Spectrum of the Metropolis-Hastings chain with an application to geometric ergodicity

Yves F. Atchadé*[†] François Perron*[‡]

CRM-2862

September 2002

*Dept. of Mathematics and Statistics, University of Montreal.

[†]Research supported in part by ISM and NSERC scholarships.

[‡]Research supported in part by NSERC.

Abstract

We study the spectrum of the operator induced by the Metropolis-Hastings algorithm. For a general Harris ergodic Metropolis-Hastings algorithm, we show (under an additional condition) that if the probability of rejection is bounded away from unity, then the chain has a spectral gap. Using this result, we propose an adaptation of the Symetric Random Walk Metropolis-Hastings algorithm which is shown to be geometrically ergodic for any bounded continuous density with tail lighter than the proposal distribution's tail. In the particular case of the Independent Metropolis-Hastings algorithm, we explicitly derive the spectrum of the induced operator therefore giving a solution to Liu's conjecture (Liu (1996)).

AMS Classification: 65C40, 60J22, 60J10.

Keywords: Metropolis-Hastings algorithm, Markov chains operator, Spectrum of an operator, Geometric ergodicity.

1 Introduction.

The Metropolis-Hastings (MH) algorithm initiated by Metropolis et al. (1953) 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. If the chain is ergodic, then for n sufficiently large, X_n is taken as approximately distributed as π. From a statistical point of view, geometric ergodicity is a desirable property for such algorithms. Specifically, under geometric ergodicity, the Central Limit Theorem for empirical sum of functions of X_n is (theoretically) easy to check as in Roberts & Rosenthal (1997). Geometric ergodicity can also be used to design convergence diagnostic techniques, a big issue in practical use of MCMC methods (see Brooks & Roberts (1997)). There is an interesting discussion about the usefulness of geometric ergodicity of MCMC algorithms in Roberts & Rosenthal (1998).

In this paper, our aim is to show that under a compacity condition, an Harris ergodic Metropolis-Hastings chain is geometrically ergodic if the rejection probability of the chain is bounded away from unity. As we shall see, this result can be useful to design new algorithms which are geometrically ergodic. Actually, it is a partial converse of Roberts & Tweedie (1996) who proved that if a Metropolis-Hastings chain is geometrically ergodic, then the rejection probability of the chain is bounded away from unity.

Our techniques of proof differ from those in Roberts & Tweedie (1996) as we mainly use results from Hilbert spaces operators theory. 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 or the chain denoted r) and an integral operator U. Under a compacity condition on U, the Weyl's theorem (Berberian (1970)) states that the "continuous" part of the spectrum of K_0 is the same with that of M_r which is ess-ran(r), the essential range of r. Then the reversibility and the Harris ergodicity of the chain is used to assure that the "discret" part of the spectrum of K_0 remains bounded away from unity.

In the particular case of the Independent Metropolis-Hastings algorithm (IMH algorithm), the compacity condition is always satisfied, but in this case more can be said. In theorem (3.1), we show that the spectrum of the IMH is exactly the essential range of the rejection probability of the chain. Liu (1996) proved this result when the state space of the chain is finite or discrete (with an additional regularity condition on π). He also conjectured that the result holds in a more general setting.

In general, it may be difficult to choose a good proposal to use with the MH algorithm. This is certainly one reason why the Symetric Random Walk Metropolis-Hastings (SRWMH) is still widely used. But as shown by Jarner & Tweedie (2001), exponential or lighter tail is necessary for the SRWMH algorithm to be geometrically ergodic. For heavy tailed distribution, polynomial rate of convergence is possible. In this direction, Jarner & Roberts (2002) have shown that for a symetric target density with polynomial tail, and for any proposal density with tail that also recedes polynomially, the SRWMH algorithm has a polynomial rate of convergence.

