RATES OF CONVERGENCE OF THE HASTINGS AND METROPOLIS ALGORITHMS[1]

By K. L. Mengersen and R. L. Tweedie

Queensland University of Technology and Colorado State University

We apply recent results in Markov chain theory to Hastings and Metropolis algorithms with either independent or symmetric candidate distributions, and provide necessary and sufficient conditions for the algorithms to converge at a geometric rate to a prescribed distribution π. In the independence case (in $\mathbb{R}^k$) these indicate that geometric convergence essentially occurs if and only if the candidate density is bounded below by a multiple of π; in the symmetric case (in $\mathbb{R}$ only) we show geometric convergence essentially occurs if and only if π has geometric tails. We also evaluate recently developed computable bounds on the rates of convergence in this context: examples show that these theoretical bounds can be inherently extremely conservative, although when the chain is stochastically monotone the bounds may well be effective.

1. Hastings and Metropolis algorithms and Markov chains. The Hastings [6] and Metropolis [9] algorithms allow simulation of a probability density $\pi(x)$ which is only known up to a factor, that is, when only $\pi(x)/\pi(y)$ is known. This is especially relevant when π is the posterior distribution in a Bayesian context: see [3], [21], [20], [13], [22] and [4] for a more detailed introduction.

In this paper we are concerned with the rate of convergence of these algorithms for distributions on $\mathbb{R}^k$, and in particular with finding conditions under which the convergence is geometrically fast.

In order to describe the Hastings and Metropolis algorithms, we first consider a *candidate transition kernel $Q(x, A)$*, $x \in \mathbb{R}^k$, $A \in \mathscr{B}(\mathsf{X})$, satisfying

$$Q(x, A) \geq 0, \qquad Q(x, \mathbb{R}^k) = 1,$$

which generates potential transitions for a discrete-time Markov chain evolving on $\mathsf{X} = \mathbb{R}^k$ equipped with the Borel σ-algebra $\mathscr{B}(\mathsf{X})$. We usually assume that $Q(x, \cdot)$ is absolutely continuous with density $q(x, y)$ with respect to Lebesgue measure μ^{Leb}, except perhaps for an atom $Q(x, \{x\}) > 0$, although this is not necessary and the results below will hold in more general contexts also.

In the Hastings algorithm, introduced in [6], a "candidate transition" generated according to the law Q is then accepted with probability $\alpha(x, y)$ given by

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

Thus actual transitions of the Hastings chain, which we shall denote by $\Phi = \{\Phi_n, n \in \mathbb{Z}_+\}$, take place according to a law P with transition density

$$
(2) \qquad p(x, y) = q(x, y)\alpha(x, y), \qquad y \neq x,
$$

and with probability of remaining at the same point given by

$$
(3) \qquad P(x, \{x\}) = \int q(x, y)[1 - \alpha(x, y)] \mu^{\mathrm{Leb}}(dy) + Q(x, \{x\}).
$$

With this choice of α we have that π satisfies

$$
(4) \qquad \pi(A) = \int \pi(dx)P(x, A) \quad \text{for all } A \in \mathscr{B}(\mathsf{X}).
$$

The Metropolis algorithm, which dates back to [9], is a special case of the Hastings algorithm, utilizing a symmetric candidate transition Q, that is, one for which $q(x, y) = q(y, x)$.

The questions we address here are whether the n-step transition probabilities of Φ, defined by

$$
P^n(x, A) = \mathrm{P}(\Phi_n \in A | \Phi_0 = x), \qquad n \in \mathbb{Z}_+, x \in \mathsf{X}, A \in \mathscr{B}(\mathsf{X}),
$$

converge to π at a geometric rate, and if so what bounds can be placed on that rate.

As has been observed often (cf. [20], [22], and [5]), the Hastings algorithms (and other algorithms of the Markov chain Monte Carlo type, such as the Gibbs sampler) can often be analyzed using the theory of φ-*irreducible* Markov chains; that is, chains for which there exists a measure φ such that

$$
(5) \qquad \varphi(A) > 0 \Rightarrow \sum_n P^n(x, A) > 0, \qquad x \in \mathsf{X}.
$$

For chains with the structure (2) and (3), it is simple to check the following result.

LEMMA 1.1.   *The chain Φ is π-irreducible if*

$$
(6) \qquad \pi(y) > 0 \Rightarrow q(x, y) > 0, \qquad x \in \mathsf{X}.
$$

Weaker but less simple conditions are possible (see [5], [13] and [16], among others).

The theory of φ-*irreducible* chains is described in Nummelin [12] or Meyn and Tweedie [10]. As discussed there, in order to develop criteria for rates of convergence, we need the concepts of *small sets* and *aperiodicity*. It is known

as in [10], Chapter 5, that for a φ-irreducible chain, any set A with $\varphi(A) > 0$ contains a small set; that is, a set C such that, for some $\delta > 0$ and $n > 0$ and some probability measure ν concentrated on C,

$$(7) \qquad P^n(x, \cdot) \geq \delta\nu(\cdot), \qquad x \in C.$$

The chain is called aperiodic if, for some one (and then every) small set with $\varphi(C) > 0$ the g.c.d. of the n for which (7) holds is 1.

Again, these concepts are easy to verify for Hastings chains under suitable conditions.

LEMMA 1.2. *The Hastings chain Φ is aperiodic and every compact set C with $\mu^{\mathrm{Leb}}(C) > 0$ is small if $\pi(x)$ and $q(y, x)$ are positive and continuous for all x, y.*

PROOF. Suppose that C is compact and nonempty; by positivity and continuity we have $d = \sup_{x \in C} \pi(x) < \infty$ and $\varepsilon = \inf_{x, y \in C} q(x, y) > 0$. Choose $B \subseteq C$. By construction, for fixed x if

$$R_x(B) = \left\{ y \in B \colon \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)} < 1 \right\}$$

and $A_x(B) = A \setminus R_x(B)$, we have

$$
\begin{aligned}
P(x, A) &= \int_{R_x(B)} q(x, y) \min\left\{ \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}, 1 \right\} \mu^{\mathrm{Leb}}(dy) \\
&\quad + \int_{A_x(B)} q(x, y) \min\left\{ \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}, 1 \right\} \mu^{\mathrm{Leb}}(dy) \\
(8) \qquad &= \int_{R_x(B)} \frac{\pi(y)}{\pi(x)} q(y, x) \mu^{\mathrm{Leb}}(dy) + \int_{A_x(B)} q(x, y) \mu^{\mathrm{Leb}}(dy) \\
&\geq \frac{\varepsilon}{d} \int_{R_x(B)} \pi(y) \mu^{\mathrm{Leb}}(dy) + \varepsilon \int_{A_x(B)} \frac{\pi(y)}{d} \mu^{\mathrm{Leb}}(dy) \\
&= \frac{\varepsilon}{d} \pi(A).
\end{aligned}
$$

Thus C is small and since (7) holds for $n = 1$ the chain is automatically aperiodic. □

The crucial observation in the Hastings context is now that, from (4), we have that the distribution π is *invariant* for Φ. Consequently, as discussed in [10], Chapter 13, it follows immediately that in the φ-irreducible aperiodic case, for π-a.e. initial condition $x \in \mathsf{X}$,

$$(9) \qquad \| P^n(x, \cdot) - \pi \| \to 0,$$

where we define the *total variation norm* of a signed measure μ by $\|\mu\| = \sup_{A \in \mathscr{B}(\mathsf{X})} |\mu(A)|$.

