_{|x|\to\infty} \frac{|\xi(x)|^2}{|x|^2} < 1. \tag{9}$$

Henceforth we will put $n(x) = x/|x|$ for $x \neq 0$.

Lemma 3.4. *Let* $K(x) = \xi(x) - x$. *Then* (9) *is satisfied if and only if*

$$\limsup_{|x|\to\infty} \left\langle \frac{K(x)}{|x|}, \frac{K(x)}{|x|} + 2n(x) \right\rangle < 0 \tag{10}$$

Proof. We have that $\xi(x) = K(x) + x$, hence

$$\begin{aligned} \frac{|\xi(x)|^2}{|x|^2} &= \frac{|K(x)|^2 + |x|^2 + 2\langle K(x), x\rangle}{|x|^2} \\ &= \frac{|K(x)|^2}{|x|^2} + 1 + 2\left\langle \frac{K(x)}{|x|}, n(x) \right\rangle \\ &= 1 + \left\langle \frac{K(x)}{|x|}, \frac{K(x)}{|x|} + 2n(x) \right\rangle, \end{aligned}$$

and the result follows. □

4. Geometric ergodicity of the local linearization

The Langevin diffusion can itself be geometrically ergodic (Roberts and Tweedie 1996a, Theorems 2.2 and 2.3). We will give another (very simple) criterion for geometric ergodicity for the diffusion, which is based on the drift function V_2. We will assume that the conditions stated in Section 2 to ensure ergodicity are fulfilled. First note that the generator of the diffusion is given by

$$\mathcal{A}f(x) = -\langle \nabla H(x), \nabla f(x) \rangle + \frac{\sigma}{2}\Delta f(x)$$

for a suitable test function f. Then

$$\mathcal{A}V_2(x) = -2\langle \nabla H(x), x\rangle + k\sigma, \tag{11}$$

and since ∇H is continuous, we have that

$$\mathcal{A}V_2(x) \leqslant -cV_2(x) + b$$

for some $c, b > 0$ if

$$\liminf_{|x|\to\infty} \frac{\langle \nabla H(x), x\rangle}{|x|^2} > 0. \tag{12}$$

The drift function V_2 tends to infinity for $|x| \to \infty$, and the compact sets are small, hence it follows from Down, *et al*. (1995, Theorem 5.2) that if (12) holds the diffusion is

V_2-geometrically ergodic, that is to say, with Q^t denoting the transition probabilities, we have that

$$\|Q^t(x, \cdot) - \pi\|_{V_2} \leqslant RV_2(x)\rho^t$$

for some constants R and $\rho < 1$. The criterion (12) is simpler than the similar criterion in Roberts and Tweedie (1996a, Theorem 2.3), which is based on the drift function $\exp(dH(x))$ for some $d \in]0, 1[$.

The question is now to what extent the local linearization preserves geometric ergodicity. If $H: \mathbb{R}^k \to \mathbb{R}$ is a twice continuously differentiable potential, we have seen in Section 2 that the local linearization with time-step $h > 0$ of the Langevin diffusion is a discrete-time Markov chain with transition probabilities Q_1 of the form (6) with mean value map

$$\xi_1(x) = x - \int_0^h \exp(-s\nabla^2 H(x))\mathrm{d}s\nabla H(x)$$

and variance map

$$\Sigma_1(x) = \sigma^2 \int_0^h \exp(-2s\nabla^2 H(x))\mathrm{d}s,$$

suppressing the dependence on h in the notation. As in Lemma 3.1, we also put

$$K_1(x) = -\int_0^h \exp(-s\nabla^2 H(x))\mathrm{d}s\nabla H(x),$$

and we will use $\lambda_{\min}(x)$ to denote the minimum eigenvalue of the symmetric matrix $\nabla^2 H(x)$.

Observe that ξ_1 as well as Σ_1 are continuous, hence, from Theorem 3.1, Q_1 is μ^{Leb}-irreducible, aperiodic and the compact sets are small.

Theorem 4.1. *The variance map Σ_1 is bounded if and only if $\lambda_{\min}(x)$ is bounded below.*

Proof. For a given $x \in \mathbb{R}^k$, let $\{\lambda_1(x), \ldots, \lambda_k(x)\}$ be the eigenvalues of $\nabla^2 H(x)$. Then it follows that the eigenvalues of $\Sigma_1(x)$ are

$$\left(\sigma^2 \int_0^h \exp(-2s\lambda_1(x))\mathrm{d}s, \ldots, \sigma^2 \int_0^h \exp(-2s\lambda_n(x))\mathrm{d}s\right),$$

and since $\Sigma_1(x)$ is symmetric $\|\Sigma_1(x)\|$ is equal to the spectral radius, that is, to the numerically largest eigenvalue, hence

$$\|\Sigma_1(x)\| = \sigma^2 \int_0^h \exp(-2s\lambda_{\min}(x))\mathrm{d}s,$$

and the claim follows. □

If the variance map is bounded, we can restrict our attention to the mean value map exclusively – at least in questions concerning geometric ergodicity. And due to Lemma 3.4 we will focus on the map K_1. If $\nabla^2 H(x)$ is invertible, it furthermore follows that

$$K_1(x) = (\exp(-h\nabla^2 H(x)) - I)(\nabla^2 H(x))^{-1}\nabla H(x).$$

The following definition will turn out to be useful.