In this paper, as another solution, we propose to modify the SRWMH algorithm by restricting the random walk behaviour on a compact set. This restriction yields a chain that is geometrically ergodic whenever the target density have a lighter tail than the proposal's tail. The idea is as follow. We fix a convex compact set Δ. Ideally, Δ contains all the modes of π. At any $x \in \Delta$, the proposal move is made from a distribution centered at x as in the SRWMH algorithm. But when the chain reaches $x \notin \Delta$, the proposal distribution is centered at $p(x)$ the orthogonal projection of x on Δ. If all the proposal distributions recede, as say, $\|x\|^{-(d+r)}$, we show in theorem (4.1) that this algorithm is geometrically ergodic for any bounded continuous density with tail decaying as

$\|x\|^{-(d+r)}$ or faster. The algorithm explores Δ essentially in the same way as the SRWMH algorithm, but its excursion length outside Δ will typically be shorter resulting in a more stable Monte Carlo estimation procedures.

In section 2, we state and prove theorem 2.1 on the existence of a spectral gap for the MH algorithm. In section 3, we derive the exact spectrum of the IMH chain. Our geometrically ergodic chain derived from the SRWMH chain is studied in section 4. A simulation study is presented to illustrate.

2 Spectral Gap for the Metropolis-Hastings Chain.

Throughout the paper, $\mathcal{S}$ is a subset of $\mathbb{R}^d$ equiped with its Borel σ-algebra. Let π be a probability measure and $Q(x,\cdot)$ a transition kernel on $\mathcal{S}$. We assume that for all x, $Q(x,\cdot)$ is mutually absolutely continuous with respect to π and write $\omega(x,y) = Q(x,dy)/\pi(dy)$.

Before going further, we recall the Metropolis-Hastings algorithm.

Algorithm 2.1. *The Metropolis-Hastings Algorithm:*
At the ith iteration, $X_i = x$.
Generate $Y_{i+1} = y$ *from* $Q(x,\cdot)$.
set

$$X_{i+1} = \begin{cases} y & \text{with probability } \alpha(x,y) \\ x & \text{with probability } 1-\alpha(x,y) \end{cases}$$

With

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

When $Q(x,\cdot) = Q(\cdot)$ for all $x \in \mathcal{S}$, we obtain the Independent Metropolis-Hastings (IHM) algorithm. And when $Q(x,\cdot)$ has a density $q(x,y)$ with respect to the Lebesgue measure with $q(x,y) = q(0,y-x)$, we obtain the Random Walk Metropolis-Hastings (RWMH) algorithm. If in addition, q is symetric ($q(0,x)$ depends on x only through $\|x\|$) we have the Symetric Random Walk Metropolis-Hastings (SRWMH). The algorithm (2.1) generates a Markov chain (X_n) with transition kernel

$$P(x,dy) = \alpha(x,y)\omega(x,y)\pi(dy) + r(x)\delta_x(dy), \tag{1}$$

where

$$r(x) = 1 - \int \alpha(x,y)\omega(x,y)\,\pi(dy) \tag{2}$$

is the probability of rejection of the chain and $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. It is well known that P is reversible with respect to π, that is:

$$\pi(dx)P(x,dy) = \pi(dy)P(y,dx), \tag{3}$$

as measures on $\mathcal{S}\times\mathcal{S}$. This implies that P admits π as an invariant distribution, $\pi P = \pi$ where

$$\pi P(A) := \int \pi(dx)P(x,A). \tag{4}$$

Below, we first review some basic concepts from Markov chain theory. For more details, we refer to Meyn & Tweedie (1993). We say that a Markov chain (X_n) with transition kernel P is $\phi-$irreducible if there exists a probability measure ϕ such that

$$\phi(A) > 0 \text{ implies that } \Pr\left(X_n \in A \text{ for some } n | X_0 = x\right) > 0 \text{ (for all } x \in \mathcal{S}). \tag{5}$$

A $\phi-$irreducible chain (X_n) is Harris recurrent if

$$\phi(A) > 0 \text{ implies that } \Pr(X_n \in A\, i.o. | X_0 = x) = 1 \text{ (for all } x \in A). \tag{6}$$

We say that a Markov chain is aperiodic if there does not exist a partition $\mathcal{S} = \mathcal{S}_1 \cup \ldots \cup \mathcal{S}_d$ with $d \geq 2$ such that $\Pr(x, \mathcal{S}_{i+1}) = 1$ for all $x \in \mathcal{S}_i$, $i = 1, \ldots, d-1$ and $\Pr(x, \mathcal{S}_1) = 1$ for all $x \in \mathcal{S}_d$.

