Necessary conditions for geometric and polynomial ergodicity of random walk-type Markov chains*

S. F. Jarner†
Lancaster University

R. L. Tweedie‡
University of Minnesota

June 22, 2001

Abstract

We give necessary conditions for geometric and polynomial convergence rates of random walk-type Markov chains to stationarity in terms of existence of exponential and polynomial moments of the invariant distribution. These results complement the use of Foster-Lyapunov drift conditions for establishing geometric and polynomial ergodicity. The results allow us to derive exact rates of convergence in situations where we are given the invariant distribution and are thus particularly useful for MCMC applications. We show that the derived bounds are tight for symmetric random walk Metropolis algorithms and Langevin algorithms with polynomial target densities.

*Work supported in part by NSF Grant DMS 9803682 and the EU TMR network ERB-FMRX-CT96-0095 on "Computational and statistical methods for the analysis of spatial data"

†Postal Address: Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, England; email: s.jarner@lancaster.ac.uk

‡Postal Address: Division of Biostatistics, School of Public Health, A460 Mayo Building, Box 303 420 Delaware Street SE Minneapolis, MN 55455-0378, USA; email: tweedie@biostat.umn.edu

[0]Keywords: Markov chains; Markov chain Monte Carlo; Metropolis algorithms; geometric and polynomial ergodicity; geometric and polynomial moments
AMS 2000 Subject Classification: Primary 60J05; 60J10
SHORT TITLE: Geometric and polynomial ergodicity of Markov chains

1 Introduction

Let $\mathbf{X} = (X_0, X_1, \ldots)$ be a discrete-time Markov chain on the d-dimensional Euclidean space $E = \mathbf{R}^d$ equipped with its Borel σ-field $\mathcal{B}$. We assume throughout that the chain is ψ-irreducible, aperiodic and positive recurrent, see [13]. Let π denote the (necessarily unique) invariant distribution. Further, let $P(x, \cdot)$ denote the Markov transition kernel and let $P^n(x, \cdot)$, $n \in N_0$, denote the n-step kernel,

$$P^n(x, A) = \mathsf{P}_x(X_n \in A) \quad (x \in E, A \in \mathcal{B}),$$

where P_x is the conditional distribution of the chain given $X_0 \equiv x$. The corresponding expectation operator will be denoted E_x. For any function V we write $PV(x)$ for the function $\int V(y)P(x, dy)$ and for any signed measure μ we write $\mu(V)$ for $\int V(y)\mu(dy)$.

Following the terminology of [13] a set $C \in \mathcal{B}$ is called small if there exist $n > 0$, $\delta > 0$ and a probability measure ν such that $P^n(x, \cdot) \geq \delta\nu(\cdot)$ for all x in C. Under our assumptions of ψ-irreducibility and aperiodicity this is the same as C being petite, see [13].

In this paper we consider geometrically and polynomially ergodic Markov chains, i.e. Markov chains for which there exists a small set C, a function $f : E \to [1, \infty)$ and a geometric or polynomial rate function $r(n)$ such that

$$\sup_{x \in C} \mathsf{E}_x[\sum_{k=0}^{\tau_C - 1} r(k) f(X_k)] < \infty, \tag{1}$$

where $\tau_C = \inf\{n \geq 1 \ : X_n \in C\}$ is the first return time of the chain to C. In the polynomial case (1) implies that for π-almost all x

$$r(n)\|P^n(x, \cdot) - \pi\|_f \to 0, \quad n \to 0, \tag{2}$$

where the f-norm is defined for a signed measure μ as $\|\mu\|_f = \sup_{|g| \leq f} |\mu(g)|$. In the geometric case the rate of convergence in (2) is exponential but generally of a lower order than $r(n)$.

The most common way of establishing geometric and subgeometric ergodicity of Markov chains on general state spaces is by verifying an associated Foster-Lyapunov type drift condition, see [15, 14, 19, 13, 18, 8]. In the Markov chain Monte Carlo (MCMC) context this approach has been successfully applied a number of times to derive sufficient conditions on π for geometric and subgeometric ergodicity of the Gibbs sampler and other Metropolis-Hastings algorithms, see e.g. [2, 11, 16, 17, 7, 5, 9]. However, there has been only very few results giving necessary conditions. In this paper we show that for random walk-type Markov chains, such as random walk Metropolis algorithms, geometric and polynomial ergodicity implies that π has certain exponential and polynomial moments. Combined with the results of [9] these results allow us to derive exact rates of convergence for certain MCMC algorithms which is something quite rare.

We say that $\mathbf{X}$ if of random walk-type if for every $\epsilon > 0$ there exists $K > 0$ such that $P(x, B(x, K)) > 1 - \epsilon$ for all x, where $B(x, K) = \{y : |y - x| < K\}$ denotes the open ball with center x and radius K w.r.t. Euclidean distance $|\cdot|$. This is equivalent to the family of increment distributions $\{P(x, \cdot) - x : x \in E\}$ being uniformly tight. Random walk-type Markov chains occur frequently in e.g. MCMC methods and queueing theory.

In Section 2 we show that a random walk-type Markov chain can only be geometrically ergodic if π has exponential moments. In the special case of a symmetric random walk Metropolis algorithm with increment proposal distribution with finite first absolute moment this has previously been proved in [11, 7] by using Wald's equation to bound the mean return time to the center of the space. The approach taken here however is more general and involves controlling only a fraction of the probability mass for which more detailed behaviour of the sample paths can be obtained in order to provide sample path bounds on the return times.