Definition 4.2. *A map $G : \mathbb{R}^k \to \mathbb{R}^k$ is said to be asymptotically homogeneous to x for $|x| \to \infty$ if there exists a function*

$$c : S^{k-1} \to \mathbb{R}$$

such that

$$\left| \frac{G(x)}{|x|} - c(n(x))n(x) \right| \to 0$$

for $|x| \to \infty$. The function c is called the asymptotic function.

We are now able to give a general condition for Q_1 to be geometrically ergodic. First, we make the following useful but trivial observation. If B is a matrix, $A = B - I$ and $|x| = 1$, then

$$|Ax|^2 + 2\langle Ax, x\rangle \leqslant \|B\|^2 - 1 \tag{13}$$

since

$$\begin{aligned} |Ax|^2 + 2\langle Ax, x\rangle &= |Bx - x|^2 + 2\langle Bx, x\rangle - 2|x|^2 \\ &= |Bx|^2 + |x|^2 - 2\langle Bx, x\rangle + 2\langle Bx, x\rangle - 2|x|^2 \\ &\leqslant \|B\|^2|x|^2 + 1 - 2 \\ &= \|B\|^2 - 1. \end{aligned}$$

Theorem 4.3. *Let H: $\mathbb{R}^k \to \mathbb{R}$ be a twice continuously differentiable potential. If*

$$\liminf_{|x|\to\infty} \lambda_{\min}(x) > 0$$

and if

$$\nabla^2 H(x)^{-1}\nabla H(x)$$

is asymptotically homogeneous to x for *$|x| \to \infty$ with the asymptotic function c satisfying*

$$0 < \inf_{|x|=1} c(x) \leqslant \sup_{|x|=1} c(x) \leqslant 1,$$

then Q_1 has V_2-geometric drift towards a compact set and is V_2-geometrically ergodic.

Proof. Start by choosing $\alpha > 0$ and $R \in [0, \infty[$ such that

$$\lambda_{\min}(x) \geqslant \alpha$$

for $|x| \geqslant R$. We will in the following only consider $x \in \mathbb{R}^k$ with $|x| \geqslant R$. Then all the eigenvalues of $\nabla^2 H(x)$ are strictly positive and $\nabla^2 H(x)$ is invertible. If we put $T_1(x) = \exp(-h\nabla^2 H(x)) - I$ and $G_1(x) = \nabla^2 H(x)^{-1}\nabla H(x)$, then

$$K_1(x) = T_1(x)G_1(x)$$

and we obtain

$$\begin{aligned}\left\langle \frac{K_1(x)}{|x|}, \frac{K_1(x)}{|x|} + 2n(x)\right\rangle &= \left| T_1(x)\frac{G_1(x)}{|x|}\right|^2 + 2\left\langle T_1(x)\frac{G_1(x)}{|x|}, n(x)\right\rangle \\ &= |T_1(x)n(x)|^2 c(n(x))^2 + 2\langle T_1(x)n(x), n(x)\rangle c(n(x)) + R(x),\end{aligned}$$

with the residual term $R(x) \to 0$ for $|x| \to \infty$. To see this, note that $R(x)$ is a sum of two differences, one of them being

$$\begin{aligned}\left| T_1(x)\frac{G_1(x)}{|x|}\right|^2 - |T_1(x)n(x)|^2 c(n(x))^2 &\leqslant \left| T_1(x)\left(\frac{G_1(x)}{|x|} - c(n(x))n(x)\right)\right|^2 \\ &\quad + |c(n(x))|\,|T_1(x)n(x)|\left| T_1(x)\left(\frac{G_1(x)}{|x|} - c(n(x))n(x)\right)\right|,\end{aligned}$$

which tends to 0 for $|x| \to \infty$, since T_1 and c are bounded. The other difference is seen to tend to 0 for $|x| \to \infty$ by a similar argument.

