On the Geometric Ergodicity of Metropolis-Hastings Algorithms

Yves F. Atchadé[1] *and François Perron*[2]

(March, 2003; revised August, 2005)

Abstract

Under a compactness assumption, we show that a ϕ-irreducible and aperiodic Metropolis-Hastings chain is geometrically ergodic if and only if its rejection probability is bounded away from unity. In the particular case of the Independence Metropolis-Hastings algorithm, we obtain that the whole spectrum of the induced operator is contained in (and in many cases equal to) the essential range of the rejection probability of the chain as conjectured by Liu (1996).

Key words: Geometric ergodicity, Markov chain operators, Metropolis-Hastings algorithm.

MSC Numbers: 65C05, 65C40, 60J27, 60J35

1 Introduction

The Metropolis-Hastings (MH) algorithm is a very flexible algorithm used to approximately sample from complicated distributions in high dimension spaces. If π is the probability distribution of interest, such an algorithm generates a Markov chain (X_n) which admits π as its stationary distribution. Geometric ergodicity characterizes a global stability property of the chain that is particularly useful from a statistical point of view. For example, if the Markov chain is geometric ergodicity, central limit theorems for empirical sums of functional of the chain are easier to obtain (see e.g. Jones (2004)). See Roberts and Rosenthal (1998) for more on the usefulness of geometric ergodicity in MCMC simulation.

There are various characterization of the concept of geometric ergodicity of Markov chains. For irreducible and aperiodic Markov chains, geometric ergodicity can be characterized by the existence of some "drift condition"; it is also known to be equivalent to an exponential decay in the tail of the distribution of return times of the chain to "small sets". A good account is given in Meyn and Tweedie (1993). These results have been used successfully to show that some specific MH algorithms converge geometrically fact (see e.g. Roberts and Rosenthal (2004) for a review). Our main objective in this note is to derive a general characterization of geometric ergodicity for MH algorithms directly in terms of the constituents of the algorithm. The "natural" target quantity is

[1]Department of Mathematics and Statistics, University of Ottawa, email: yatchade@uottawa.ca

[2]Department of Mathematics and Statistics, University of Montreal, email: perronf@dms.umontreal.ca

the rejection probability of the algorithm. It is known that if a MH chain is geometrically ergodic, then its rejection probability is necessarily bounded away from unity (Roberts and Tweedie (1996)). We shall prove that the condition is also sufficient if the linear operator generates by the absolutely continuous part of the transition kernel of the MH chain is compact. We conjecture that the result is still true even if the compactness assumption does not hold.

We then turn to the Independent Metropolis-Hastings algorithm (IMH). Liu (1996) prove that, when the state space of the IMH chain is finite or discrete (with an additional regularity condition on π), the spectrum of the IMH chain is the range of the rejection probability. He has conjectured that the result holds in general spaces. Below, in Theorem (2.2), we show that in general state spaces, the spectrum of the IMH chain is contained in (and in most practical cases equal to) the essential range of the rejection probability of the chain.

Our proofs mainly rely on operators theory in Hilbert spaces. We decompose K_0, the operator induced by the Metropolis-Hastings chain, as the sum of a multiplication operator M_r (multiplication by the rejection probability of the chain denoted r) and an integral operator U. We use a result (Weyl's criteria) from perturbation theory to relate the spectrum of K_0 to that of M_r.

2 Geometric Ergodicity of the Metropolis-Hastings Chain

Throughout we fix $(\mathcal{S}, \mathcal{F}, \pi)$ a probability space, where π is our target probability measure. Let $Q(x, \cdot)$ be a transition kernel on $\mathcal{S}$. We shall assume that for all x, $Q(x, \cdot)$ is absolutely continuous with respect to π and we write $Q(x, dy) = \omega(x, y)\pi(dy)$. Let P be the transition of the MH algorithm.

$$P(x, dy) = \alpha(x, y)\omega(x, y)\pi(dy) + r(x)\delta_x(dy), \tag{2.1}$$

where

$$r(x) = 1 - \int \alpha(x, y)\omega(x, y)\, \pi(dy), \tag{2.2}$$

and $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. A Markov chain with transition kernel P can be simulated as follows. Start with some arbitrary point $X_0 \in \mathcal{S}$. Given $X_i = x$, $i \geq 0$, generate Y from $Q(x, \cdot)$ and set $X_{i+1} = Y$ with probability $\alpha(x, Y)$, $X_{i+1} = x$ with probability $1 - \alpha(x, Y)$, where

$$\alpha(x, y) = \begin{cases} Min\left(1, \frac{\omega(y,x)}{\omega(x,y)}\right) & \text{if } \omega(x, y) \neq 0 \\ 1 & \text{if } \omega(x, y) = 0. \end{cases}$$