In Section 3 this idea is taken further to show the existence of polynomial moments of π when $\mathbf{X}$ is polynomially ergodic. We say that $\mathbf{X}$ is *polynomially ergodic of order* (α, β), where $\alpha, \beta \geq 0$, if (1) holds with $r(k) = (k+1)^\beta$ and $f(x) = (|x| + 1)^\alpha$, and we show that this implies

$$\int_{\mathbf{R}^d} |x|^{\alpha+\eta\beta} \pi(dx) < \infty, \tag{3}$$

where $0 < \eta \leq 2$ depends on the tail behaviour and drift of the family of increment distributions $\{P(x, \cdot) - x\}$. The case $\eta = 1$ corresponds to uniform integrability of the increment distributions, and the case $\eta = 2$ corresponds to the family of increment distributions having uniformly bounded variance and drift towards the center of the space of order at most $|x|^{-1}$.

Section 4 uses these results to show that the polynomial rates of convergence of symmetric random walk Metropolis algorithms and Langevin algorithms with polynomial target densities found in [9] are the exact rates. Knowing the exact rates instead of only lower bounds has obvious advantages when comparing different algorithms.

The technique used in this paper of controlling only part of the probability mass in order to derive lower bounds on expectations of functionals of the Markov chain seems quite general and suitable for providing necessary conditions also for more general (f, r)-ergodic chains than those considered here. Indeed, similar ideas are used in [10] to get necessary and sufficient conditions for specific convergence rates of a Markov chain associated with the mean of a Dirichlet process.

2 Geometric ergodicity

The Markov chain $\mathbf{X}$ is called geometrically ergodic if (1) holds with $r(n) = \rho^n$ for some $\rho > 1$. By Theorem 15.0.1 of [13] an equivalent condition is that there exists a small set C, constants $\lambda < 1$ and $b < \infty$ and a function $V \geq 1$ finite for at least one $x_0 \in E$ satisfying

$$PV \leq \lambda V + b 1_C. \tag{4}$$

Showing the Foster-Lyapunov drift condition (4) is often the easiest way to prove geometric ergodicity. Note, that by Theorem 14.3.7 of [13] any function V satisfying (4) has finite expectation w.r.t. π. In particular, V is finite π-a.e. We will use these properties below.

The next theorem shows that for random walk-type Markov chains as defined in the introduction geometric ergodicity implies the existence of exponential moments of π. For symmetric random walk Metropolis algorithms this has previously been proved in [11, 7] under the additional assumption that the family of increment distributions has finite first absolute moment but this assumption is not needed for our approach.

We need the following simple lemma from [7].

Lemma 2.1 *If $\mathbf{X}$ is a random walk-type Markov chain then every small set is bounded.*

Theorem 2.2 *Let $\mathbf{X}$ be a random walk-type Markov chain. If $\mathbf{X}$ is geometrically ergodic then there exists $s > 0$ such that*

$$\int_{\mathbf{R}^d} e^{s|x|} \pi(dx) < \infty. \tag{5}$$

Proof Since $\mathbf{X}$ is geometrically ergodic there exist small set C, constants $\lambda < 1$ and $b < \infty$ and function $V \geq 1$ finite π-a.e. such that (4) holds. Choose M so large that $A = \{V \leq M\}$ has positive π-measure. By Theorems 15.2.6 and 15.2.1 of [13] A is then a small V-Kendall set, and by Theorem 15.2.4 of [13] there then exist $\kappa > 1$, $\tilde{\lambda} < 1$ and $\tilde{b} < \infty$ such that

$$P\tilde{V} \leq \tilde{\lambda}\tilde{V} + \tilde{b} 1_A, \tag{6}$$

where

$$\tilde{V}(x) = \begin{cases} V(x) & \text{for } x \in A, \\ \mathsf{E}_x[\sum_{k=0}^{\tau_A} \kappa^k V(X_k)] & \text{for } x \in A^c. \end{cases}$$

Since $\tilde{V}$ satisfies the drift condition (6) we have $\pi(\tilde{V}) < \infty$, and (5) thus follows if we can find an exponential lower bound on $\tilde{V}(x)$ for $|x|$ sufficiently large.

Choose R so large that $A \subset B(0, R)$; this can be done because A is a small set and hence bounded by Lemma 2.1. Since $V \geq 1$ we have the lower bound $\tilde{V}(x) \geq \mathsf{E}_x[\kappa^{\tau_A}]$ for $x \in A^c$ and thus in particular for $|x| \geq R$.

Choose $\epsilon > 0$ so small that $\kappa(1-\epsilon) > 1$ and then, using the random walk-structure, choose K such that $P(x, B(x,K)) > 1-\epsilon$ for all x. For any real number z let $\lceil z \rceil$ denote the smallest integer equal to or larger than z. For $|x| \geq R$ we then have

$$\mathsf{E}_x[\kappa^{\tau_A}] \geq (\kappa(1-\epsilon))^w, \tag{7}$$

where $w = \lceil (|x| - R)/K \rceil$, because $(1-\epsilon)^w$ is a lower bound for the probability that the next w jumps are at most of length K and on this event $\tau_A \geq w$. It follows that we can find $s > 0$ and $c > 0$ such that $\tilde{V}(x) \geq \mathsf{E}_x[\kappa^{\tau_A}] \geq ce^{s|x|}$ for $|x| \geq R$ and we are finished. ■

3 Polynomial ergodicity

Recall that the Markov chain $\mathbf{X}$ is polynomially ergodic of order (α, β), where $\alpha, \beta \geq 0$, if there exists a small set C such that

$$\sup_{x \in C} \mathsf{E}_x[\sum_{k=0}^{\tau_C - 1} (k+1)^\beta (|X_k| + 1)^\alpha] < \infty. \tag{8}$$

By Theorem 14.0.1 of [13] this implies in particular that $\mathbf{X}$ is positive recurrent with invariant distribution π and $\pi(|x|^\alpha) < \infty$.

For random walk-type Markov chains we show below under varying additional conditions that polynomial ergodicity implies polynomial moments of π. The results take the general form that polynomial ergodicity of order (α, β) implies that

$$\int_{\mathbf{R}^d} |x|^{\alpha + \eta\beta} \pi(dx) < \infty, \tag{9}$$

where $0 < \eta \leq 2$ depends on the heaviness of the tails and the drift of the family of increment distributions $\{P(x, \cdot) - x\}$.