If we let $\gamma = \exp(-h\alpha) < 1$, then since $\exp(-h\nabla^2 H(x)) \leqslant \gamma I$ and since $0 < \inf_{|y|=1} c(y) \leqslant c(n(x)) \leqslant 1$, it follows that

$$\begin{aligned}|T_1(x)n(x)|^2 c(n(x))^2 + 2\langle T_1(x)n(x), n(x)\rangle c(n(x)) &\leqslant \big(|T_1(x)n(x)|^2 + 2\langle T_1(x)n(x), n(x)\rangle\big) c(n(x)) \\ &\leqslant (\gamma^2 - 1)c(n(x)) \\ &\leqslant (\gamma^2 - 1)\inf_{|y|=1} c(y),\end{aligned}$$

where the second inequality follows from (13). Therefore we have that

$$\limsup_{|x|\to\infty}\left\langle \frac{K_1(x)}{|x|}, \frac{K_1(x)}{|x|} + 2n(x)\right\rangle \leqslant (\gamma^2 - 1)\inf_{|y|=1} c(y) < 0,$$

and from Lemmas 3.2 and 3.4 we have that Q_1 has V_2-geometric drift towards a compact set, and since the compact sets are small, Q_1 is V_2-geometrically ergodic. □

The theorem above will be used in the next section to show that Q_1 is geometrically ergodic for a special class of potentials.

Remark 4.4. *Geometric interpretation*. That the matrix $\nabla^2 H(x)$ should become positive definite as $|x| \to \infty$ has a geometric interpretation in terms of the contour manifolds of H becoming convex as $|x| \to \infty$. Assume that $H(x) \to \infty$ for $|x| \to 0$ and that H has no stationary points outside some ball $B(0, R)$, that is, that $\nabla H(x) \neq 0$ for $|x| \geqslant R$. Then by choosing $\rho > 0$ large enough, the contour manifold

$$C_\rho = \{x \in \mathbb{R}^k | H(x) = \rho\}$$

satisfies $C_\rho \cap B(0, R) = \varnothing$, and C_ρ has a normal field N: $C_\rho \to S^{k-1}$ defined by

$$N(x) = \frac{\nabla H(x)}{|\nabla H(x)|}.$$

Differentiation yields

$$\nabla N(x) = \frac{\nabla^2 H(x)}{|\nabla H(x)|} - \frac{N(x)}{|\nabla H(x)|}(\nabla|\nabla H(x)|)^{\mathrm{T}}.$$

If T_x denotes the tangent hyperplane to C_ρ at x, then, for $v \in T_x$,

$$\langle \nabla N(x)v, v\rangle = \frac{1}{|\nabla H(x)|}\langle \nabla^2 H(x)v, v\rangle$$

since $v \perp N(x)$. Now $\nabla N(x)$ restricted to T_x is a symmetric operator, with the eigenvalues being the principal curvatures of C_ρ at x, and if we assume that $\nabla^2 H(x)$ is positive definite for $|x|$ large enough, all the principal curvatures are strictly positive. It follows that locally C_ρ lies exclusively on one side of T_x, and hence that C_ρ is convex at x. Strict positivity of the principal curvatures actually implies that C_ρ is strictly convex at x, that is, that there is a neighbourhood O of x in C_ρ, such that $O \cap T_x = \{x\}$.

If $\lambda_{\min}(x) \to \infty$ for $|x| \to \infty$ instead of just being bounded away from 0 it is actually possible to prove geometric ergodicity with a less restrictive claim on the asymptotic function c.

Theorem 4.5. *Let H: $\mathbb{R}^k \to \mathbb{R}$ be a twice continuously differentiable potential. If*

$$\lambda_{\min}(x) \to \infty$$

and if

$$\nabla^2 H(x)^{-1}\nabla H(x)$$

is asymptotically homogeneous to x for $|x| \to \infty$ with the asymptotic function c satisfying

$$0 < \inf_{|x|=1} c(x) \leqslant \sup_{|x|=1} c(x) < 2,$$

then Q_1 has V_2-geometric drift towards a compact set and is V_2-geometrically ergodic.

Proof (sketch). One can use the fact that $\lambda_{\min}(x) \to \infty$ for $|x| \to \infty$ to show that K_1 is asymptotically homogeneous to x for $|x| \to \infty$ with asymptotic function c if and only if $\nabla^2 H(x)^{-1}\nabla H(x)$ is asymptotically homogeneous to x for $|x| \to \infty$ with asymptotic function $-c$, and one can easily show that since

$$-2 < \inf_{|x|=1} - c(x) \leqslant \sup_{|x|=1} - c(x) < 0$$

and since $\xi_1(x) = K_1(x) + x$, this gives that

$$\limsup_{|x|\to\infty} \frac{|\xi_1(x)|}{|x|} < 1,$$

in which case (9) is satisfied, and the claim follows from Corollary 3.3 □

4.1. A polynomial class of potentials

To illustrate the use of Theorem 4.3 in the previous section, we will consider a class of potentials on $\mathbb{R}^k$ having a polynomial form.