The proposal kernel Q can be chosen in various ways. When $Q(x, \cdot) = Q(\cdot)$ for all $x \in \mathcal{S}$, the algorithm obtained is the Independent Metropolis-Hastings sampler (IMH). If $Q(x, dy) = q(x, y)dy$

(with respect to some measure) and $q(x,y) = q(y,x)$ we obtain the Metropolis sampler. When $\mathcal{S} = \mathbb{R}^d$ and $Q(x,\cdot)$ has a density with respect to the Lebesgue measure with $Q(x,dy) = q(|y-x|)dy$, where $|x|$ denotes the Euclidean norm, we obtain the Random Walk Metropolis algorithm. More details can be find in Tierney (1994).

By construction, P admits π as its invariant distribution: $\pi P = \pi$ where

$$\pi P(A) := \int \pi(dx)P(x,A). \tag{2.3}$$

We say that a transition kernel P with invariant distribution π is ergodic if

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{var} \longrightarrow 0, \text{ for } \pi\text{-a.e. } x \in \mathcal{S}, \text{ as } n \to \infty, \tag{2.4}$$

where $\|\mu\|_{var} := \frac{1}{2}\sup_{|f|\leq 1}\left|\int \mu(dy)f(y)\right|$ is the total variation norm of a signed measure μ.

Geometric ergodicity essentially says that the convergence in (2.4) takes place at a geometric rate. More precisely, a transition kernel P with invariant distribution π is geometrically ergodic if there exist $\rho < 1$ and a function $V : \mathcal{S} \longrightarrow [1,\infty]$ finite π-a.e. such that:

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{var} \leq V(x)\rho^n, \text{ for } \pi\text{-a.e } x \in \mathcal{S}. \tag{2.5}$$

The transition kernel P of the MH algorithm induces a linear bounded operator K on $L^2(\pi)$ the Hilbert space of real-valued square integrable functions defined on $\mathcal{S}$ by:

$$Kf(x) := \int f(y)P(x,dy). \tag{2.6}$$

Instead of the operator K defined in (2.6), we shall mainly work with K_0, the restriction of K to $L_0^2(\pi) := \{f \in L^2(\pi) : \pi(f) := \int f\, d\pi = 0\}$. It has been shown (see e.g. Roberts and Rosenthal (1997) and Roberts and Tweedie (2001)) that for reversible Markov chains, geometric ergodicity is equivalent to:

$$\sup_{f\in L_0^2(\pi),\ \|f\|\leq 1} \|K_0 f\| \leq \rho, \tag{2.7}$$

for some $\rho < 1$, where, for $f \in L^2(\pi)$, $\|f\| := \{\int f^2(x)\pi(dx)\}^{1/2}$.

Define the spectrum of K_0 by $\sigma(K_0) := \{\lambda \in \mathbb{R} : K_0 - \lambda I \text{ is non invertible}\}$ where I is the identity operator of $L_0^2(\pi)$, and write $r(K_0) := \sup\{|\lambda| : \lambda \in \sigma(K_0)\}$ for the spectral radius of K_0. Then it is well known that (2.7) is equivalent to $r(K_0) = \|K_0\| < 1$, where $\|K_0\|$ is the norm of the operator K_0 defined by $\|K_0\| := \sup_{\|f\|\leq 1}\|K_0 f\|$. Whenever $\|K_0\| < 1$, we say that the chain has a spectral gap.

In the study of the spectrum of K_0, we make the distinction between $\sigma_d(K_0)$, the discrete spectrum of K_0 and $\sigma_{ess}(K_0) = \sigma(K_0) \setminus \sigma_d(K_0)$ the essential spectrum of K_0. the discrete spectrum

$\sigma_d(K_0)$ is defined as those λ in $\sigma(K_0)$ which are eigenvalues of K_0 and are isolated point of $\sigma(K_0)$ and such that $\dim\ker(K_0 - \lambda I)$, the dimension of the null space of $K_0 - \lambda I$, is finite. We shall also denote by

$$\text{ess-ran}(r) := \{\lambda \in \mathbb{R} : \pi\{x : |r(x) - \lambda| < \epsilon\} > 0 \text{ for all } \epsilon > 0\}$$

the essential range of r. If ess-inf(r) (respectively ess-sup(r)) is the essential (with respect to π) infimum (respectively essential supremum) of the function r, it is easily seen that ess-ran$(r) \subseteq$ [ess-inf(r), ess-sup(r)] and that both ess-inf(r) and ess-sup(r) belong to ess-ran(r).