And a Markov chain (X_n) with transition P is ergodic if

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{var} \longrightarrow_{n \longrightarrow \infty} 0$$

for $\pi-$a.e $x \in \mathcal{S}$, where $\|\mu\|_{var} = \frac{1}{2} \sup_{|f| \leq 1} |\int \mu(dy) f(y)|$ is the total variation norm of a signed measure μ. An Harris recurrent and ergodic chain will be called Harris ergodic. It is well known that if a ϕ-irreducible Markov chain admits an invariant distribution, and is aperiodic then it is ergodic.

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

$$Kf(x) = \int f(y) P(x, dy). \tag{7}$$

Instead of the operator K defined in (7), we shall mainly work with K_0, the restriction of K to $L_0^2(\pi) = \{f \in L^2(\pi) : \pi(f) := \int f \, d\pi = 0\}$. We say that the chain is geometrically ergodic if there exist $\rho < 1$ and a function V with $V(f) < \infty$ for all $f \in L_0^2(\pi)$ such that

$$\|K_0^n f\|_{L^2(\pi)} \leq V(f) \rho^n, \tag{8}$$

for all function $f \in L_0^2(\pi)$.
What we called geometric ergodicity is called L^2-geometric ergodicity by Roberts & Rosenthal (1997). And as in the main stream Markov chain theory, they use the term geometric ergodicity for convergence involving the total variation norm (L^1 norm). We do not make this distinction here as Roberts & Tweedie (2000) have shown that for reversible chains, L^1 and L^2 geometric ergodicity are equivalent concepts.

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 equation (8) 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 Roberts & Rosenthal (1997), P acts on the space of signed measure $\{\mu \ll \pi : \int \left(\frac{d\mu}{d\pi}\right)^2 d\pi < \infty\}$ by transforming μ into μP as defined in equation (4). For a π-reversible chain, it is well known that the two operators are isomorphically equivalent.

In the study of the spectrum of K_0, we make the distinction between $\sigma_d(K_0)$, the discret spectrum of K_0 and $\sigma_{ess}(K_0) = \sigma(K_0) \setminus \sigma_d(K_0)$ the essential spectrum of K_0. the discret 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 $\dim\ker(K_0 - \lambda) < \infty$. We shall also denote by ess-ran$(r) = \{\lambda \in \mathbb{R} : \pi\{x : |r(x) - \lambda| < \epsilon\} > 0$ for all $\epsilon >$ 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 & Geyer (1994) will be useful later.

Proposition 2.1. *If the chain (X_n) generated by the Metropolis-Hastings algorithm is Harris ergodic then K_0 has no eigenvalue with modulus 1.*

The operator K_0 acts on $f \in L^2_0(\pi)$ as:

$$K_0 f(x) = \int f(y) P(x, dy) \quad (9)$$
$$= M_r f(x) + U f(x). \quad (10)$$

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. The simplest way to study the spectrum of such operator is to make U a compact operator. This will be done through the following condition:

Compacity Condition 2.1. *The operator U is a compact $L^2_0(\pi)$ operator.*

The following lemma gives a sufficient condition on the proposal transition kernel Q for this compacity condition to hold.

Lemma 2.1. *If*

$$\int Q(x, dy) Q(y, dx) < \infty, \quad (11)$$

then U is an Hilbert-Schimdt operator, thus is compact.