Definition 4.6. *Let $\mathcal{P}_d$ be the class of potentials H on $\mathbb{R}^k$, which can be written as*

$$H(x) = p(x) + r(x),$$

where p is a homogeneous polynomial of degree d and r is twice continuously differentiable and satisfies

$$\|\nabla^2 r(x)\| = o(|x|^{d-2})$$

for $|x| \to \infty$. Let $\mathcal{P}_d^+$ be the subset of $\mathcal{P}_d$ for which $\nabla^2 p(x)$ is also positive definite for all x.

Remark 4.7. If $r : \mathbb{R}^k \to \mathbb{R}$ satisfies $\|\nabla^2 r(x)\| = o(|x|^a)$ for $|x| \to \infty$ with $a \geqslant 0$, then it follows that

$$|\nabla r(x)| = o(|x|^{a+1}),$$

$$|r(x)| = o(|x|^{a+2}).$$

We skip the proof since it is elementary, but note that it is crucial that the restriction on the growth of r is on the second-order derivative and not on r itself, since it is impossible to bound the derivative of a function in terms of the function itself.

Recall that we can write a homogeneous polynomial p of degree d as a sum of monomials, that is,

$$p(x) = \sum_{|\alpha|=d} c_\alpha x^\alpha,$$

where we have $c_\alpha \in \mathbb{R}$. Here we have also used the multi-index notation: $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$, $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and $x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$. Furthermore, the homogeneous polynomials are completely determined by their values on the unit sphere since, for $x \neq 0$,

$$p(x) = |x|^d p(n(x))$$

if p has degree d. Rather trivial calculations show that for p a homogeneous polynomial of degree d, we have that

$$\langle \nabla p(x), x \rangle = d\, p(x), \tag{14}$$

$$\nabla^2 p(x) x = (d-1) \nabla p(x), \tag{15}$$

$$\langle \nabla^2 p(x) x, x \rangle = d(d-1)\, p(x). \tag{16}$$

If p is a homogeneous polynomial of degree d, it follows that ∇p is a vector of homogeneous polynomials of degree $d-1$ and $\nabla^2 p$ is a matrix of homogeneous polynomials of degree $d-2$ (if $d \geqslant 1$ and $d \geqslant 2$, respectively). We thus have, for $x \neq 0$, that

$$\nabla p(x) = |x|^{d-1} \nabla p(n(x)) \tag{17}$$

and

$$\nabla^2 p(x) = |x|^{d-2} \nabla^2 p(n(x)). \tag{18}$$

From this we see that $\nabla^2 p(x)$ is positive definite for all $x \in S^{k-1}$ if and only if $\nabla^2 p(x)$ is positive definite for all $x \in \mathbb{R}^k$. It also follows that if $\lambda_{\min}(x)$ is the least eigenvalue of $\nabla^2 p(x)$, then

$$\gamma = \min_{|x|=1} \lambda_{\min}(x) > 0$$

if and only if $\nabla^2 p(x)$ is positive definite for all x.

Example 4.8. A subclass of potentials contained in $\mathcal{P}_d$ is the polynomials of degree d. If $H(x)$ is a polynomial of degree d we see that $r(x)$ is a polynomial of degree $d-1$, and we have that

$$r(x) = \sum_{j=0}^{d-1} c_j q_j(x),$$

where q_j is a homogeneous polynomial of degree j. It follows that

$$\begin{aligned} \|\nabla^2 r(x)\| &\leqslant \sum_{j=2}^{d-1} |c_j| \|\nabla^2 q_j(x)\| \\ &\leqslant \sum_{j=2}^{d-1} |x|^{j-2} |c_j| \max_{|z|=1} \|\nabla^2 q_j(z)\|, \end{aligned}$$

from which we clearly have that

$$\|\nabla^2 r(x)\| = o(|x|^{d-2}).$$

Example 4.9. Another subclass of $\mathcal{P}_d$ is the potentials of the form $H(x) = p(x) + \log(q(x))$, where q is a strictly positive homogeneous polynomial and p a homogeneous polynomial of degree d. Then $r(x) = \log q(x)$, and we obtain

$$\nabla r(x) = \frac{1}{q(x)} \nabla q(x)$$

and

$$\nabla^2 r(x) = \frac{1}{q(x)} \nabla^2 q(x) - \frac{1}{q(x)^2} \nabla q(x) \nabla q(x)^{\mathrm{T}}.$$

Since q was a homogenous polynomial, then in fact we have that

$$\|\nabla^2 r(x)\| \to 0$$

for $|x| \to \infty$, which can easily be seen from (17) and (18), and therefore we also have that $\|\nabla^2 r(x)\| = o(|x|^{d-2})$. Notice, however, that a similar statement is *not* in general true if q is a non-homogeneous polynomial.