From [10] we now outline two results which will enable us to analyze the convergence properties of Hastings algorithms. First, we note that, although the convergence in (9) is not in general uniform in the starting point of the chain, there is a useful class of chains for which such uniformity does hold; that is, for which there exists a sequence $r(n) \to 0$ as $n \to \infty$ such that, for all x,

$$(10) \qquad \|P^n(x, \cdot) - \pi\| \le r(n).$$

Such chains are called *uniformly ergodic*.

The following criteria for uniform ergodicity are taken from [10], Chapter 16.

THEOREM 1.3. *For any Markov chain* Φ *the following are equivalent*:

(i) *The chain is aperiodic and Doeblin's condition holds*; *that is, there is a probability measure* ϕ *on* $\mathscr{B}(\mathsf{X})$ *and* $\varepsilon < 1$, $\delta > 0$, $m \in \mathbb{Z}_+$ *such that, whenever* $\phi(A) > \varepsilon$,

$$(11) \qquad \inf_{x \in \mathsf{X}} P^m(x, A) > \delta.$$

(ii) *The whole state space* X *is small*; *that is, for some* $\delta > 0$ *and some* $m \ge 1$ *and some probability measure* ν,

$$(12) \qquad P^m(x, A) \ge \delta \nu(A), \qquad x \in \mathsf{X}, \ A \in \mathscr{B}(\mathsf{X}).$$

(iii) *For some small set* C *we have*

$$\sup_{x \in \mathsf{X}} \mathsf{E}_x[\tau_C] < \infty,$$

where τ_C *is the first return time to* C.

(iv) *The chain* Φ *is uniformly ergodic*.

When any of (i)–(iv) *hold the convergence in* (10) *is geometrically fast*, *and for any x we can bound the rate of convergence using* (12) *by*

$$(13) \qquad \|P^n(x, \cdot) - \pi\| \le (1 - \delta)^{[n/m]}.$$

In the next two sections we shall apply this result to the Hastings algorithm with Q chosen independently of x and to the Metropolis algorithm in which Q is chosen to be symmetric.

Except on compact spaces, however, most chains are not uniformly ergodic: this is intuitively clear from Theorem 1.3(iii). Even so it is possible to find conditions under which the convergence in (9) takes place at a geometric rate even if there is dependence on the initial state Φ_0. The following theorem is taken from [10], Chapters 15 and 16.

THEOREM 1.4. *Suppose that* Φ *is* φ-*irreducible and aperiodic. Then the following are equivalent*:

(i) *There is a function* $V \ge 1$, *finite at least for one x, and a small set* C *such that for some* $\lambda_C < 1$, $b_C < \infty$ *the "geometric drift condition"*

$$(14) \qquad PV \le \lambda_C V + b_C \mathbb{1}_C$$

is satisfied.

(ii) *For some small set C with $\varphi(C) > 0$, there exists $\kappa > 1$ such that*

$$\sup_{x \in C} \mathsf{E}_x[\kappa^{\tau_C}] < \infty. \tag{15}$$

(iii) *The chain is geometrically ergodic in the sense that there is a function $V \geq 1$, finite π-a.e. [which can be taken as the V in (i)], and constants $r > 1$ and $R < \infty$ such that, for all $n \in \mathbb{Z}_+$ and all x such that $V(x)$ in (14) is finite,*

$$\|P^n(x, \cdot) - \pi\|_V \leq RV(x)r^{-n}, \tag{16}$$

where for any signed measure μ the V-norm is defined as $\|\mu\|_V = \sup_{|g| \leq V} |\int \mu(dy)g(y)|$.

It is valuable to note explicitly that this result shows:

(a) When (14) holds for an everywhere finite function V, then (16) and thus also (9) hold for all rather than almost all x.

(b) The theorem covers convergence, not just of total variation norms and hence of expectations of bounded functions, but also typically of unbounded moments: it ensures that for all functions $g \leq V$ we have $\mathsf{E}_x[g(\Phi_n)] \to \mathsf{E}_\pi[g(\Phi_n)]$ in a uniform manner, and in very many cases (see Section 5 for example) V will be such that this covers higher-order moments in the usual sense.

We apply both the necessary and sufficient aspects of the drift condition (14) to the Metropolis algorithm in Theorems 3.2 and 3.3, while the necessity of (15) is applied in Theorem 2.1 in the case of Q independent of x. In the related Gibbs sampling framework, Chan [5] has also recently exploited (14).

In later sections we also address the question of bounding the rate of convergence in a computable way in (16) using some new Markov chain methods.

2. Convergence of Hastings independence algorithms. We first consider the simplest algorithm defined by (2). This is the *independence* case in which

$$q(x, y) = q(y), \qquad x, y \in \mathsf{X}. \tag{17}$$

To avoid complications, we assume $\pi(x) > 0$ and $q(x) > 0$ for all x so that the acceptance probabilities here take the form

$$\alpha(x, y) = \min\left\{\frac{\pi(y)}{\pi(x)}\frac{q(x)}{q(y)}, 1\right\}. \tag{18}$$

We prove, using Theorems 1.3 and 1.4, the following result.

THEOREM 2.1. *The independent Hastings algorithm is uniformly ergodic if there exists $\beta > 0$ such that*

$$\frac{q(y)}{\pi(y)} \geq \beta, \qquad y \in \mathsf{X}, \tag{19}$$

and then

(20) $$\|P^n(x, \cdot) - \pi\| \le (1 - \beta)^{[n]}.$$

Conversely, if $\operatorname{ess\,inf}\{q(y)/\pi(y)\} = 0$ *in* π-*measure, then the algorithm is not even geometrically ergodic.*

PROOF.   Suppose that (19) holds. For fixed x, we have either $\alpha(x, y) = 1$, in which case $p(x, y) = q(y) \ge \beta\pi(y)$; or else

$$p(x, y) = q(y)\frac{\pi(y)}{\pi(x)}\frac{q(x)}{q(y)} = \frac{q(x)}{\pi(x)}\pi(y) \ge \beta\pi(y).$$

In either case (12) holds and the result follows from Theorem 1.3.