Let h be a non-increasing function $h : [0, \infty) \to [0, 1]$ such that for all $x \in \mathbf{R}^d$ and all $y \geq 0$

$$P(x, B(x,y)^c) \leq h(y). \tag{10}$$

Since we can always use $h \equiv 1$ such a function exists for any Markov chain. The Markov chain is of random walk-type if and only if there exists h with $h(y) \to 0$ as $y \to \infty$ such that (10) holds.

We first consider conditions in terms of how quickly h tends to zero. Theorems 3.2 and 3.3 show that if h is integrable then (9) holds with $\eta = 1$, while if h tends to zero at a non-integrable polynomial rate we get (9) with $0 < \eta < 1$. In Section 3.1 we assume more structure and show in Theorems 3.4 and 3.5 that (9) holds with $1 < \eta \leq 2$ for random walk-type Markov chains where the family of increment distributions has uniformly bounded moments of order η and drift to the center of the space of order at most $|x|^{1-\eta}$.

For any sequence r we define the sequence Δr by

$$\Delta r(0) = r(0), \quad \Delta r(k) = r(k) - r(k-1) \quad \text{for } k = 1, 2, \ldots$$

From Theorems 2.1 and 2.3 of [18] and the trivial bound $|x|^\alpha \leq (|x|+1)^\alpha$ we get the following lemma on which all the following results rely.

Lemma 3.1 *If* $\mathbf{X}$ *is polynomially ergodic of order* (α, β) *then there exists a small set* C *such that*

$$\int_{\mathbf{R}^d} \mathsf{E}_x[\sum_{k=0}^{\tau_C - 1} \Delta r(k)\ |X_k|^\alpha] \pi(dx) < \infty, \tag{11}$$

where $r(k) = (k+1)^\beta$.

Note that for $\beta = 0$ (11) reduces to $\pi(|x|^\alpha) < \infty$ since in this case $\Delta r(k) = 0$ for $k > 0$.

Theorem 3.2 *Assume* $\mathbf{X}$ *is of random walk-type and that (10) holds with* $\int_0^\infty h(y)dy < \infty$. *If* $\mathbf{X}$ *is polynomially ergodic of order* (α, β) *then*

$$\int_{\mathbf{R}^d} |x|^{\alpha+\beta} \pi(dx) < \infty. \tag{12}$$

Proof Assumption (10) with $\int_0^\infty h(y)dy < \infty$ implies that there exists a sequence of i.i.d. random variables $Y_n > 0$ with finite mean $\mu = \mathsf{E}(Y_n)$ such that for all $x \in \mathbf{R}^d$ and all $y \geq 0$

$$P(x, B(x,y)^c) \leq \mathsf{P}(Y_n \geq y). \tag{13}$$

By the weak law of large numbers we have for any $\epsilon > 0$

$$\mathsf{P}(S_n < (\mu + \epsilon)n) \to 1, \quad n \to \infty,$$

where $S_n = Y_1 + \ldots + Y_n$. Hence we can choose N so large that for $n \geq N$

$$\mathsf{P}(S_n < 2\mu n) \geq 1/2.$$

Using (13) this shows by a stochastic comparison argument that for all $x \in \mathbf{R}^d$ and all $n \geq N$

$$\mathsf{P}_x(X_i \in B(x, 2\mu n) \text{ for } i = 0, \ldots, n) \geq 1/2. \tag{14}$$

For $|x|$ so large that $|x|/4\mu \geq N$ it follows from (14) with $n = \lfloor |x|/4\mu \rfloor \geq N$ that

$$\mathsf{P}_x(X_i \in B(x, |x|/2) \text{ for } i = 0, \dots, \lfloor |x|/4\mu \rfloor) \geq 1/2. \tag{15}$$

By Lemmas 3.1 and 2.1 (11) holds for a small and hence bounded set C. For $|x|$ so large that $B(x, |x|/2) \subset C^c$ we have on the above event

$$\tau_C - 1 \geq \lfloor |x|/4\mu \rfloor, \quad |X_k|^\alpha \geq (|x|/2)^\alpha \quad \text{for } k = 0, \dots, \lfloor |x|/4\mu \rfloor,$$

and thereby also

$$\sum_{k=0}^{\tau_C - 1} \Delta r(k) \; |X_k|^\alpha \geq \sum_{k=0}^{\lfloor |x|/4\mu \rfloor} \Delta r(k) \; |X_k|^\alpha \geq \frac{|x|^\alpha}{2^\alpha} r(\lfloor |x|/4\mu \rfloor) \geq \frac{|x|^{\alpha+\beta}}{2^\alpha (4\mu)^\beta},$$

where $r(k) = (k+1)^\beta$. For $|x|$ sufficiently large this event has probability at least 1/2 by (15) and hence for $|x|$ sufficiently large

$$\mathsf{E}_x[\sum_{k=0}^{\tau_C - 1} \Delta r(k) \; |X_k|^\alpha] \geq \frac{|x|^{\alpha+\beta}}{2^{\alpha+1} (4\mu)^\beta},$$

and (12) now follows from (11). ∎

When the dominating function h decays to zero at a polynomial rate but is not integrable we can use a stable law limit result instead of the weak law of large numbers to get the following result.