Proof. It is sufficient to check that if (11) holds then

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

It follows from $\alpha(x, y)\omega(x, y) = Min\,(\omega(x, y), \omega(y, x))$, that $\alpha^2(x, y)\omega^2(x, y) \leq \omega(x, y)\omega(y, x)$. Therefore:

$$\alpha^2(x, y)\omega^2(x, y)\pi(dx)\pi(dy) \leq Q(x, dy)Q(y, dx).$$

□

The next theorem is our main result. We have used the Weyl's perturbation theorem (Berberian (1970)) together with some basic Hilbert space operator theory to show that K_0 has a spectral gap when ess-sup$(r) < 1$.

Theorem 2.1. *Suppose that the compacity condition on U holds. Suppose also that the Markov chain generated by the Metropolis-Hastings algorithm (algorithm 2.1) with proposal kernel Q is Harris ergodic. Then it has a spectral gap if and only if ess-sup$(r) < 1$. The essential supremum being taken with respect to π.*

Proof. Since U is a compact operator, by the Weyl's theorem (Berberian (1970), $\sigma_{ess}(K_0) = \sigma_{ess}(M_r)$. The spectrum of the multiplication operator is well known (see Conway (1985) example 2.6 page 271). $\sigma(M_r) =$ ess-ran(r) and λ is an eigenvalue for M_r if and only if $\pi\{y : r(y) = \lambda\} > 0$ and the indicator function of the set $\{y : r(y) = \lambda\}$ is an associated eigenfunction. Thus

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

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 $\sigma(K_0)$ is bounbed by $\|K_0\|$. Suppose that ess-sup$(r) < 1$. Then if ess-sup$(r) < \|K_0\|$, $\|K_0\| \in \sigma_d(K_0)$ (or $-\|K_0\| \in \sigma_d(K_0)$). But for an Harris ergodic chain, we know from proposition (2.1) that ± 1 cannot be eigenvalues for K_0. Thus $\|K_0\| < 1$.

The necessary part is proposition 5.1 of Roberts & Tweedie (1996). □

As, we mentioned above, the necessary part of theorem (2.1) holds even without the compacity condition (2.1) (Roberts & Tweedie (1996)).

In practice, it is not very hard to construct an Harris ergodic MH algorithm. The following proposition is adapted from Tierney (1994) and Roberts & Tweedie (1996) and gives simple conditions under which the MH kernel is Harris ergodic. Weaker conditions are possible.

Proposition 2.2. *1. Suppose that:*

$$\omega(x, y) > 0 \text{ for all } x, y \in \mathcal{S}. \tag{12}$$

Then the MH kernel is Harris recurrent.

2. In addition to (12), suppose that there exist $\varepsilon > 0$ and a Borel set $A \subseteq \mathcal{S}$, with $\pi(A) > 0$ such that:

$$\omega(x, y) > \varepsilon \text{ for all } x, y \in A. \tag{13}$$

Then the MH kernel is Harris ergodic.

Condition (12) is not a restricted requirement and will be satisfied in many situations. If $\omega(x, y)$ is continuous and satisfy (12), then (13) also will be satisfied and we can take A to be any non empty compact subset of $\mathcal{S}$.

3 Spectrum of the IMH chain.

In the case of the IMH, (11) of lemma (2.1) is always satisfy. Therefore we have the following well known result on the geometric ergodicity of the IMH algorithm (Tierney (1994), Liu (1996), Mengersen & Tweedie (1996)).

Corollary 3.1. *Let r be the probability of rejection of the IMH chain as given by equation (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(ω) > 0.*

Note that by proposition 2.2, ess-inf(ω) > 0 implies that the IMH is Harris ergodic. In fact, more can be said about the spectrum of the IMH chain. We shall prove the following:

Theorem 3.1. *For the IMH algorithm, $\sigma(K_0) \subseteq$ ess-ran(r). And if for all $\alpha \in$ ess-ran(r), $\pi\{y : r(y) = \alpha\} = 0$ then $\sigma(K_0) =$ ess-ran(r).*

Proof. We know from theorem (2.1) that $\sigma_{ess}(K_0) \subseteq$ ess-ran(r) with equality if for all $\alpha \in$ ess-ran(r), $\pi\{y : r$ $\alpha\} = 0$. It remains to show that K_0 has no eigenvalue outside ess-ran(r).