Suppose now that (19) fails, so that $\pi(D_n) > 0$ for all n, where

$$D_n = \left\{x\colon \frac{q(x)}{\pi(x)} \le n^{-1}\right\}.$$

Suppose the chain is geometrically ergodic, and let C be a small set satisfying (15). Since the chain is π-irreducible by construction, for any n there is at least one $x_n \in C$ and some $m \in \mathbb{Z}_+$ such that

(21) $$\mathsf{P}_{x_n}(\Phi_m \in D_n, \tau_C > m) > 0.$$

For fixed x, write $A_x = \{y\colon [\pi(y)/q(y)][q(x)/\pi(x)] \ge 1\}$ and $R_x = \{y\colon [\pi(y)/q(y)][q(x)/\pi(x)] < 1\}$ for the regions where a transition using Q is accepted and is possibly rejected, respectively. Suppose that $x \in D_n$. Then we have, for $y \in A_x$,

$$1 \le [\pi(y)/q(y)][q(x)/\pi(x)] \le n^{-1}[\pi(y)/q(y)]$$

and so

$$P(x, A_x \setminus \{x\}) = \int_{A_x} q(y)\mu^{\text{Leb}}(dy) \le \int_{A_x} n^{-1}\pi(y)\mu^{\text{Leb}}(dy) \le n^{-1}.$$

Similarly, for $y \in R_x$,

$$R(x, R_x) = \int_{R_x} q(y)[\pi(y)/q(y)][q(x)/\pi(x)]\mu^{\text{Leb}}(dy)$$

$$\le \int_{R_x} n^{-1}\pi(y)\mu^{\text{Leb}}(dy) \le n^{-1}.$$

Thus we have $P(x, \{x\}) \ge 1 - 2/n$ for every $x \in D_n$, and so by iterating

$$\mathsf{P}_x(\tau_C \ge k) \ge (1 - 2/n)^k.$$

Thus the radius of convergence of

$$\mathsf{E}_x[z^{\tau_C}] = \sum_k \mathsf{P}_x(\tau_C \ge k)z^k$$

is smaller than $n/(n-2)$. But now from (21) it follows that for x_n we also have that $\mathsf{E}_{x_n}[z^{\tau_c}]$ diverges outside $n/(n-2)$. Since we can take n arbitrarily large, it follows that (15) cannot hold for any $\kappa > 1$. Thus we have a contradiction and the chain is not geometrically ergodic. $\square$

The sufficiency of (19) was noted in [22], but the necessity appears to be new. The fact that without this bound the chain may tend to "stick" in regions of low density is of considerable practical importance and is not merely a curiosity: seemingly sensible procedures give this behavior, as the following example illustrates.

EXAMPLE 1 (Estimating normal densities using independent normal candidates).   Consider the situation in which π is an $N(0,1)$ distribution. If we do not know the true mean, but do know that the underlying distribution is normal, then one choice for the candidate density might be an independent normal density centered at some known fixed value: so, for example, $Q(x, \cdot)$ might be taken as an $N(1,1)$ distribution for each x.
    The acceptance ratios then follow from the calculation that

$$[\pi(y)/\pi(x)][q(x)/q(y)] = e^{x-y}.$$

Hence $R_x = (x, \infty)$ so that moves to the right are possibly rejected, but moves to the left are *always* accepted, due to the relative positions of the means of the candidate and target distributions.
    Now (19) fails in this example, and we find that the chain moves to the negative axis with positive probability and then leaves the sets $(-\infty, -n]$ more and more slowly if it enters them, and the algorithm cannot converge geometrically quickly.
    If, on the other hand, we know the mean but not the variance of π, then we might take $Q(x, \cdot)$ to be an $N(0, \sigma^2)$ distribution for some known $\sigma^2 > 1$. Then we have that $q(x)/\pi(x) \geq \sigma^{-1}$ and we can use (20) to find that convergence occurs at rate at least $(1 - \sigma^{-1})^n$, thus quantifying the role of choosing a σ close to the true value.

3. Convergence of Metropolis algorithms.   We next consider the algorithm defined by (2) in the *symmetric* case where

$$(22) \qquad\qquad q(x, y) = q(y, x), \qquad x, y \in \mathsf{X}.$$

The most common usage of such chains occurs (cf. [22]) if Q is not merely symmetric but satisfies

$$(23) \qquad\qquad q(x, y) = q(x - y) = q(y - x)$$

for some fixed density q; that is, the candidate increment distributions are identical, and the chain would be a random walk if it were not for the acceptance probabilities, which take the form

$$(24) \qquad \alpha(x, y) = \begin{cases} \min\left\{ \dfrac{\pi(y)}{\pi(x)}, 1 \right\}, & \pi(x)q(x, y) > 0, \\ 1, & \pi(x)q(x, y) = 0. \end{cases}$$

This Metropolis algorithm typically fails to be uniformly ergodic when the state space is not bounded. We have the following result.

THEOREM 3.1.   *If Q satisfies* (23) *on* $\mathbb{R}^k$, *then the Metropolis algorithm is not uniformly ergodic for any π.*

PROOF.   It is easiest to see this through the necessity of (12). Suppose the chain is uniformly ergodic. It is obvious by considering rejected moves that, for any A not containing x,

$$P^m(x, A) \le \sum_{j=1}^{m} Q^j(x, A).$$

Considering the j-step densities (with respect to Lebesgue measure), we thus have from (12) that we can find some bounded nonempty set C', and $\delta' > 0$, such that we would have

$$\sum_{j=1}^{m} q^j(x, y) \ge \delta'$$

for $y \in C'$ and for every x not in C'. But then, by the symmetry of Q and its iterates, we would have for $y \in C'$ that $\sum_{j=1}^{m} q^j(y, x) \ge \delta'$ for every x not in C', which is impossible since the space is not bounded.   □

This result holds for many asymmetric candidate distributions also. To see this intuitively, consider the condition in Theorem 1.3(iii). Note that the Hastings algorithm will often move more slowly to hit a set C in the "center" (due to moves "away") than a random walk that rejects *all* moves "away": it is well known [10] that for all random walks the number of steps needed to reach the center of the space is essentially linear in the distance from the center, and so the number of steps needed for the Hastings algorithm to reach the center of the space is also unbounded. Therefore, Theorem 1.3(iii) cannot hold for such algorithms.

Although, as just shown, symmetric or Metropolis algorithms are not uniformly ergodic in general, Theorem 1.4 can provide useful criteria for geometric ergodicity under some widely used conditions. In the remainder of this section we shall consider only distributions on $\mathbb{R}$. Some of the results can be extended, although not simply, to higher dimensions, and this has been done recently in [16].