Theorem 3.3 *Assume* $\mathbf{X}$ *is of random walk-type and that there exists* $0 < \eta < 1$ *and constant* $c > 0$ *such that (10) holds with* $h(y) = cy^{-\eta}$ *for* y *sufficiently large. If* $\mathbf{X}$ *is polynomially ergodic of order* (α, β) *then*

$$\int_{\mathbf{R}^d} |x|^{\alpha + \eta\beta} \pi(dx) < \infty. \tag{16}$$

Proof By (10) and the assumption on h there exists a sequence of i.i.d. random variables $Y_n > 0$ with distribution function F satisfying $1 - F(y) = cy^{-\eta}$ for y sufficiently large such that for all $x \in \mathbf{R}^d$ and all $y \geq 0$

$$P(x, B(x, y)^c) \leq \mathsf{P}(Y_n \geq y). \tag{17}$$

From Chapters XVII.5 and XIII.6 of [4] it follows that there exists $\delta > 0$ such that for all $z > 0$

$$\mathsf{P}(S_n \leq z\delta n^{1/\eta}) \to G_\eta(z), \quad n \to \infty,$$

where $S_n = Y_1 + \ldots + Y_n$ and G_η is the cdf of a stable law with index η. In particular, we can find $\nu > 0$ and N such that for $n \geq N$

$$\mathsf{P}(S_n \leq \nu n^{1/\eta}) \geq 1/2.$$

Thus by (17) and a stochastic comparison argument it holds for all $x \in \mathbf{R}^d$ and all $n \geq N$

$$\mathsf{P}_x(X_i \in B(x, \nu n^{1/\eta}) \text{ for } i = 0, \ldots, n) \geq 1/2. \tag{18}$$

For $|x|$ so large that $(|x|/2\nu)^\eta \geq N$ it follows from (18) with $n = \lfloor (|x|/2\nu)^\eta \rfloor \geq N$ that

$$\mathsf{P}_x(X_i \in B(x, |x|/2) \text{ for } i = 0, \ldots, \lfloor (|x|/2\nu)^\eta \rfloor) \geq 1/2, \tag{19}$$

from which (16) follows by the same arguments as in the proof of Theorem 3.2. ∎

3.1 Random walk-type Markov chains with moments and drift

In this section we obtain higher polynomial moments of π for random walk-type Markov chains with increment distributions having uniformly bounded moments and polynomial drift. For the Markov kernel $P(x, \cdot)$ let $D(x, \cdot)$ be the distribution of $|X_1| - |x|$, i.e

$$D(x, (-\infty, z)) = P(x, B(0, |x| + z)) \quad (x \in \mathbf{R}^d,\ z \in \mathbf{R}).$$

For $z \leq -x$ the ball $B(0, |x| + z)$ is the empty set and the probability on the right hand side is zero. We assume that there exists a family of distributions $(H(r))_{r \geq 0}$ on $\mathbf{R}$ such that for all $r \geq 0$ and all $|x| \geq r$

$$H(r) \overset{\text{st}}{\leq} D(x, \cdot), \tag{20}$$

where for two distributions Q_1 and Q_2 on $\mathbf{R}$ we write $Q_1 \overset{\text{st}}{\leq} Q_2$ when the corresponding distribution functions F_1 and F_2 satisfy $F_1(y) \geq F_2(y)$ for all y.

Let $Y(r)$ be a random variable with distribution $H(r)$. The assumption is then that whenever the Markov chain is at distance at least r from the origin the amount by which the distance increases after one iteration is stochastically larger than $Y(r)$. If $\mathbf{X}$ is of random walk-type this assumption is always satisfied and in fact we can choose $H(r) = H$ independent of r such that (20) holds. In general, however, we want to choose $H(r)$ depending on r and "as large as possible" for the given r.

In Theorems 3.4 and 3.5 below we assume that for some $1 < \eta \leq 2$ and r sufficiently large $Y(r)$ has uniformly bounded moments of order η and the drift $\mathsf{E}[Y(r)]$ is bounded from below by $-cr^{1-\eta}$, $c > 0$. We show that under these assumptions polynomial ergodicity of order (α, β) implies (9). These assumptions imply that the drift of the Markov chain towards the center of space gets smaller and smaller such that it looks more and more like an unbiased random walk the further away it gets from the origin. As shown in Section 4 Langevin and symmetric random walk Metropolis algorithms with polynomial target densities behave like this. Since we only make assumptions about the limit behaviour of $Y(r)$ it is enough that (20) holds for r sufficiently large.

Theorem 3.4 *Assume* $\mathbf{X}$ *is of random walk-type and that there exists a constant* $c > 0$ *such that*

$$\mathsf{E}[Y(r)] \geq -\frac{c}{r} \quad \text{for } r \text{ sufficiently large,} \tag{21}$$

and

$$\limsup_{r\to\infty} \operatorname{Var}(Y(r)) < \infty. \tag{22}$$

If $\mathbf{X}$ *is polynomially ergodic of order* (α, β) *then*

$$\int_{\mathbf{R}^d} |x|^{\alpha+2\beta} \pi(dx) < \infty. \tag{23}$$

Proof For any x we have by (20) the stochastic ordering $S(|x|/2) \overset{\text{st}}{\leq} D(y,\cdot)$ for all $|y| \geq |x|/2$ and by a stochastic comparison argument we then also have

$$\hat{\tau}_{(-\infty,|x|/2)} \overset{\text{st}}{\leq} \tau_{B(0,|x|/2)}, \tag{24}$$

where $\tau_{B(0,|x|/2)}$ is the first return time to $B(0,|x|/2)$ of the Markov chain $\mathbf{X}$ started at x and $\hat{\tau}_{(-\infty,|x|/2)}$ is the first return time to the interval $(-\infty,|x|/2)$ of the random walk (W_i) on $\mathbf{R}$ given by

$$\begin{aligned} W_0 &= |x|, \\ W_i &= W_{i-1} + Z_i \quad (i \geq 1), \end{aligned}$$

where (Z_i) is an i.i.d. sequence of random variables with distribution $H(|x|/2)$. To ease the notation we are suppressing the dependency of W_i and Z_i on x.