The following result from Chan and Geyer (1994) will be useful later.

Proposition 2.1. *Suppose that P is an ergodic transition kernel on $(\mathcal{S}, \mathcal{F}, \pi)$ with invariant distribution π. Then K_0 as defined above has no eigenvalue with absolute value 1.*

In the case of the Metropolis-Hastings algorithm, the operator K_0 acts on $f \in L_0^2(\pi)$ as follows:

$$K_0 f(x) = \int f(y) P(x, dy) \tag{2.8}$$

$$= M_r f(x) + U f(x). \tag{2.9}$$

With $M_r f(x) = r(x) f(x)$ and $U f(x) = \int \alpha(x, y)\omega(x, y) f(y)\pi(dy)$.

In other words, the Metropolis-Hastings operator is a multiplication operator perturbed by an integral operator. We shall assume that this integral operator is compact.

Theorem 2.1. *Assume that U is compact. Suppose that the transition kernel P of the MH algorithm is ergodic. Then it is geometrically ergodic if and only if ess-sup$(r) < 1$. The essential supremum being taken with respect to π.*

Proof. The necessary part is proposition 5.1 of Roberts and Tweedie (1996). For the sufficient part, it is sufficient to show that $\|K_0\| < 1$. If $\|K_0\| \leq$ ess-sup(r), then the theorem is trivially true. Let us assume that $\|K_0\| >$ ess-sup(r). Then we claim that $\|K_0\|$ (or $-\|K\|_0$) is an eigenvalue for K_0. But for an ergodic chain, we know from proposition (2.1) that ± 1 cannot be eigenvalues for K_0. Thus $\|K_0\| < 1$.

Proof of the claim:

The compactness of the operator U implies (by Weyl's theorem (Berberian (1970)), that $\sigma_{ess}(K_0) = \sigma_{ess}(M_r)$). The spectrum of the multiplication operator M_r is $\sigma(M_r) =$ ess-ran(r). Thus

$$\begin{aligned} \sigma_{ess}(M_r) &\subseteq \text{ess-ran}(r), \\ &\subseteq [\text{ess-inf}(r), \text{ess-sup}(r)]. \end{aligned}$$

Since K_0 is self-adjoint, either $\|K_0\| \in \sigma(K_0)$ or $-\|K_0\| \in \sigma(K_0)$ (Halmos (1957) section 34, theorem 2) and the claim follows because $\|K_0\| > \text{ess-sup}(r)$. □

Remark 2.1. 1. In practice, it is very difficult to show that the operator U is compact. Actually, for most MH algorithms used in practice, U is not compact when the state space is unbounded. Nevertheless, we think that Theorem 2.1 is useful in the sense that it clarifies further what is known about the MH algorithm. Even if U is not compact, we conjecture that Theorem 2.1 remains true.

2. From a theoretical point of view, one easy way to obtain the compactness of U is to require that U be Hilbert-Schmidt that is to assume that

$$\int \alpha^2(x,y)\omega^2(x,y)\pi(dx)\pi(dy) < \infty.$$

It is easy to see that this later condition holds if:

$$\int Q(x,dy)Q(y,dx) < \infty. \tag{2.10}$$

3. Theorem 2.1 requires the ergodicity of the Markov chain (X_n). In practice, it is not very hard to construct ergodic MH algorithms. See for example Tierney (1994), Roberts and Tweedie (1996) for some easily verifiable conditions.

In the case of the IMH algorithm, (2.10) is clearly always satisfied. Therefore we have the following well known result on the geometric ergodicity of the IMH algorithm (Tierney (1994), Mengersen and Tweedie (1996)).