Let us define the class $\mathscr{M}$ of densities with the properties that $\pi(x)$ is continuous and $\pi(x) > 0$ for all $x \in \mathbb{R}$. Call such a density π *log-concave in the tails* if there exists $\alpha > 0$ and some x_1 such that, for all $y \geq x \geq x_1$,

$$(25) \qquad\qquad \log \pi(x) - \log \pi(y) \geq \alpha(y - x),$$

and similarly, all $y \leq x \leq -x_1$,

$$(26) \qquad\qquad \log \pi(x) - \log \pi(y) \geq \alpha(x - y).$$

THEOREM 3.2. *Suppose $\pi(x) \in \mathscr{M}$ is log-concave in the tails and symmetric. Then for any Q which satisfies (23) with continuous density $q(x) > 0$ the Metropolis algorithm is geometrically ergodic, with*

$$\| P^n(x, \cdot) - \pi \|_V \leq R r^{-n} V(x)$$

for some $R < \infty$, $r > 1$ and $V(x) = e^{s|x|}$ for any $s < \alpha$, where α is as in (25) and (26).

If π is not symmetric, then the same conclusion holds provided that the candidate density satisfies also $q(x) \leq b e^{-\alpha|x|}$ for some finite b.

PROOF. The positivity and continuity conditions on π and q ensure that Lemmas 1.1 and 1.2 hold. Hence every compact set of positive measure is small and so it suffices to check (14) with $C = [-x, x]$ for some $x > 0$.

For each x define again the sets

$$A_x = \{ y : \pi(x) \leq \pi(y) \}, \qquad R_x = \{ y : \pi(x) > \pi(y) \}$$

for the acceptance and (possible) rejection regions for the chain started from x. If $\pi \in \mathscr{M}$ and is log-concave in the tails, these sets are relatively simple. We have immediately from the construction of the algorithm that when π is symmetric then there exists x_0 such that $A_x = \{|y| \leq |x|\}$ for $|x| > x_0$, while in the nonsymmetric case there exists x_0 such that for $x > x_0$ the set $(x, \infty) \subseteq R_x$ and the set $(-x, x) \subseteq A_x$; similarly for $x < -x_0$ the set $(-\infty, x) \subseteq R_x$ and the set $(x, -x) \subseteq A_x$.

Let us choose a suitably large $x^* \geq x_0 \vee x_1$, where x_1 is the point at which log-concavity begins; the other properties of x^* become apparent below. We first consider the case of symmetric π. We will apply Theorem 1.3(ii) using the test function $V(x) = e^{s|x|}$ for any $s < \alpha$.

Identifying moves to A_x, R_x and $\{x\}$ separately, we have, for $x \geq x^*$,

$$\lambda_x := \int P(x, dy)V(y)/V(x)$$

$$= \int_{|y| \leq x} Q(x, dy)\exp(s(|y| - x))$$

$$+ \int_{|y| > x} Q(x, dy)\exp(s(|y| - x))[\pi(y)/\pi(x)]$$

(27)

$$+ \int_{|y| > x} Q(x, dy)[1 - \pi(y)/\pi(x)]$$

$$= 1 + \int_{|y| \leq x} Q(x, dy)[\exp(s(|y| - x)) - 1]$$

$$+ \int_{|y| > x} Q(x, dy)[\exp(s(|y| - x)) - 1][\pi(y)/\pi(x)].$$

Now, using the fact that log-concavity implies, for $y \geq x \geq x^*$,

(28)
$$\frac{\pi(y)}{\pi(x)} \leq e^{-\alpha(y-x)},$$

we have, for $x \geq x^*$ and $s < \alpha$,

$$\lambda_x \leq 1 + \int_0^x Q(x, dy)[\exp(s(y - x)) - 1]$$

(29)

$$+ \int_x^{2x} Q(x, dy)\exp(-\alpha(y - x))[\exp(s(y - x)) - 1]$$

$$+ Q(x, (2x, \infty)) + Q(x, (-\infty, 0)).$$

The last two terms in (29) total less than $2\int_x^\infty q(z)\mu^{\text{Leb}}(dz)$ and so for large x^* can be made arbitrarily small. Thus by symmetry we will have drift toward $C = (-x^*, x^*)$ as in (14) if we have the sum of the second and third terms in (29) strictly bounded below 0 for all $x > x^*$. Using (23), we can write these terms as

$$\int_0^x q(z)[e^{-sz} - 1 + e^{-(\alpha - s)z} - e^{-\alpha z}]\mu^{\text{Leb}}(dz)$$

$$= -\int_0^x q(z)[1 - e^{-sz}][1 - e^{-(\alpha - s)z}]\mu^{\text{Leb}}(dz)$$

and since the integrand on the right is positive and increasing as z increases, we find that, for suitably large x^*, λ_x in (29) is strictly less than 1 as required.

For $0 \leq x \leq x^*$, the right-hand side of (27) is similarly less than

$$1 + 2\int_{x^*}^\infty q(z)\mu^{\text{Leb}}(dz) + 2\exp(sx^*)\int_0^{x^*} q(z)\mu^{\text{Leb}}(dz).$$

For negative x the same calculations are valid by symmetry, and so (14) holds and the result is proved.

In the case of nonsymmetric π, the proof is similar. We need now, however, to control the term (for $x > 0$)

$$\int_{y \leq -x} Q(x, dy)\left[\exp(s(|y| - x)) - 1\right]\left[1 \vee \pi(y)/\pi(x)\right]$$

in (27), and this is guaranteed to be negligible if $q(x) \leq b \exp(-\alpha|x|)$ as we have assumed. □

This theorem shows that a criterion for geometric ergodicity in $\mathbb{R}$ is that the target distribution π have exponentially decreasing tails. In higher dimensions this also seems plausible, but in [16] we show that if indeed π decreases exponentially as one moves away from the origin, then the Metropolis algorithm is geometrically ergodic, provided also that π has smooth contours in an explicit way. Without the extra contour condition, the chain can fail to be geometrically ergodic.

We now show in $\mathbb{R}$, geometric convergence essentially only occurs if the tails of π decrease exponentially.

THEOREM 3.3. *Suppose that $\pi \in \mathcal{M}$ and that Q satisfies (23) and also has finite absolute first moment $\int |z| q(z) \mu^{\mathrm{Leb}}(dz) = \gamma < \infty$.*

If the Metropolis algorithm is geometrically ergodic, then there exists some $s > 0$ such that $\int \pi(dy) e^{s|y|} < \infty$.