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

$$\begin{aligned} \int \alpha(x, y)\omega(y)f(y)\pi(dy) &= \int_{\{y:\,\omega(x)\geq\omega(y)\}} \omega(y)f(y)\pi(dy) \\ &\quad + \int_{\{y:\,\omega(x)<\omega(y)\}} \omega(x)f(y)\pi(dy) \\ &= -\int_{\{y:\,\omega(y)\leq\omega(x)\}} (\omega(x) - \omega(y))\, f(y)\pi(dy). \end{aligned} \tag{14}$$

Now, take $\lambda \notin$ ess-ran(r) and suppose that there is a none zero $f_0 \in L_0^2(\pi)$ such that $K_0 f_0(x) = \lambda f_0(x)$. Using equation (14) in equation (10), this is equivalent to:

$$\int_{\{y:\,\omega(y)\leq\omega(x)\}} \frac{\omega(x) - \omega(y)}{r(x) - \lambda} f_0(y)\pi(dy) = f_0(x). \tag{15}$$

Consider T the operator $Tf(x) = \int_{\{y:\,\omega(y)\leq\omega(x)\}} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f(y)\pi(dy)$. Then equation (15) says that 1 is an eigenvalue for T with eigenfunction f_0. We shall now show that this is not possible and the theorem will be proved.

Note $\underline{\omega} = \text{ess-inf}(\omega(x))$ and $\kappa = \text{ess-inf}\,(|r(x)-\lambda|) > 0$. Since $f_0 \neq 0$, we can find $u > \underline{\omega}$ sufficiently large such that $f_0 \neq 0$ on $\{x \in \mathcal{S} : \underline{\omega} \leq \omega(x) < u\}$. For any partition $I_n = (u_0 \leq u_1 \leq \ldots \leq u_n)$ of $[\underline{\omega}, u]$, with $u_0 = \underline{\omega}$ and $u_n = u$, we note $D(u_i) = \{x \in \mathcal{S} : u_{i-1} \leq \omega(x) < u_i\}$ 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 M_{D_i} be the multiplication operator by $\chi_{D(u_i)}$, the indicator function of the set $D(u_i)$. For $h \in L_0^2(\pi)$, we write h_{D_i} for $h\chi_{D(u_i)} = M_{D_i}h$. With these notations, it is easy to see that equation (15) implies:

$$\begin{cases} M_{D_1}TM_{D_1}f_{0,D_1} &= f_{0,D_1} \\ M_{D_2}TM_{D_2}f_{0,D_2} &= f_{0,D_2} - M_{D_2}h_2 \\ \quad\vdots & \\ M_{D_n}TM_{D_n}f_{0,D_n} &= f_{0,D_n} - M_{D_n}h_n \end{cases} \tag{16}$$

where $h_j(x) = \sum_{k=1}^{j-1} \int_{D(u_k)} \frac{\omega(x)-\omega(y)}{r(x)-\lambda} f_0(y)\pi(dy)$, for $j = 2, \ldots, n$. Equation (16) implies that $r\left(M_{D_i}TM_{D_i}\right) \geq 1$ for at least one i in $1, \ldots, n$.

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

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

Therefore $\|M_{D_i}TM_{D_i}\| \leq (u_i - u_{i-1})/\kappa$. >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 can write for $i = 1, \ldots, n$:

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

Thus the theorem is proved. □

4 Geometric ergodicity of a restricted Symetric Random Walk Metropolis-Hastings Algorithm.

4.1 Restricting the Symetric Random Walk Metropolis-Hastings.

In practice, the Symetric Random Walk Metropolis-Hastings (SRWMH) algorithm remains one of the most used MH algorithm. But for many common distribution, the SRWMH will fail to be geometrically ergodic. In fact, as we mentioned in the introduction, exponential or lighter tail is necessary for the SRWMH to be geometrically ergodic. Here, we propose a MH algorithm that behaves as the SRWMH on a compact set. We show in theorem 4.1 that the proposed algorithm is geometrically ergodic for any bounded positive and continuous density which tail decays at least as faster as the proposal density tail.