Theorem 4.10. *If $H \in \mathcal{P}_{\mathrm{d}}^{+}$ then Q_1 is V_2-geometrically ergodic.*

Proof. First notice that since $\nabla^2 p(x)$ is positive definite, it is invertible. Furthermore, $H(x) = p(x) + r(x)$ and, for $z, y \in S^{k-1}$,

$$\begin{aligned} \langle \nabla^2 H(x) z, y \rangle &= \langle \nabla^2 p(x) z, y \rangle + \langle \nabla^2 r(x) z, y \rangle \\ &\geqslant \lambda_{\min, \nabla^2 p}(n(x)) |x|^{d-2} - \|\nabla^2 r(x)\| \\ &= \left(\lambda_{\min, \nabla^2 p}(n(x)) - \frac{\|\nabla^2 r(x)\|}{|x|^{d-2}} \right) |x|^{d-2}, \end{aligned}$$

with $\lambda_{\min, \nabla^2 p}(x)$ denoting the minimal eigenvalue of $\nabla^2 p(x)$. Since $\nabla^2 p(x)$ was positive definite for all x, if $d > 2$ we have

$$\lambda_{\min, \nabla^2 p}(n(x)) \geqslant \gamma > 0,$$

and therefore

$$\lambda_{\min}(x) = \min_{|z|=|y|=1} \langle \nabla^2 H(x) z, y \rangle \to \infty$$

for $|x| \to \infty$, and if $d = 2$ we have

$$\limsup_{|x| \to \infty} \lambda_{\min}(x) \geqslant \gamma > 0.$$

We need to show that $\nabla^2 H(x)^{-1} \nabla H(x)$ is asymptotically homogeneous to x for $|x| \to \infty$, and we have that

$$\begin{aligned} \nabla^2 H(x)^{-1} \nabla H(x) &= (\nabla^2 p(x) + \nabla^2 r(x))^{-1} \nabla p(x) + (\nabla^2 p(x) + \nabla^2 r(x))^{-1} \nabla r(x) \\ &= \frac{1}{d-1} (\nabla^2 p(x) + \nabla^2 r(x))^{-1} \nabla^2 p(x) x \\ &\quad + (\nabla^2 p(x) + \nabla^2 r(x))^{-1} \nabla^2 p(x) \nabla^2 p(x)^{-1} \nabla r(x), \end{aligned}$$

where the last equality follows from (15). First, we see that

$$\nabla^2 p(x)^{-1}\nabla^2 r(x) = \nabla^2 p(n(x))^{-1}\frac{\nabla^2 r(x)}{|x|^{d-2}} \to 0$$

for $|x| \to \infty$, and since inversion is continuous on the invertible matrices,

$$\begin{aligned}(\nabla^2 p(x) + \nabla^2 r(x))^{-1}\nabla^2 p(x) &= \left(\nabla^2 p(x)^{-1}(\nabla^2 p(x) + \nabla^2 r(x))\right)^{-1} \\ &= \left(I + \nabla^2 p(x)^{-1}\nabla^2 r(x)\right)^{-1} \to I\end{aligned}$$

for $|x| \to \infty$. Furthermore,

$$\nabla^2 p(x)^{-1}\frac{\nabla r(x)}{|x|} = \nabla^2 p(n(x))^{-1}\frac{\nabla r(x)}{|x|^{d-1}} \to 0$$

for $|x| \to \infty$, since $|\nabla r(x)| = o(|x|^{d-1})$ from Remark 4.7. All in all, we obtain that $\nabla^2 H(x)^{-1}\nabla H(x)$ is asymptotically homogeneous to x for $|x| \to \infty$ with the asymptotic function being constantly $1/(d-1)$, and since $1/(d-1) \in]0, 1[$ for $d \geqslant 2$, Theorem 4.3 shows that Q_1 is V_2-geometrically ergodic. □

Combining the fact that

$$\lambda_{\min}(x) = \min_{|z|=|y|=1}\langle\nabla^2 H(x)z,\ y\rangle$$

and (15) shows that

$$\begin{aligned}\liminf_{|x|\to\infty}\frac{\langle\nabla H(x),\ x\rangle}{|x|^2} &= \liminf_{|x|\to\infty}\langle\nabla^2 H(x)n(x),\ n(x)\rangle \\ &\geqslant \liminf_{|x|\to\infty}\lambda_{\min}(x).\end{aligned}$$

This implies that if $H \in \mathcal{P}_{\mathrm{d}}^+$ then (12) is satisfied and the Langevin diffusion is geometrically ergodic. In this case we have shown that the local linearization is geometrically ergodic, too. On the other hand, using (18), it is easily shown that

$$\inf_{|x|=1} p(x) > 0 \tag{19}$$

is sufficient for (12) to be satisfied and hence for the Langevin diffusion to be geometrically ergodic. Thus it is not necessary that H is in $\mathcal{P}_d^+$ for the Langevin diffusion itself to be geometrically ergodic.