PROOF. Since $\pi \in \mathcal{M}$, as in the previous proof all sets $(-x, x)$ are small. Thus for any $x > 0$ by Theorem 1.4(ii) there exists a function $V \geq 1$ and $\lambda < 1$ such that $PV(y) \leq \lambda V(y)$ for $y \geq x$, so that by Jensen's inequality

$$\mathsf{E}_y\left[\log(V(\Phi_1)/V(y))\right] \leq \log\left(\mathsf{E}_y\left[V(\Phi_1)/V(y)\right]\right) \leq \log \lambda < 0.$$

Thus we have

$$(30) \qquad \mathsf{E}_y\left[\log(V(\Phi_1))\right] \leq \log(V(y)) + \log \lambda, \qquad y \geq x;$$

that is, $\log V$ is itself a Foster–Lyapunov function for (simple) ergodicity. From Theorem 11.3.4 of [10] it follows from (30) that, for $\varepsilon = -\log \lambda$,

$$(31) \qquad \log V(y) \geq \mathsf{E}_y[\tau_x]/\varepsilon,$$

where τ_x denotes the first hitting time of Φ on $(-\infty, x)$.

Now let $\hat{Q}$ denote the transition law of the one-sided random walk with

$$(32) \qquad \hat{Q}(\{0\}) = 1/2, \qquad \hat{Q}(A) = Q(A), \qquad A \subseteq (-\infty, 0),$$

and let $\hat{\tau}_x$ denote the first hitting times on $(-\infty, x)$ of the one-sided random walk. Then we have

$$(33) \qquad \mathsf{P}_y(\tau_x \leq m) \leq \mathsf{P}_y(\hat{\tau}_x \leq m)$$

by a stochastic monotonicity argument since, starting from $y > x$, the Metropolis chain is always (in distribution) stochastically larger than the one-sided random walk.

The one-sided random walk with law (32) has mean increment $\gamma/2$. By a standard random walk argument we have that, for large y,

$$\mathsf{E}_y[\hat{\tau}_x] \sim \frac{\gamma}{2}(y - x).$$

The stochastic ordering (33) implies that $\mathsf{E}_y[\tau_x] \geq \mathsf{E}_y[\hat{\tau}_x]$, and so it follows from (31) that, for all large y and some $\eta > 0$,

$$V(y) \geq \exp\left(\frac{\gamma}{2\varepsilon + \eta}(y - x)\right).$$

A symmetric argument for the left tail shows that for large values of $-y$ we also have

$$V(y) \geq \exp\left(\frac{\gamma}{2\varepsilon + \eta}(-y + x)\right).$$

Finally, we use the fact that, since V satisfies the drift condition (14), then from Theorem 14.0.1 of [10], we must have $\int \pi(dy)V(y) < \infty$, and the result is proved. $\square$

Note that this necessity result demands no monotonicity properties on π at all.

If the density $\pi(x)$ is sufficiently regular, then we have an elegant dichotomy into geometrically and nongeometrically ergodic chains. Let $\mathcal{M}^+$ denote those symmetric densities in $\mathcal{M}$ such that $\theta(x) := (d/dx)\log(\pi(x))$ is defined for all sufficiently large x, and for which

$$\lim_{x \to \infty} \theta(x) = \theta_\infty$$

is defined (although possibly $-\infty$). Then we have the following result.

THEOREM 3.4. *Suppose $\pi \in \mathcal{M}^+$ and Q satisfies (23), with q continuous and with finite absolute first moment. Then the Metropolis algorithm is geometrically ergodic if and only if $\theta_\infty < 0$.*

PROOF.   To see sufficiency, note that for all sufficiently large $x < y$ we have

(34) $\log \pi(y) - \log \pi(x) = \int_x^y \theta(z) \mu^{\mathrm{Leb}}(dz) \leq \frac{\theta_\infty}{2}(y - x),$

so that (25) holds, and by symmetry of the left tail we have that Theorem 3.2 holds.

Conversely, if $\theta_\infty = 0$, then, for any $\delta > 0$, for large x and all $z > 0$, as in (34),

$$\log \pi(x + z) - \log \pi(x) \geq -\delta z,$$

so that we have

$$\pi(x + z)\exp(\delta z) \geq \pi(x).$$

Integrating both sides shows that $\int_x^\infty \pi(y)\exp(\delta(y))\mu^{\mathrm{Leb}}(dy) = \infty$; and since δ is arbitrary we have from Theorem 3.3 that the chain cannot be geometrically ergodic. $\square$

Note, in particular, that we have shown that if the tails of π are at least exponentially decreasing, then the Metropolis algorithm is geometrically ergodic; but if the tails are only polynomial, with say $\pi(x) \propto x^{-n}$, then since $\theta(x) \propto n/x$ the Metropolis algorithm is not geometrically ergodic. Thus if π is normal we have geometric convergence, but if π is a t-distribution this will not occur.

We conjecture that convergence rates in the case of polynomial tails of order n will be polynomial of order $n - 1$; this is plausible based on similar results in [23].

4. Computable rates of convergence: chains with atoms. Unlike (13), there exist no simple conditions in general for rates of convergence in the nonuniform case. However, recent results in [11], [7] and [19] enable us to derive bounds on the rate of convergence which are actually computable.

The bounds are simpler when the state space is countable or in some other way admits a reachable atom in the space, that is, essentially, when there is a single point α such that $\pi(\alpha) > 0$. They are even simpler for chains possessing a stochastic monotonicity property, that is, where we have a chain on $[0, \infty)$ rather than on $\mathbb{R}^k$, and where, for all $z \geq 0$,

$$(35) \qquad P(x,[0, z]) \geq P(y,[0, z]), \qquad x \leq y.$$

The next theorem gives the explicit form for these bounds which we can use when conditions permit.

THEOREM 4.1.   *Suppose that for some atom $\alpha \in \mathscr{B}(\mathsf{X})$ we have $\lambda < 1$, $b < \infty$ and a function $V \geq 1$ such that*

$$(36) \qquad \int P(x, dy)V(y) \leq \lambda V(x) + b\,\mathbb{1}_\alpha(x)$$

and also that the atom is strongly aperiodic, that is, for some δ,

$$(37) \qquad P(\alpha, \alpha) > \delta.$$

(i) *For any $r < [\vartheta(j)]^{-1}$,*

$$(38) \qquad \|P^n - \pi\|_V \leq V(x)\frac{1}{1 - r\vartheta(j)}r^{-n}, \qquad n \in \mathbb{Z}_+,$$

where $\vartheta(j) = 1 - M(j)^{-1}$, and, for each $j = 1, 2, 3$,

$$(39) \qquad M(j) = \frac{1}{(1-\lambda)^2}\left[1 - \lambda + b + b^2 + \zeta_\alpha(j)\big(b(1-\lambda) + b^2\big)\right];$$

here the three bounds

$$\zeta_\alpha(1) \geq \zeta_\alpha(2) \geq \zeta_\alpha(3)$$

are given by

$$(40) \qquad\qquad\qquad \zeta_\alpha(1) = \frac{32 - 8\delta^2}{\delta^3}\left(\frac{b}{1-\lambda}\right)^2,$$

$$(41) \qquad\qquad\qquad \zeta_\alpha(2) = \left(\frac{b}{1-\lambda}\right)^2 \frac{1}{2\,\delta p(2)} - 1,$$

where $p(2) = \mathsf{P}_\alpha(\tau_\alpha = 2)$, and

$$(42) \qquad\qquad\qquad \zeta_\alpha(3) = \frac{1}{2[\pi(\alpha)]^2 \delta p(2)} - 1.$$