Corollary 2.1. *Let r be the probability of rejection of the IMH chain as given by Equation (2.2). Then ess-inf$(r) = 0$ and ess-sup$(r) = 1 -$ ess-inf(ω). Therefore the IMH algorithm has a spectral gap if and only if ess-inf$(\omega) > 0$.*

In fact, more can be said about the spectrum of the IMH chain. We shall prove the following:

Theorem 2.2. *For the IMH algorithm, $\sigma(K_0) \subseteq$ ess-ran(r). The equality holds if for all $\alpha \in$ ess-ran(r), $\pi\{y : r(y) = \alpha\} = 0$.*

Proof. For the IMH algorithm, condition (2.10) is clearly satisfied. Therefore the operator U in the decomposition (2.9) is compact and we have, by Weyl's perturbation theorem, that $\sigma_{ess}(K_0) = \sigma_{ess}(M_r) \subseteq \text{ess-ran}(r)$. Next, we show that for any eigenvalue λ of K_0, $\lambda \in \text{ess-ran}(r)$ and conclude that $\sigma(K_0) \subseteq \text{ess-ran}(r)$.

First note that for $f \in L_0^2(\pi)$,

$$\begin{aligned} Uf(x) &= \int \alpha(x,y)\omega(y)f(y)\pi(dy) \\ &= \int_{\{y:\,\omega(x)\ge\omega(y)\}} \omega(y)f(y)\pi(dy) + \int_{\{y:\,\omega(x)<\omega(y)\}} \omega(x)f(y)\pi(dy) \\ &= \int_{\{y:\,\omega(y)\le\omega(x)\}} \left(\omega(y)-\omega(x)\right) f(y)\pi(dy). \end{aligned} \tag{2.11}$$

Now, take $\lambda \notin$ ess-ran(r) and suppose that there is a none zero $f_0 \in L_0^2(\pi)$ such that $K_0 f_0(x) = \lambda f_0(x)$. We shall prove that this leads to a contradiction and the result will be proved.

From (2.11) and (2.9), λ being an eigenvalue of K_0 with eigenfunction f_0 is equivalent to:

$$\int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\pi(dy) = f_0(x). \tag{2.12}$$

Consider T the operator $Tf(x) = \int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f(y)\pi(dy)$.

Note $\underline{\omega} =$ ess-inf$(\omega(x))$ and $\kappa =$ ess-inf$\left(|r(x)-\lambda|\right)$. Since $\lambda \notin$ ess-ran(r), $\kappa > 0$. Because f_0 is $\pi-$integrable and is not identically null, we can find $u > \underline{\omega}$ sufficiently large such that f_0 is not null on $\{x \in \mathcal{S} :\, \underline{\omega} \le \omega(x) < u\}$. For any partition $I_n = (u_0 \le u_1 \le \ldots \le u_n)$ of $[\underline{\omega}, u]$, with $u_0 = \underline{\omega}$ and $u_n = u$, we note $D(u_i) := \{x \in \mathcal{S} :\, u_{i-1} \le \omega(x) < u_i\}$ and $L_i^2(\pi) := \{h \in L_0^2(\pi) :\, h(x) = 0 \text{ for } x \notin D(u_i)\}$, $i = 1, \ldots, n$. $L_i^2(\pi)$ is an Hilbert space as a closed subspace of $L_0^2(\pi)$. Let $\chi_{D(u_i)}$ be the indicator function of the set $D(u_i)$ and let M_{D_i} be the multiplication operator defined by $\chi_{D(u_i)}$. Note that $M_{D_i}M_{D_i} = M_{D_i}$. For $h \in L_0^2(\pi)$, we write h_{D_i} for $h\chi_{D(u_i)} = M_{D_i}h$. Note that if T is an operator on $L_0^2(\pi)$, then $M_{D_i}TM_{D_i}$ is an operator on $L_i^2(\pi)$. With these notations, by applying M_{D_1} on both side of (2.12) we obtain that $M_{D_1}Tf_0 = M_{D_1}f_0$. But:

$$\begin{aligned} M_{D_1}Tf_0(x) &= \int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\chi_{D_1}(x)\pi(dy) \\ &= \int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} M_{D_1}f_0(y)\chi_{D_1}(x)\pi(dy) \\ &= M_{D_1}TM_{D_1}f_{0,D_1}. \end{aligned}$$