Example 4.11. We will illustrate by example what can happen if $\nabla^2 p(x)$ is not positive definite. Consider the homogeneous polynomial on $\mathbb{R}^2$ given by

$$p(x_1,\ x_2) = x_1^4 + x_2^4 - x_1^2 x_2^2.$$

For $(x_1, x_2) \neq 0$ we have $p(x_1, x_2) > 0$ and (19) gives that the Langevin diffusion with potential $H = p$ is geometrically ergodic. The second-order derivative is

$$\nabla^2 p(x_1, x_2) = \begin{pmatrix} 12x_1^2 - 2x_2^2 & -4x_1x_2 \\ -4x_1x_2 & 12x_2^2 - 2x_1^2 \end{pmatrix},$$

and if $x_2 = 0$ and $x_1 \neq 0$, the eigenvalues at $(x_1, 0)$ are seen to be $12x_1^2$ and $-2x_1^2$. In particular, $\nabla^2 p(x_1, 0)$ is not positive definite. In the direction given by (1, 0) the matrix $\nabla^2 H$ tends to $-\infty$ in the lower left corner and for the variance matrix we find (using $\sigma = 1$) that

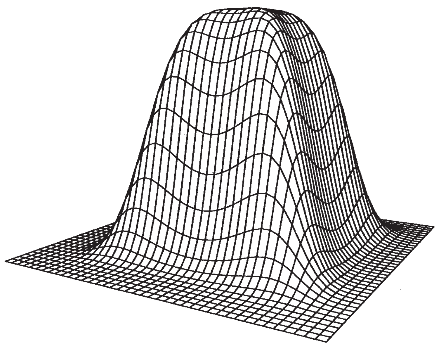
$$\Sigma_1(x_1, 0) = \begin{pmatrix} \dfrac{1 - \exp(-24hx_1^2)}{24x_1^2} & 0 \\ 0 & \dfrac{\exp(4hx_1^2) - 1}{4x_1^2} \end{pmatrix},$$

for which the trace tends to infinity very rapidly (exponentially). This prevents us from using the techniques presented in this paper to check whether some kind of drift equation is satisfied. Moreover, if we look at the behaviour of the Markov chain close to (but not on) the axis given by (1, 0) we find that the mean value will grow rapidly in the second coordinate, too.

In Figure 1 we consider the point $(x_1, x_2) = (3.3, 0.2)$, in which case $\xi_1(3.3, 0.2) = (2.306, 6.261)$ and

$$\Sigma_1(3.3, 0.2) = \begin{pmatrix} 0.0399 & 2.077 \\ 2.077 & 119.56 \end{pmatrix}.$$

This covariance matrix results in a move to a point which almost certainly has a numerically large second coordinate. The same problem is present in the direction along the other axis,



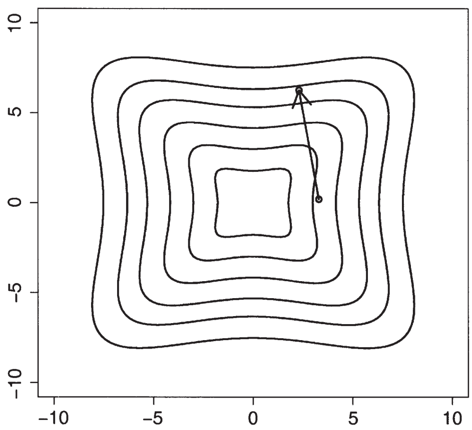


Figure 1. The density $\pi(x) \propto \exp(-2H(x))$ from Example 4.11 (left), and the vector $K_1(3.3, 0.2)$ together with the contour curves for the potential H (right). Notice that K_1 is pointing in a completely wrong direction, and note also the non-convexity of the contour curves.

and the existence of these large jumps away from the centre leads us to question geometric ergodicity of the local linearization Q_1 – it even makes us question ergodicity.

Although we have not been able to prove or disprove ergodicity, we have certainly been able to show some unpleasant features of the local linearization for potentials with (asymptotically) non-positive definite second derivative.