Let $R > 0$ be so large that (21) holds for $r \geq R$ and so large that $K = \sup_{r\geq R} \operatorname{Var}(Y(r)) < \infty$. For $|x| \geq 2R$ we then have by Kolmogorov's inequality for any $a > 0$ and any $n > 0$

$$\mathsf{P}(\max_{1\leq k\leq n} |S_k - m_k| > a) \leq \frac{\operatorname{Var}(S_n)}{a^2} \leq \frac{nK}{a^2}, \tag{25}$$

where $S_k = Z_1 + \ldots + Z_k$ and $m_k = \mathsf{E}[S_k] = k\mathsf{E}[Y(|x|/2)]$.

Let $\delta = \min\{1/32K, 1/8c\}$. From (25) with $n = \lfloor \delta|x|^2 \rfloor$ and $a = |x|/4$ it follows that

$$\mathsf{P}(\max_{1\leq k\leq \lfloor\delta|x|^2\rfloor} |S_k - m_k| > |x|/4) \leq \frac{\lfloor\delta|x|^2\rfloor 16K}{|x|^2} \leq 1/2. \tag{26}$$

Since $m_k \geq -2ck/|x| \geq -2c\lfloor \delta |x|^2 \rfloor / |x| \geq -|x|/4$ for $k = 1, \ldots, \lfloor \delta |x|^2 \rfloor$ it follows that

$$\mathsf{P}(\hat{\tau}_{(-\infty,|x|/2)} > \lfloor \delta |x|^2 \rfloor) \geq 1/2, \tag{27}$$

and then by (24) also

$$\mathsf{P}_x(\tau_{B(0,|x|/2)} > \lfloor \delta |x|^2 \rfloor) \geq 1/2. \tag{28}$$

By Lemmas 3.1 and 2.1 (11) holds for a small and hence bounded set C. For $|x|$ so large that $C \subset B(0, |x|/2)$ we have on the above event

$$\tau_C - 1 \geq \lfloor \delta |x|^2 \rfloor, \quad |X_k|^\alpha \geq (|x|/2)^\alpha \quad \text{for } k = 0, \ldots, \lfloor \delta |x|^2 \rfloor,$$

and thereby also

$$\sum_{k=0}^{\tau_C - 1} \Delta r(k)\ |X_k|^\alpha \geq \sum_{k=0}^{\lfloor \delta |x|^2 \rfloor} \Delta r(k)\ |X_k|^\alpha \geq \frac{|x|^\alpha}{2^\alpha} r(\lfloor \delta |x|^2 \rfloor) \geq \frac{|x|^{\alpha+2\beta}\delta^\beta}{2^\alpha},$$

where $r(k) = (k+1)^\beta$. For $|x|$ sufficiently large this event has probability at least 1/2 by (28) and hence for $|x|$ sufficiently large

$$\mathsf{E}_x[\sum_{k=0}^{\tau_C - 1} \Delta r(k)\ |X_k|^\alpha] \geq \frac{|x|^{\alpha+2\beta}\delta^\beta}{2^{\alpha+1}},$$

and (23) now follows from (11). ■

It is well known that for a one-dimensional symmetric random walk with finite variance the return time to the center increases like $|x|^2$, see e.g. Chapter III of [3]. The theorem above says that this is still the case if the random walk is biased of order $|x|^{-1}$.

Theorem 3.5 *Assume* $\mathbf{X}$ *is of random walk-type and that there exists* $1 < \eta < 2$ *and constant* $c > 0$ *such that*

$$\mathsf{E}[Y(r)] \geq -\frac{c}{r^{\eta-1}} \quad \textit{for } r \textit{ sufficiently large}, \tag{29}$$

and

$$\limsup_{r \to \infty} \mathsf{E}|Y(r)|^\eta < \infty. \tag{30}$$

If $\mathbf{X}$ *is polynomially ergodic of order* (α, β) *then*

$$\int_{\mathbf{R}^d} |x|^{\alpha+\eta\beta} \pi(dx) < \infty. \tag{31}$$

Proof As in the proof of Theorem 3.4 we seek to bound the return time $\hat{\tau}_{(-\infty,|x|/2)}$ of the random walk

$$W_0 = |x|, \qquad W_i = W_{i-1} + Z_i \quad (i \geq 1),$$

where (Z_i) is an i.i.d. sequence of random variables with distribution $H(|x|/2)$. Let $\tilde{Z}_i = Z_i 1_{(|Z|_i \leq |x|)}$ be the random variable Z_i truncated at $|x|$, and let $\tilde{\tau}_{(-\infty,|x|/2)}$ be the return time to $(-\infty, |x|/2)$ of the random walk $(\tilde{W}_i)$ given by

$$\tilde{W}_0 = |x|, \qquad \tilde{W}_i = \tilde{W}_{i-1} + \tilde{Z}_i \quad (i \geq 1).$$

Let $R > 0$ be so large that (29) holds for $r \geq R$ and so large that $K = \sup_{r \geq R} \mathsf{E}|Y(r)|^\eta < \infty$. For $|x| \geq 2R$ we then have the bounds

$$\mathsf{P}(\tilde{Z}_i \neq Z_i) = \mathsf{P}(|Z_i| > |x|) \leq \frac{\mathsf{E}|Z_i|^\eta}{|x|^\eta} \leq \frac{K}{|x|^\eta},$$

$$|\mathsf{E}[\tilde{Z}_i] - \mathsf{E}[Z_i]| \leq \mathsf{E}|\tilde{Z}_i - Z_i| = \mathsf{E}[|Z_i| 1_{(|Z_i|>|x|)}] \leq \frac{\mathsf{E}|Z_i|^\eta}{|x|^{\eta-1}} \leq \frac{K}{|x|^{\eta-1}},$$

$$\mathsf{E}[\tilde{Z}_i] \geq \mathsf{E}[Z_i] - |\mathsf{E}[\tilde{Z}_i] - \mathsf{E}[Z_i]| \geq -\frac{2^{\eta-1}c + K}{|x|^{\eta-1}} = -\frac{d}{|x|^{\eta-1}},$$

where $d = 2^{\eta-1}c + K$, and

$$\mathrm{Var}(\tilde{Z}_i) \leq \mathsf{E}\tilde{Z}_i^2 = \mathsf{E}[|Z_i|^2 1_{(|Z_i| \leq |x|)}] \leq |x|^{2-\eta} \mathsf{E}|Z_i|^\eta \leq K|x|^{2-\eta}.$$