Therefore, (2.12) implies that $M_{D_1}TM_{D_1}f_{0,D_1} = f_{0,D_1}$. In the same way, for $2 \le i \le n$, (2.12) implies that $M_{D_i}Tf_0 = M_{D_i}f_0$, and as before, we have:

$$\begin{aligned} M_{D_i}Tf_0(x) &= \int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\chi_{D_i}(x)\pi(dy) \\ &= \sum_{k=1}^{i-1}\int_{\{y:\,y\in D(u_k)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\chi_{D_i}(x)\pi(dy) \\ &\quad + \int_{\{y:\,u_{i-1}<\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\chi_{D_i}(x)\pi(dy) \\ &= M_{D_i}h_i(x) + \int_{\{y:\,\omega(y)\le\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} M_{D_i}f_0(y)\chi_{D_i}(x)\pi(dy) \\ &= M_{D_i}h_i(x) + M_{D_i}TM_{D_i}f_{0,D_i}, \end{aligned}$$

where $h_i(x) = \sum_{k=1}^{i-1}\int_{D(u_k)} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\pi(dy)$.

This development shows that (2.12) implies:

$$\begin{cases} M_{D_1}TM_{D_1}f_{0,D_1} = f_{0,D_1} \\ M_{D_2}TM_{D_2}f_{0,D_2} = f_{0,D_2} - M_{D_2}h_2 \\ \quad\vdots \\ M_{D_n}TM_{D_n}f_{0,D_n} = f_{0,D_n} - M_{D_n}h_n. \end{cases} \tag{2.13}$$

Now we shall prove that (2.13) implies that $r\left(M_{D_i}TM_{D_i}\right) \ge 1$ for at least one $i \in \{1,\ldots,n\}$, where $r\left(M_{D_i}TM_{D_i}\right)$ denotes the spectral radius of $M_{D_i}TM_{D_i}$. To see why, assume that $r\left(M_{D_1}TM_{D_1}\right) < 1$. Therefore $M_{D_1}TM_{D_1}f_{0,D_1} = f_{0,D_1}$ implies that $f_{0,D_1} \equiv 0$ which in turn implies that $h_2(x) \equiv 0$. Note that from the second equation of (2.13), $h_2 \equiv 0$ implies that $M_{D_2}TM_{D_2}f_{0,D_2} = f_{0,D_2}$. If in addition, $r\left(M_{D_2}TM_{D_2}\right) < 1$, then since we have $M_{D_2}TM_{D_2}f_{0,D_2} = f_{0,D_2}$, we can assert that $f_{0,D_2} \equiv 0$ which together with $f_{0,D1} \equiv 0$ implies that $h_3 \equiv 0$. Continuing this way, we can see that if $r\left(M_{D_i}TM_{D_i}\right) < 1$ for all $i = 1,\ldots,n$ then $f_{0,D_1} = \ldots = f_{0,D_n} \equiv 0$ which contradicts the fact that f_0 is not the null function on $\{x \in \mathcal{S} : \underline{\omega} \le \omega(x) < u\}$ as chosen.

Until now, what we have proved is that if $\lambda \notin$ ess-ran(r) is an eigenvalue of K_0 with eigenfunction f_0 then for any $u > \underline{\omega}$ taken sufficiently large such that f_0 is not identically zero on $\{x \in \mathcal{S} : \underline{\omega} \le \omega(x) < u\}$ and any partition $I_n = (u_0 \le u_1 \le \ldots \le u_n)$ of $[\underline{\omega}, u]$, with $u_0 = \underline{\omega}$ and $u_n = u$, there is at least one $i \in \{1,\ldots,n\}$ such that $r\left(M_{D_i}TM_{D_i}\right) \ge 1$. To show the contradiction promised at the begining, we shall simply show that by taking a partition with a sufficiently small increment, we can make all the $r\left(M_{D_i}TM_{D_i}\right) < 1$.

For $g \in L_i^2(\pi)$ with $\|g\| = 1$, and using the Cauchy-Schwarz inequality, we can write:

$$\begin{aligned}
\|M_{D_i} T M_{D_i} g\|^2 &= \int_{D(u_i)} \left\{ \int_{\{y:\, \omega(y) \leq \omega(x)\}} \frac{\omega(x) - \omega(y)}{r(x) - \lambda} g_{D_i}(y) \pi(dy) \right\}^2 \pi(dx) \\
&\leq \left(\frac{u_i - u_{i-1}}{\kappa} \right)^2 \int_{D(u_i)} g^2(y) \pi(dy) \\
&\leq \left(\frac{u_i - u_{i-1}}{\kappa} \right)^2,
\end{aligned}$$

and therefore $\|M_{D_i} T M_{D_i}\| \leq (u_i - u_{i-1})/\kappa$. From this, by taking a partition $I_n = (u_0 \leq u_1 \leq \ldots \leq u_n)$ with $\max_{1 \leq i \leq n}(u_i - u_{i-1}) < \kappa$, we have for $i = 1, \ldots, n$:

$$\begin{aligned}
r\left(M_{D_i} T M_{D_i}\right) &= \lim_{n \to \infty} \|(M_{D_i} T M_{D_i})^n\|^{\frac{1}{n}} \\
&\leq \|M_{D_i} T M_{D_i}\| \\
&\leq \frac{u_i - u_{i-1}}{\kappa} \\
&< 1.
\end{aligned}$$