(ii) *If the chain is stochastically monotone on $[0, \infty)$, as in (35), and (36) holds for the atom $\{0\}$ and monotone V, then, for any $r \leq \lambda^{-1}$,*

$$(43) \qquad \|P^n(x, \cdot) - \pi\|_V \leq \frac{2}{1 - r\lambda}\left[V(x) + \frac{b}{1-\lambda}\right] r^{-n}, \qquad n \in \mathbb{Z}_+.$$

The result using $\zeta_\alpha(1)$ is in Theorems 2.1 and 2.2 of [11], and requires only that we know λ, b and δ; the results using $\zeta_\alpha(2)$ or $\zeta_\alpha(3)$ follow from (42) of [11], and require that we know at least a lower bound on $p(2)$, and for $\zeta_\alpha(3)$ that we also know $\pi(\alpha)$. This will sometimes be the case, but in general we have only the coarse bound in (40) available.

The stochastic monotonicity result holds also for nonmonotone V with the same rate λ^n, but the constant is slightly messier. This version is from Theorem 2.3 of [19]: the main thrust of the result is proved in Theorem 3.1 of [7]. The value of λ as a bound on the rate of convergence is in many cases best possible, as is also shown in [7].

There have recently been a number of other results for rates of convergence of Markov chains. Rosenthal [17] uses an elegant coupling method to give bounds on the total variation norm, although they do not cover convergence of unbounded functions of the chain. Baxendale [2] has similar bounds, using methods related to those in [11]. Neither of these seem to be uniformly better than those we will apply here [14]. Amit [1] considers quite different approaches in an L_2 setting, as do Schervish and Carlin [18], but these appear rather more specialized in implementation.

We illustrate Theorem 4.1 with an application to a distribution concentrated on the nonnegative integers $\mathbb{Z}_+$. This gives some idea of the difference in effectiveness of the two bounds given by (43) and (39).

EXAMPLE 2 (Estimating a geometric distribution using a symmetric candidate). Suppose that π is the geometric distribution given by

$$\pi(j) = (1 - p)p^j, \qquad j = \{0, 1, 2, \ldots\},$$

and that we use as a candidate distribution the simple random walk with transitions given by

$$q(i, i - 1) = q(i, i + 1) = 1/2, \qquad i > 0; \qquad q(0, 1) = q(0, 0) = 1/2.$$

It is easily verified that this is a symmetric candidate and so the acceptance probabilities are given by $\alpha(0, 1) = p$ and

$$\alpha(i, i - 1) = 1, \qquad \alpha(i, i + 1) = p, \qquad i \geq 1.$$

For this Metropolis algorithm we then find the transition law is given by

$$
\begin{aligned}
p(i, i - 1) &= 1/2, \qquad p(i, i + 1) = p/2, \\
p(i, i) &= (1 - p)/2, \qquad i > 0; \\
p(0, 0) &= 1 - p/2, \quad p(0, 1) = p/2.
\end{aligned}
\tag{44}
$$

It is clear by construction that this is stochastically monotone chain. In order to verify geometric ergodicity and to compare the bounds in Theorem 4.1 for the rate of convergence, we use (36) and the drift function $V(j) = \beta^j$, where the value of β is to be determined. We have, for $j > 0$, that the drift constant in (36) for each β is given by

$$\lambda_\beta := \sum_j p(i, j)V(j)/V(i) = 1/2\beta + (1 - p)/2 + p\beta/2. \tag{45}$$

To use Theorem 4.1(ii), we need to minimize λ_β and this occurs at $\beta = 1/\sqrt{p}$, which gives a value of $\lambda_{\min} = \sqrt{p} + (1 - p)/2$.

From Theorem 4.1(ii), $\lambda_{\min}$ gives a best value (for this choice of geometric function V) of the bound on the rate of convergence.

At 0, we then get the value of b in (36) given by $b = (1 - \sqrt{p})/2$, and we can also use (39) to get upper bounds on the rate of convergence, ignoring the stochastic monotonicity structure, since we know $\delta = p(0, 0)$ in this case; we also know that $\pi(0) = 1 - p$ and $p(2) = p/4$, so that we can use the version of the bounds with $\vartheta(2)$ and $\vartheta(3)$ using (41) and (42).

Table 1 shows the calculation of bounds for various values of the parameter p of the target geometric distribution. Note from the converse results of

TABLE 1

p	$\lambda_{\min}$	$\vartheta(3)$	$\vartheta(2)$	$\vartheta(1)$
0.8	0.99	0.99992	0.99997	0.999999
0.6	0.97	0.99892	0.99960	0.999974
0.4	0.93	0.99562	0.99808	0.999750
0.2	0.85	0.99046	0.99503	0.998344
0.1	0.77	0.99026	0.99416	0.995750

[7] that this is a case where the exact rate of convergence is indeed $\lambda_{\min}$. It is now obvious that the bounds from Theorem 4.1(i) are much worse than $\lambda_{\min}$; for example, at $p = 0.4$ we know that the algorithm converges at rate $\lambda = 0.93$, but the bounds from (39) range from $\vartheta(3) = 0.9956$ to $\vartheta(1) = 0.99975$. This is due, as noted in [11], largely to the large value of $\zeta_\alpha(j)$, and more work needs to be done to improve this if the bounds are to be of value in practice in evaluating Hastings–Metropolis algorithms.

One can also look at this effect by defining $N_\varepsilon(\rho)$ to be the number of steps needed to bring the total variation distance below a given specified ε, that is, by setting $\rho^{N_\varepsilon(\rho)} = \varepsilon$. This is tabulated in Table 2 for the three bounds and for values of $\varepsilon = 0.01$ and $\varepsilon = 0.001$. The techniques illustrated by this example extend to more general monotone decreasing π, since the stochastic monotonicity will continue to hold provided that also $\pi(i + 1)/\pi(i)$ is nonincreasing. Thus the effective rate of convergence results using Theorem 4.1(ii) can be applied to a reasonable class of target distributions, and for these realistic and usable bounds can be found.

5. Rates of convergence: general case.   Even when there is no atom in the space, it is shown in [11], [2] and [17] that under very general conditions it is still possible to get analytical bounds on the rates of convergence which are actually computable and which are applicable to general Hastings algorithms where only ratios of densities are known. These extend those of Theorem 4.1(i): from Theorem 2.3 of [11] we have the following result.