For $|x| \geq 2R$ we have by Kolmogorov's inequality for any $a > 0$ and any $n > 0$

$$\mathsf{P}(\max_{1 \leq k \leq n} |\tilde{S}_k - \tilde{m}_k| > a) \leq \frac{\mathrm{Var}(\tilde{S}_n)}{a^2} \leq \frac{nK|x|^{2-\eta}}{a^2}, \tag{32}$$

where $\tilde{S}_k = \tilde{Z}_1 + \ldots + \tilde{Z}_k$ and $\tilde{m}_k = \mathsf{E}[\tilde{S}_k] = k\mathsf{E}[\tilde{Z}_1]$.

Let $\delta = \min\{1/32K, 1/4d\}$. From (32) with $n = \lfloor \delta |x|^\eta \rfloor$ and $a = |x|/4$ we get

$$\mathsf{P}(\max_{1 \leq k \leq \lfloor \delta |x|^\eta \rfloor} |\tilde{S}_k - \tilde{m}_k| > |x|/4) \leq \frac{\lfloor \delta |x|^\eta \rfloor |x|^{2-\eta} 16K}{|x|^2} \leq 1/2. \tag{33}$$

Since $\tilde{m}_k \geq -kd/|x|^{\eta-1} \geq -\lfloor \delta |x|^\eta \rfloor d/|x|^{\eta-1} \geq -|x|/4$ for $k = 1, \ldots, \lfloor \delta |x|^\eta \rfloor$ it follows that

$$\mathsf{P}(\tilde{\tau}_{(-\infty,|x|/2)} > \lfloor \delta |x|^\eta \rfloor) \geq 1/2. \tag{34}$$

Further,

$$\mathsf{P}(\tilde{Z}_k = Z_k \text{ for } k = 1, \ldots, \lfloor \delta |x|^\eta \rfloor) = 1 - \mathsf{P}(\cup_{k=1}^{\lfloor \delta |x|^\eta \rfloor} (\tilde{Z}_k \neq Z_k)) \geq 1 - \frac{\lfloor \delta |x|^\eta \rfloor K}{|x|^\eta} \geq 31/32,$$

and then

$$\begin{aligned}\mathsf{P}(\hat\tau_{(-\infty,|x|/2)} > \lfloor \delta|x|^\eta \rfloor) &\geq \mathsf{P}(\tilde\tau_{(-\infty,|x|/2)} > \lfloor \delta|x|^\eta \rfloor \text{ and } \tilde Z_k = Z_k \text{ for } k = 1,\ldots,\lfloor \delta|x|^\eta \rfloor) \\ &\geq 15/32.\end{aligned}$$

As in the proof of Theorem 3.4 we then also have

$$\mathsf{P}_x(\tau_{B(0,|x|/2)} > \lfloor \delta|x|^\eta \rfloor) \geq 15/32,$$

from which conclusion (31) follows as in the proof of Theorem 3.4. ∎

In Theorems 3.4 and 3.5 the drift is assumed to be bounded from below by $-cr^{1-\eta}$. This assumption is only made to match the random fluctuations such that after n iterations the distance travelled due to drift and due to random fluctuations are of the same order. However, the arguments used in the proofs of Theorems 3.4 and 3.5 can also be used if the drift is allowed to be larger but in this case the drift will dominate the random fluctuations and the inferred polynomial moments of π will be smaller.

4 Applications to MCMC

We will give two MCMC applications of the results in the previous sections: one concerning the symmetric random walk Metropolis algorithm and one concerning the Langevin algorithm, both of which are special cases of the Metropolis-Hastings algorithm, [6]. In both cases we will show that the polynomial ergodicity results obtained in [9] are tight.

The Metropolis-Hastings algorithm is an algorithm for constructing a Markov chain with a prescribed invariant distribution π referred to as the target distribution. We assume that the state space is $\mathbf{R}^d$ equipped with its Borel σ-field, and that the target distribution π has density, also denoted by π, w.r.t. Lebesgue measure μ^{Leb}. The algorithm is based on a *candidate transition kernel* $Q(x,\cdot)$ which generates proposed moves for the Markov chain $\mathbf{X}$. We assume that $Q(x,\cdot)$ has density $q(x,y)$ w.r.t. μ^{Leb}. If the current state is x a proposed move to y, generated according to the density $q(x,y)$, is then accepted with probability

(35) $$\alpha(x,y) = \begin{cases} \min\left\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right\}, & \text{when } \pi(x)q(x,y) > 0, \\ 1, & \text{when } \pi(x)q(x,y) = 0. \end{cases}$$

Thus the Markov transition kernel P for the Markov chain $\mathbf{X}$ is given by

(36) $$P(x,dy) = p(x,y)\mu^{\text{Leb}}(dy) + r(x)\delta_x(dy),$$

where $p(x,y) = \alpha(x,y)q(x,y)$ for $x \neq y$ and 0 otherwise, δ_x is the point mass at x and

$$r(x) = \int (1 - \alpha(x,y))q(y-x)\mu^{\text{Leb}}(dy) \tag{37}$$

is the probability of staying at x. The kernel P is reversible w.r.t. π and hence has π as its invariant distribution.

We first consider the special case of this algorithm known as the symmetric random walk Metropolis algorithm, [12], in which q has the form

$$q(x,y) = q(|x-y|), \tag{38}$$

i.e. the proposed increments are generated according to the same symmetric distribution $Q(dx) = q(x)\mu^{\text{Leb}}(dx)$. In this case the acceptance probability simplifies to

$$\alpha(x,y) = \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}. \tag{39}$$

The symmetric random walk Metropolis algorithm is clearly of random walk-type as defined in the introduction.