Next, take $\lambda \in$ ess-ran(r) such that $\pi\{y : r(y) = \lambda\} = 0$. Then $\lambda \in \sigma_{ess}(M_r) = \sigma_{ess}(K_0) \subseteq \sigma(K_0)$. This shows that if for all $\lambda \in$ ess-ran(r), $\pi\{y : r(y) = \lambda\} = 0$ then $\sigma(K_0) =$ ess-ran(r). $\square$

Remark 2.2. 1. A typical case for theorem (2.2) will be that $\mathcal{S} = \mathbb{R}^d$ and π and Q are absolutely continuous with respect to the Lebesgue measure: $\pi(dx) = \pi(x)dx$ and $Q(dx) = q(x)dx$. If in addition, $\pi(x)$ and $Q(x)$ are continuously differentiable, then $\{x : r(x) = \lambda\}$ is a sub-manifold of dimension $d-1$ of $\mathbb{R}^d$ thus $\pi\left(\{x : \pi(x) = \lambda\}\right) = 0$ and we have $\sigma(K_0) = [0, 1 - \text{ess-inf}(\omega)]$.

2. One well known consequence of theorem (2.2) is that it is not possible to take advantage of the correlation of the Markov chain to decrease the asymptotic variance of the Monte Carlo estimate. Suppose that (X_n) is a geometrically ergodic Markov chain with stationary distribution π, generated using the IMH algorithm. For any $f \in L_0^2(\pi)$, writting e_f to denote the spectral measure of K_0 on f and $\rho := 1 - \text{ess-inf}(\omega)$, we have: $\lim_{n \to \infty} \frac{1}{n} E\left(\sum_{k=0}^{n-1} f(X_k)\right)^2 = \int_0^\rho \frac{1+\lambda}{1-\lambda} e_f(d\lambda) \geq \|f\|^2$.

Acknowledgements:

We are grateful to the Editor and an anonymous referee for their very helpful comments and for improving the presentation of this paper.

References

Berberian, S. K., 1970. The weyl spectrum of an operator. Indiana University Mathematics Journal 20 (6), 529–544.

Chan, K., Geyer, G., 1994. Discussion paper. Ann. Statist. 22 (4), 1747–1758.

Halmos, P. R., 1957. Introduction to Hibert space. Chelsea Publishing Company, New York.

Jones, G., 2004. On the markov chain central limit theorem. Probability surveys 1, 299–320.

Liu, J. S., 1996. Metropolized independent sampling with comparaisons to rejection sampling and importance sampling. Statistics and Computing 6, 113–119.

Mengersen, K. L., Tweedie, R. L., 1996. Rates of convergence of the Hastings and Metropolis algorithms. Ann. Statist. 24 (1), 101–121.

Meyn, S. P., Tweedie, R. L., 1993. Markov chains and stochastic stability. Springer-Verlag London Ltd., London.

Roberts, G. O., Rosenthal, J. S., 1997. Geometric ergodicity and hybrid Markov chains. Electron. Comm. Probab. 2, no. 2, 13–25 (electronic).

Roberts, G. O., Rosenthal, J. S., 1998. Markov-chain Monte Carlo: some practical implications of theoretical results. Canad. J. Statist. 26 (1), 5–31, with discussion by Hemant Ishwaran and Neal Madras and a rejoinder by the authors.

Roberts, G. O., Rosenthal, J. S., 2004. General state space markov chain and mcmc algorithms. Technical Report, Dept of Statistics, University of Toronto .

Roberts, G. O., Tweedie, R. L., 1996. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika 83 (1), 95–110.

Roberts, G. O., Tweedie, R. L., 2001. Geometric l^2 and l^1 convergence are equivalent for reversible markov chains. J. Appl. Prob. 38 A, 37–41.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Statist. 22 (4), 1701–1762, with discussion and a rejoinder by the author.