5. Implicit and explicit schemes

As described in Mattingly *et al*. 2002, implicit schemes are also a possible way around numerical instability of the Euler scheme. They derive, for a number of SDEs, geometric ergodicity of the solution to the SDE, which they extend to different approximation schemes. Considering some (in fact quite simple) implicit schemes instead of explicit schemes, they are able to treat more general SDEs

$$\mathrm{d}X(t) = f(X(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{20}$$

where f is not necessarily a gradient and where the noise $W(t)$ might be degenerate. It follows from Mattingly *et al*. (2002, Section 8.2) that the split-step backward Euler as well as the backward Euler scheme for the solution of (20) possess geometric drift with a standard quadratic drift function whenever the vector field f is sufficiently smooth, satisfies the dissipativity condition

$$\langle f(x), x\rangle \leqslant -\beta|x|^2 + \alpha \tag{21}$$

for some $\alpha, \beta > 0$, and the time-steps are chosen sufficiently small. Condition (21) is equivalent to condition (12), which is essentially what we need for the solution of the SDE itself to be geometrically ergodic. So unpleasant features like those of Example 4.11 do not occur when using these implicit schemes.

The geometric ergodicity is, however, only proved for sufficiently small time-steps, whereas the local linearization is geometrically ergodic for all time-steps – whenever it is geometrically ergodic. Also the probabilistic structure of the local linearization is explicitly given as Gaussian transition probabilities, whereas for the implicit schemes we do not know the transition probabilities. So since we also need to verify ϕ-irreducibility, or in the framework of Mattingly *et al*. (2002) some minorization condition, to completely prove geometric ergodicity for the schemes, we might expect more trouble doing so for the implicit schemes than we had for the explicit ones.

6. Langevin diffusions and the Metropolis–Hastings algorithm

Discretizations of the Langevin diffusion with invariant probability measure π have been considered a useful approach to the construction of proposals for the Metropolis–Hastings algorithm (Roberts and Tweedie 1996a; Stramer and Tweedie 1999a; 1999b), although some problems have been encountered. For instance, Roberts and Tweedie (1996a) found that the use of the Euler scheme does not lead to any fruitful results if π has light tails. The local linearization considered in this paper was initially studied with the purpose of avoiding some of these difficulties. The general hope was that if the discretization to the diffusion is good enough, geometric ergodicity would be inherited from the diffusion to the approximation and then to the Metropolis–Hastings algorithm – for instance, in the way described by Stramer and Tweedie (1999b). Counter-examples can, however, be given, for instance using the density $\pi(x) \propto \exp(-x^6)$ on $\mathbb{R}$. Then it follows by Theorem 4.10 that the local linearization is geometrically ergodic, but (tedious) calculations show that the Metropolis–Hastings algorithm is not geometrically ergodic. The details are omitted, but the essence is to show that rejection probabilities tend to 1 for $|x| \to \infty$ and then use Roberts and Tweedie (1996b, Proposition 5). *Ad hoc* corrections to the discretization can be made so that geometric ergodicity is possible for the Metropolis–Hastings algorithm, but then the relation to the original diffusion becomes rather obscure.

Another objection to the construction of proposals using discretization schemes concerns the length h of the time-steps. If one is supposed to obtain a reasonable good approximation of the diffusion, the length h should be rather small, but this will not lead to a good proposal, since the suggested jumps will then be small too. To suggest jumps of a useful size the choice of h should be much larger than seems to be reasonable, if we want the discretization to approximate the diffusion.

Anyway, the use of a proposal like $N(x + (h/2)\nabla \log \pi(x), hI)$ seems perfectly natural from a geometric point of view, since it suggests moves in the direction where the log-density increases the most. As with optimization algorithms, $\nabla \log \pi(x)$ might not have the correct length and some kind of length modification should be used. This has been considered in Christensen *et al*. (2001).

7. Concluding remarks

We have shown that the local linearization considered in this paper behaves better than the simpler Euler discretization for the multidimensional Langevin diffusion, in the sense that geometric ergodicity of the diffusion to a larger extent is preserved by the local linearization. We have also shown for a large and rather concrete class of potentials that the local linearization is geometrically ergodic. It seems, however, that even the local linearization does not inherit geometric ergodicity in complete generality. One should possibly turn to actually quite simple implicit schemes instead to obtain an even better preservation of geometric ergodicity, though the implicit nature complicates some of the analysis since the transition probabilities for the discrete-time approximation are not explicitly known.

Furthermore, geometric ergodicity of the discretization of the Langevin diffusion does not in general lead to geometric ergodicity of the Metropolis–Hastings algorithm, and the diffusion-discretization approach to the construction of proposals is not considered particularly fruitful.

Appendix: Discrete-time Markov chain theory

We briefly review some results from the discrete-time Markov chain theory. We refer to Meyn and Tweedie (1996) for further results and references on the theory of Markov chains on a general state space. The notation follows Meyn and Tweedie (1996) closely.