We assume that π is bounded away from zero and infinity on bounded sets and that q is bounded away from zero in some region around zero, that is there exists $\delta_q > 0$ and $\epsilon_q > 0$ such that $q(x) \geq \epsilon_q$ for $|x| \leq \delta_q$. Under these assumptions P is μ^{Leb}-irreducible and aperiodic by Theorem 2.2 of [16].

By Theorem 2.2 it follows that exponential or lighter tails of π is a necessary condition for geometric ergodicity of the symmetric random walk Metropolis algorithm irrespective of the proposal distribution Q. In one dimension this is essentially also a sufficient condition while in higher dimensions additional assumptions on the contour manifolds are needed, see [11, 16, 7].

If π has polynomial tails the algorithm will be only polynomially ergodic. It is somewhat surprising, however, that the order of polynomial ergodicity depends on the tails of both π and Q. For ease of exposition we consider only a stylised one-dimensional case.

Assume π is a continuous, strictly positive density on the half-line $[0, \infty)$ and that there exists $r > 0$ such that

$$\pi(x) \propto \frac{1}{x^{1+r}} \quad \text{for } x \text{ sufficiently large.} \tag{40}$$

Proposition 4.1 *Assume π is of form (40) and that Q has finite variance. For $\alpha, \beta \geq 0$ the symmetric random walk Metropolis algorithm is polynomially ergodic of order (α, β) if and only if $\alpha + 2\beta < r$.*

Proof The "if" part follows from Proposition 3.1 of [9]. The "only if" part will follow from Theorem 3.4 if we can show that (21) and (22) are satisfied.

By the assumptions on π we have that for x sufficiently large all proposed moves to the left are accepted. Also, by (40) we have that for x sufficiently large the acceptance probability of any positive increment y is an increasing function in x, i.e. $\alpha(x, x+y)$ is increasing in x for any $y \geq 0$. Thus for z sufficiently large we have for all $x \geq z$

$$H(z) \overset{\text{st}}{\leq} P(x, \cdot) - x, \tag{41}$$

where $H(z, dy) = h(z, y)\mu^{\text{Leb}}(dy)$ for $y \neq 0$ with $h(z, y)$ given by

$$h(z, y) = \begin{cases} q(y), & \text{for } y < 0, \\ q(y)\left(\frac{z}{z+y}\right)^{1+r}, & \text{for } y > 0, \end{cases} \tag{42}$$

and $H(z, \{0\}) = 1 - \int h(z, y)dy$. Let $Y(z)$ be a random variable with distribution $H(z)$. Since Q is assumed to have variance (22) is satisfied and it only remains to show (21). Now using that for any $u \geq 0$

$$\left(\frac{1}{1+u}\right)^{1+r} - 1 \geq -(1+r)u,$$

we find

$$\mathsf{E}[Y(z)] = \int_0^\infty q(y)y\left[\left(\frac{z}{z+y}\right)^{1+r} - 1\right]dy \geq -\frac{(1+r)}{z}\int_0^\infty q(y)y^2 dy,$$

and since $\int_0^\infty q(y)y^2dy < \infty$ this shows that (21) holds and we are finished. ∎

For symmetric random walk Metropolis algorithms with proposal distribution without variance we have the following result. Recall that a function l is normalised slowly varying if for all $a > 0$, $x^a l(x)$ is eventually increasing and $x^{-a}l(x)$ is eventually decreasing.

Proposition 4.2 *Assume π is of form (40) and that there exists $0 < \eta \leq 2$ such that for $|x|$ sufficiently large $q(x)$ can be written*

$$q(x) = \frac{l(|x|)}{|x|^{1+\eta}}, \tag{43}$$

where l is a normalised slowly varying function.

The symmetric random walk Metropolis algorithm is polynomially ergodic of order (α, β) for all $\alpha, \beta \geq 0$ with $\alpha + \eta\beta < r$, and not polynomially ergodic of order (α, β) for any $\alpha, \beta \geq 0$ with $\alpha + \eta\beta > r$.

Proof That the symmetric random walk Metropolis algorithm is polynomially ergodic of order (α, β) when $\alpha + \eta\beta < r$ follows from Proposition 3.2 of [9] and the remarks following it. For the second part of the statement assume by way of contradiction that the algorithm is polynomially ergodic of order (α, β) with $\alpha + \eta\beta > r$.

Consider first the case where $0 < \eta \leq 1$. Choose $0 < \eta' < \eta$ such that also $\alpha + \eta'\beta > r$. By (43) we have for $|x|$ sufficiently large

$$q(x) \leq \frac{1}{|x|^{1+\eta'}}, \tag{44}$$

and it then follows from Theorem 3.3 that $\pi(|x|^{\alpha+\eta'\beta}) < \infty$ which contradicts $\alpha + \eta'\beta > r$.