THEOREM 5.1.   *Suppose that in* (14) *the set C satisfies* (7) *with $n = 1$ so that*

(46) $P(x, \cdot) \geq \delta_C \nu(\cdot), \qquad x \in C,$

and assume

(47) $\eta := \inf_{x \in C} P(x, C) - \delta_C > 0, \qquad v_C = \sup_{x \in C} V(x) < \infty.$

Define the constants

$$b_C^* = [1 - \delta_C]^{-1}[b_C + \delta_C(\lambda_C v_C - \nu(V))],$$

(48) $\gamma_C = [\delta_C \eta]^{-1}[1 - \delta_C]b_C^*, \qquad b_\alpha^* = \nu(V) - \lambda_C,$

$$\check{\lambda} = [\lambda_C + \gamma_C]/[1 + \gamma_C] < 1, \qquad \check{b} = b_\alpha^* + \gamma_C < \infty.$$

TABLE 2

p	$N_\alpha(\lambda_{\min})$	$N_\alpha(\vartheta(3))$	$N_\alpha(\vartheta(2))$	$N_\alpha(\vartheta(1))$	$N_\alpha(\lambda_{\min})$	$N_\alpha(\vartheta(3))$	$N_\alpha(\vartheta(2))$	$N_\alpha(\vartheta(1))$
		$\varepsilon = 0.01$				$\varepsilon = 0.001$		
0.8	824	56192	179379	5534447	1236	84287	269068	8301671
0.6	179	4269	11373	180046	268	6403	17060	270070
0.4	66	1048	2396	18403	99	1572	3594	27605
0.2	28	480	925	2778	42	720	1387	4168
0.1	17	471	786	1081	26	706	1179	1622

If we now define $\vartheta = 1 - M_C^{-1}$, *where*

$$(49) \qquad M_C = \frac{1}{(1-\check{\lambda})^2}\left[1 - \check{\lambda} + \check{b} + \check{b}^2 + \bar{\zeta}_C\big(\check{b}(1 - \check{\lambda}) + \check{b}^2\big)\right]$$

and

$$(50) \qquad \bar{\zeta}_C = \frac{1 - \eta^2}{2\delta_C^4\eta}\left(\frac{b_C}{1 - \lambda_C}\right)^2,$$

then, for any $\rho \in (\vartheta, 1)$ *and all* $x \in \mathsf{X}$,

$$(51) \qquad \|P^n(x, \cdot) - \pi\|_V \le V(x)(1 + \gamma_C)\frac{\rho}{\rho - \vartheta}\rho^n, \qquad n \in \mathbb{Z}_+.$$

The value ζ_C could be improved with more knowledge of the model in a similar manner to the improvement through using $\zeta_\alpha(2)$ or $\zeta_\alpha(3)$ in place of $\zeta_\alpha(1)$ in Theorem 4.1, but this requires considerable extra work and also requires information not readily available in many models.

We now apply Theorem 5.1 to assess bounds on the rate of convergence of two Metropolis algorithms, in which π is again an $N(0, 1)$ distribution. These examples do indicate that, as in Example 2, the bounds are not in general yet of practical value: they do, however, give an indication of how to construct the drift inequalities that are used in all of the bounds such as those of Theorem 5.1 and [2] and [17].

EXAMPLE 3 (Estimating normal densities using normal candidates). If π is $N(0, 1)$, then we first choose as a symmetric candidate distribution the normal centered on the current value x with unit variance. Such a model satisfies all the conditions of Theorem 3.2 above, and hence is geometrically ergodic. We now calculate the constants in (48) in order to bound the rate of convergence. This illustrates strikingly the different emphases between practical calculation of the constants and the steps in the existence proof in Theorem 3.2, and the loss of effectiveness without existence of an atom.

We will apply (14) for the small set $C_\Delta = (-x_\Delta, x_\Delta)$ for some $x_\Delta > 0$ to be chosen. The size of the bound (49) may be best controlled by first specifying x_Δ, which determines δ_C and η, which give dominant terms in M_C. If one were interested only in establishing the existence of geometric ergodicity, ν could be chosen to be Lebesgue so that δ_C would be given by $\exp(-2x_\Delta^2)$. A considerable improvement is obtained by choosing the minorizing distribution with density

$$\nu(x) = \mathbb{1}_{C_\Delta}(x)e^{-x^2}/N_\Delta,$$

with N_Δ the normalizing constant $N_\Delta = \int_{-x_\Delta}^{x_\Delta} e^{-x^2}\,dx$. This gives

$$\delta_C = [2\pi]^{-1/2}N_\Delta\exp(-x_\Delta^2).$$

The infimum η occurs at $x = x_\Delta$, so that

$$\eta = 1/2 - \Phi(-2x_\Delta) - [2\pi]^{-1/2}N_\Delta\exp(-x_\Delta^2).$$

With the small set and its constants calculated, we will consider the test function $V(y) = e^{s|y|}$ for some s to be chosen also. Then, following (27),

$$K(x, s) := \int P(x, y) V(y) / V(x)\, dy$$

$$= \exp\left(\frac{s^2}{2}\right)(\Phi(-s) - \Phi(-x - s)$$

$$+ \exp(-2xs)(\Phi(-x + s) - \Phi(-2x + s)))$$

$$+ \frac{1}{\sqrt{2}}\left(\exp\left(\frac{x^2 - 6xs + s^2}{4}\right)\Phi((-3x + s)/\sqrt{2})\right.$$

$$\left. + \exp\left(\frac{(x - s)^2}{4}\right)\Phi((-x + s)/\sqrt{2})\right) + \frac{1}{2}$$

$$+ \Phi(-2x) - \frac{1}{\sqrt{2}}\exp\left(\frac{x^2}{4}\right)(\Phi(-3x/\sqrt{2}) + \Phi(-x/\sqrt{2})).$$

Now, instead of specifying s and calculating λ_C, it is more effective to specify λ_C and then determine s which then gives the remaining constants.

We find s by equating $\sup_{x \in C_\Delta^c} K(x, s)$ to λ_C, which is straightforward since the supremum occurs at x_Δ; since $PV(x) \leq \lambda_C V(x) + (K(x, s) - \lambda_C)V(x)$ for all $x \in C_\Delta$, we then find

$$b_C = \sup_{x \in C_\Delta} e^{s|x|}(K(x, s) - \lambda_C),$$

which has a supremum at 0 given by

$$b_C = \sqrt{2}\,\Phi(s/\sqrt{2})\exp\left(\frac{s^2}{4}\right) + 1 - 1/\sqrt{2} - \lambda_C.$$

We then have

$$b_\alpha^* = \left(\left(\sqrt{2}\exp\left(\frac{s^2}{4}\right)\right)/N_\Delta\right)(\Phi(\sqrt{2}(x_\Delta - s/2)) + \Phi(s/\sqrt{2}) - 1) - \lambda_C.$$

The terms b_c^*, λ_c, $\check{\lambda}$, $\check{b}$, $\bar{\zeta}$, M_C then follow as in (48), (49) and (50).