We will consider ergodic Markov chains on a state space $(E, \mathbb{E})$ with transition probabilities given by a Markov kernel P. The convergence of P^n towards the invariant distribution π is the main interest – especially geometric convergence of P^n in the V-norm.

Definition A.1. *Let $V : E \to [1, \infty[$ be a measurable function. If ν is a signed measure on $(E, \mathbb{E})$ such that V is ν-integrable we define the V-norm of ν as*

$$\|\nu\|_V = \sup_{|f| \leqslant V} \left| \int f \, \mathrm{d}\nu \right| = \sup_{|f| \leqslant V} |\nu(f)|.$$

The set of signed measures making V integrable forms a subspace of the vector space of signed measures on which $\| \cdot \|_V$ is indeed a norm.

Note that since $V \geqslant 1$, the V-norm is stronger than the total variation norm.

Definition A.2. *Let $V : E \to [1, \infty[$ be a measurable function. A Markov kernel P with invariant probability measure π is V-geometrically ergodic if there exist constants R and $\rho < 1$ such that*

$$\|P^n(x, \cdot) - \pi\|_V \leqslant RV(x)\rho^n$$

for all $x \in E$.

Since $\rho < 1$ we obtain that $\|P^n(x, \cdot) - \pi\|_V \to 0$ with a geometric rate.

In concrete cases it is often possible to verify geometric ergodicity by using the Foster–Lyapunov drift conditions. A Markov kernel P is said to have V-*geometric drift* towards a set C if $V : E \to [1, \infty[$ is a measurable function satisfying

$$PV(x) = \int V(y)P(x, \mathrm{d}y) \leqslant \beta V(x) + b1_C(x)$$

for all $x \in E$, some $\beta < 1$ and some $b \in [0, \infty[$. The function V is often called the drift function. Meyn and Tweedie (1996, Theorem 15.0.1) implies:

Theorem A.3. *If P is a ϕ-irreducible and aperiodic Markov kernel with V-geometric drift towards a small set, then P has an invariant probability measure π, $\pi(V) < \infty$ and P is V-geometrically ergodic.*

Acknowledgements

I would like to thank Ernst Hansen for carefully reading and commenting on earlier versions of the manuscript. Thanks are also due to the anonymous referees for providing useful references, in particular to the highly relevant paper by Mattingly *et al.* (2002).

References

Biskey, R., Jimenez, J.C., Riera, J.J. and Valdes, P.A. (1996) Local linearization method for the numerical solution of stochastic differential equations. *Ann. Inst. Statist. Math.*, **48**, 631–644.

Christensen, O.F., Møller, J. and Waagepetersen, R.P. (2001) Geometric ergodicity of Metropolis–Hastings algorithms for conditional simulation in generalised linear mixed models. *Methodol. Comput. Appl. Probab.*, **3**, 309–327.

Down, D., Meyn, S.P. and Tweedie, R.L. (1995) Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, **23**, 1671–1691.

Gardiner, C.W. (1983) *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Berlin: Springer-Verlag.

Ikeda, N. and Watanabe, S. (1989) *Stochastic Differential Equations and Diffusion Processes*. Amsterdam: North-Holland.

Jacobsen, M. (1991) Homogeneous Gaussian diffusions in finite dimensions. Preprint 3, Department of Statistics and Operations Research, University of Copenhagen.

Jacobsen, M. (1994) A brief account of the theory of homogeneous Gaussian diffusions in finite dimension. In H. Niemi, G. Högnas, A.N. Shiryaev and A.V. Melnikov (eds), *Proceedings of the Third Finnish–Soviet Conference on Probability Theory and Mathematical Statistics*, Frontiers Pure Appl. Probab. 1, pp. 86–94. Utrecht: VSP.

Kloeden, P.E. and Platen, E. (1992) *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer-Verlag.

Mattingly, J.C., Stuart, A.M. and Higham, D.J. (2002) Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, **101**, 185–232.

Meyn, S.P. and Tweedie, R.L. (1996) *Markov Chains and Stochastic Stability.* Berlin: Springer-Verlag.

Ozaki, T. (1992) A bridge between non-linear time series models and non-linear stochastic dynamical systems. A local linearization approach. *Statist. Sinica*, **2**, 113–135.

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

Roberts, G.O. and Tweedie, R.L. (1996b) Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, **83**, 95–110.

Stramer, O. and Tweedie, R.L. (1999a) Langevin-type models I: Diffusions with given invariant distributions and their discretizations. *Methodol. Comput. Appl. Probab.* **1**, 283–306.

Stramer, O. and Tweedie, R.L. (1999b) Langevin-type models II: Self-targeting candidates for MCMC algorithms. *Methodol. Comput. Appl. Probab.*, **1**, 307–328.

Received February 2002 and revised July 2002