Consider next the case where $1 < \eta \leq 2$. As in the proof of Proposition 4.1 we have that (41) holds with $H(z, dy) = h(z, y)\mu^{\mathrm{Leb}}(dy)$ for $y \neq 0$ where $h(z, y)$ is given by (42). Let $Y(z)$ be a random variable with distribution $H(z)$. By the same argument as above it is clear that for any $1 < \eta' < \eta$

$$\limsup_{z\to\infty} \mathsf{E}|Y(z)|^{\eta'} < \infty.$$

Now choose $1 < \eta' < \eta$ such that $\alpha + \eta'\beta > r$, and let K be so large that (44) holds for $|x| \geq K$. We then have

$$\mathsf{E}[Y(z)] = I_1(z) + I_2(z), \tag{45}$$

where

$$I_1(z) = \int_0^K q(y)y\left[\left(\frac{z}{z+y}\right)^{1+r} - 1\right]dy, \quad I_2(z) = \int_K^\infty q(y)y\left[\left(\frac{z}{z+y}\right)^{1+r} - 1\right]dy.$$

As in the proof of Proposition 4.1 $I_1(z) \geq -c_1/z$ for some constant $c_1 > 0$. For $I_2(z)$ we get using (44) and the transformation $u = y/z$

$$I_2(z) \geq \int_K^\infty \frac{1}{y^{\eta'}}\left[\left(\frac{z}{z+y}\right)^{1+r} - 1\right]dy = \frac{1}{z^{\eta'-1}}\int_{K/z}^\infty \frac{1}{u^{\eta'}}\left[\left(\frac{1}{1+u}\right)^{1+r} - 1\right]dy \geq -\frac{c_2}{z^{\eta'-1}},$$

where

$$c_2 = \int_0^\infty \frac{1}{u^{\eta'}}\left[1 - \left(\frac{1}{1+u}\right)^{1+r}\right]dy < \infty.$$

The integral is finite since the integrand looks like $u^{-\eta'}$ at infinity and $\eta' > 1$, and like $u^{1-\eta'}$ at the origin and $1 - \eta' > -1$. Hence by (45) there exists constant $c > 0$ such that $\mathsf{E}[Y(z)] \geq -c/z^{\eta'-1}$ for z sufficiently large. By Theorem 3.5 it then follows that $\pi(|x|^{\alpha+\eta'\beta}) < \infty$ which again contradicts $\alpha + \eta'\beta > r$. ∎

Whereas the sufficiency part of, in particular, the last proposition is not straightforward to extend to higher dimensions the necessity part can easily be extended to $\mathbf{R}^d$ and to more general polynomially decaying target densities.

4.1 The Langevin algorithm

When π is a positive, twice differentiable density it has been proposed by e.g. [1] to use the candidate transition kernel

$$Q(x, \cdot) = N(x + \frac{h}{2}\nabla \log \pi(x), h), \tag{46}$$

where $h > 0$, N denotes the normal distribution and ∇ is the differential operator. As shown by [17] this choice performs well when π has exponentially decaying or Gaussian tails giving geometrically ergodic algorithms in both these cases. They also show that the algorithm fails to be geometrically ergodic when $\nabla \log \pi(x) \to 0$ as $|x| \to \infty$, e.g. when π has polynomial tails. In fact, when π has polynomial tails the results of this paper are particularly easy to apply and we get the following result which shows that in this case the convergence properties of the Langevin algorithm are exactly the same as for a symmetric random walk Metropolis algorithm with finite variance.

Let P denote the Markov transition kernel for the Langevin algorithm, i.e. the Metropolis-Hastings algorithm with Q given by (46).

Proposition 4.3 *Assume π is a strictly positive, twice differentiable density on $[0, \infty)$ of form (40). For any $h > 0$ the Langevin algorithm is polynomially ergodic of order (α, β) if and only if $\alpha + 2\beta < r$.*

Proof The "if" part follows from Proposition 4.1 of [9]. For the "only if" part first note that P is μ^{Leb}-irreducible and aperiodic and of random walk-type since $|\nabla \log \pi(x)|$ is bounded away from infinity. Further, it is shown in the proof of Proposition 4.1 of [9] that for x sufficiently large all positive increments are accepted while negative increments are possibly rejected. Thus for x sufficiently large we have for all $y \geq x$

$$N(-\frac{h(1+r)}{2x}, h) \overset{\text{st}}{\leq} P(y, \cdot) - y.$$

It then follows from Theorem 3.4 that P cannot be polynomially ergodic of order (α, β) for any $\alpha, \beta \geq 0$ with $\alpha + 2\beta \geq r$. ■

Acknowledgements

We are grateful to Gareth Roberts for useful discussions. The first author would also like to thank the Danish Research Training Council for financial support including a visit to the University of Minnesota.

References

[1] J. Besag. Comments on "Representations of knowledge in complex systems" by U. Grenander and M. Miller. *J. Royal Stat. Soc. B* **56**, pages 591–592, 1994.

[2] K.S. Chan. Asymptotic behaviour of the Gibbs sampler. *JASA* **88**, pages 320–326, 1993.

[3] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. I, third edition.* Wiley, New York, 1968.

[4] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II, second edition.* Wiley, New York, 1971.

[5] G. Fort and E. Moulines. V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Stat. Prob. Letters* **49**, pages 401–410, 2000.

[6] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, pages 97–109, 1970.

[7] S.F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stoch. Proc. Appl.* **85**, pages 341–361, 2000.

[8] S.F. Jarner and G.O. Roberts. Polynomial convergence rates of Markov chains. To appear in *Ann. Appl. Prob.*

[9] S.F. Jarner and G.O. Roberts. Convergence of heavy tailed MCMC algorithms. http://www.statslab.cam.ac.uk/~mcmc. Submitted for publication.

[10] S.F. Jarner and R.L. Tweedie. Convergence rates and moments of Markov chains associated with the mean of Dirichlet processes. Submitted for publication.

[11] K.L. Mengersen and R.L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Stat.* **24**, pages 101–121, 1996.

[12] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, pages 1087–1091, 1953.

[13] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, London, 1993.

[14] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge, 1984.

[15] E. Nummelin and P. Tuominen. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **12**, pages 187–202, 1982.

[16] G.O. Roberts and R.L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, pages 95–110, 1996.

[17] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**, pages 341–364, 1996.

[18] P. Tuominen and R.L. Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. Appl. Prob.* **26**, pages 775–798, 1994.

[19] R.L. Tweedie. Criteria for rates of convergence of Markov chains with application to queueing and storage theory. In *Probability, Statistics and Analysis*. Ed. J.F.C. Kingman and G.E.H. Reuter, London Mathematics Society Lecture Note Series, Cambridge University Press, 1983.