Using these computations, we found a minimum value of $M_C = 6.29 \times 10^7$, obtained for $x_\Delta = 1.20$, $\lambda_c = 0.99999$ and $s = 2.9 \times 10^{-5}$, that is, for a very flat V. This leads to a value of $\vartheta = 0.999999984$, and, in the notation above, $N_{0.01}(\vartheta) = 289{,}665{,}203$. The corresponding values for other terms in this calculation are as follows:

δ_C	b_C	b_α^*	$b_{\check{C}}^*$
0.15	2.6×10^{-5}	2.9×10^{-5}	3.4×10^{-5}

η	$\check{\lambda}$	$\check{b}$	$\bar{\zeta}$
0.34	$1 - 10^{-5}$	5.9×10^{-4}	1.6×10^4

The symmetric model just considered is based on a zero-mean random walk candidate. We conclude by considering a modified or adaptive chain which moves relatively quickly to the center of the space; this is similar to a model analyzed by Schervish and Carlin [18], which was shown to converge at rate $1/2$, and we might hope to find a faster theoretical rate of convergence here.

Again, take π as $N(0, 1)$, and now let us choose the candidate density by $Q(x, \cdot) = N(x/2, 1)$. It is straightforward to calculate that the acceptance region A_x is again $|y| \leq |x|$, and that the acceptance probabilities for transitions outside this region are given by

$$\alpha(x, y) = \exp\left(\frac{x^2 - y^2}{8}\right).$$

We now consider the same steps for calculating M_C, and use the two test functions

$$V_1(y) = 1 + s|y|,$$
$$V_2(y) = \exp(sy^2)$$

in order to see whether a steeper or slower choice of V gives a tighter bound.

Using V_1, a best value of $M_C = 2.4 \times 10^6$ was obtained at $x_\Delta = 1.55$, $\lambda_C = 0.98$. As expected, this was improved using V_2, for which the minimum value of M_C was 7.6×10^5 at $x_\Delta = 1.70$, $\lambda_C = 0.98$. These give, respectively, values of $\vartheta = 0.99999958$ and $N_{0.01}(\vartheta) = 11{,}052{,}406$ for V_1, and $\vartheta = 0.9999986$ and $N_{0.01}(\vartheta) = 3{,}499{,}927$ for V_2.

6. Value of bounds on rates. Clearly, at the moment the bounds developed in the last two sections are not useful for practical purposes, except in the stochastically monotone case. It is possible to construct variants of standard algorithms that are stochastically monotone in some specialized cases, but these are rather artificial [8]. The bounds found in Section 5 are also unfortunately relatively close to the best that can be expected using this method of bounding, as can be seen by consideration of the role of δ_C and η in the expressions concerned; more details are in [8].

By carrying out the computations above, we are able to indicate the type of function V which provides an indicator of the "moments" that will converge in the sense of (16). In the case of Example 2, we had an obvious choice of V and of λ, but this is not always the case when using (39): it is perhaps counterintuitive that one chooses a value of λ_C so close to unity in the first part of Example 3, but this ensures a very low value of b_C and the interplay between the values of b and λ in (39) give the "best" values above. These computations indicate the approach to using the bounds: all methods we know of require similar computation.

It is hard to assess how accurate bounds of this nature may be. We saw in Example 2 that they are likely to be inaccurate by several orders of magnitude, and numerical work in [8] shows that the convergence of specific moments may well be much faster. An alternative approach to convergence, giving nongeometric but apparently more effective bounds in practice (at

least on the total variation norm), has recently been developed in [15], where the authors give an application to the examples in Section 5, indicating that the total variation norm distance is likely to decrease more quickly than $10^{-3}/n$ in these cases.

Nonetheless, they do have the virtue that they are "honest" bounds. If we can improve the theoretical methods in ways that use the structure of individual models a little more, as was done in constructing the rather more accurate constants $\vartheta(2)$ and $\vartheta(3)$ in Theorem 4.1, then we can expect to give some insight into the amount of iteration that really is needed to achieve convergence in the Hastings–Metropolis algorithm.

REFERENCES

[1] AMIT, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–99.

[2] BAXENDALE, P. H. (1993). Uniform estimates for geometric ergodicity of recurrent Markov processes. Technical report, Dept. Mathematics, Univ. Southern California.

[3] BESAG, J. E. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 25–38.

[4] BESAG, J. E., GREEN, P. J., HIGDON, D. and MENGERSEN, K. L. (1995). Bayesian computation and stochastic systems (with discussion). *Statist. Sci.* **10** 3–66.

[5] CHAN, K. S. (1993). Asymptotic behavior of the Gibbs sampler. *J. Amer. Statist. Assoc.* **88** 320–326.

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

[7] LUND, R. B. and TWEEDIE, R. L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.* To appear.

[8] MENGERSEN, K. L. and TWEEDIE, R. L. (1993). Rates of convergence of the Hastings and Metropolis algorithms. Technical Report 93/30, Dept. Statistics, Colorado State Univ.

[9] METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1091.

[10] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, New York.

[11] MEYN, S. P. and TWEEDIE, R. L. (1994). Computable bounds for convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011.

[12] NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Univ. Press.

[13] ROBERTS, G. O. and POLSON, N. G. (1994). A note on the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **56** 377–384.

[14] ROBERTS, G. O., MENGERSEN, K. L., SCOTT, D. J. and TWEEDIE, R. L. (1995). Bounds on convergence rates in MCMC: a comparison of methods. Unpublished manuscript.

[15] ROBERTS, G. O. and ROSENTHAL, J. S. (1995). Shift-coupling and convergence rates of ergodic averages. Unpublished manuscript.

[16] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*. To appear.

[17] ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566.

[18] SCHERVISH, M. J. and CARLIN, B. P. (1992). On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics* **1** 111–127.

[19] SCOTT, D. J. and TWEEDIE, R. L. (1996). Explicit rates of convergence of stochastically ordered Markov chains. In *Proceedings of Athens Conference on Applied Probability and Time Series*. Springer, New York. To appear.

[20] SMITH, A. F. M. and GELFAND, A. E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *Amer. Statist.* **46** 84–88.

[21] SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 3–24.

[22] TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22** 1701–1762.

[23] TUOMINEN, P. and TWEEDIE, R. L. (1994). Subgeometric rates of convergence of *f*-ergodic Markov chains. *Adv. in Appl. Probab.* **26** 775–798.

SCHOOL OF MATHEMATICS
QUEENSLAND UNIVERSITY OF TECHNOLOGY
BRISBANE, QUEENSLAND 4000
AUSTRALIA

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523