In this section, we set $\mathcal{S} = \mathbb{R}^d$, and we suppose that π has a density with respect to the Lebesgue measure. We shall also denote this density by π. Let Δ be a non empty convex compact subset of $\mathcal{S}$. We note by $p(x)$ the orthogonal projection of x on Δ. Consider the following proposal kernel, which is a student distribution kernel:

$$Q(x,y) = K \frac{c}{\left(1 + \frac{c^{2/d}}{\nu} \|p(x) - y\|^2\right)^{\frac{d+\nu}{2}}}, \tag{17}$$

where K is the normalizing constant, $\nu > 0$ the degree of freedom and $c > 0$ the scale parameter. On Δ, the MH algorithm 2.1 with proposal given by (17) behaves as a SRWMH algorithm with a student kernel. For $x \notin \Delta$, the proposal is taken from $Q(p(x), y)$ where $p(x)$ is the projection of x on Δ. Typically, if x_0 is any position parameter of π, one can take $\Delta = B(x_0, R)$, the closed ball centered at x_0 with radius $R > 0$ taking sufficiently large for Δ to include all the essential features of π. In this case, $p(x) = x_0 + (\min(1, R/\|x - x_0\|))(x - x_0)$ is easy to compute.

Clearly, the transition density becomes:

$$P(x,y)dy = \alpha(x,y)Q(p(x),y)dy + r(x)\delta_x(dy),$$

where

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)Q(p(y),x)}{\pi(x)Q(p(x),y)}\right),$$

and

$$r(x) = 1 - \int \alpha(x,y)Q(p(x),y)dy. \tag{18}$$

Theorem 4.1. *Suppose that:*

i. *the density π is bounded, everywhere positive and continuous.*

ii. $\limsup_{\|x\|\to\infty} \pi(x) \|x\|^{d+\nu} < \infty$.

Then the MH algorithm 2.1 with proposal kernel given by (17) is geometrically ergodic.

Proof. It follows from (i), (17) and proposition (2.2) that the MH algorithm with proposal given by (17) is Harris ergodic.
Showing the compacity condition 2.1 amounts to show that $\int dx \int Q(p(x),y)Q(p(y),x)dy < \infty$. Consider the function:

$$Q_0(x) = \frac{cK}{\left(1 + \frac{c^{2/d}}{\nu} \|p(x) - x\|^2\right)^{\frac{d+\nu}{2}}}.$$

Clearly, for all x and y in $\mathcal{S}$, we always have $\|p(x) - y\| \geq \|p(y) - y\|$. This implies that $Q(x,y) \leq Q_0(y)$. Therefore $\int dx \int Q(p(x),y)Q(p(y),x)dy \leq \left(\int Q_0(y)dy\right)^2 < \infty$. Then il follows from theorem 2.1 that the proposed algorithm is geometrically ergodic if and only if ess-sup $r(x) < 1$, where r is the rejection probability given by (18).

>From (i), (ii) and the compactness of Δ, it is not hard to see that:

$$\inf_{z\in\Delta, x\in\mathcal{S}} \frac{Q(z,x)}{\pi(x)} = \eta > 0,$$

and

$$\inf_{x\in\mathcal{S}} \frac{Q_0(x)}{\pi(x)} = \eta_0 > 0.$$

Therefore we have:

$$\begin{aligned}
\alpha(x,y) &= \min\left(1, \frac{\pi(y)Q(p(y),x)}{\pi(x)Q(p(x),y)}\right) \\
&\geq \min\left(1, \frac{Q(p(y),x)/\pi(x)}{Q_0(y)/\pi(y)}\right) \\
&\geq \frac{\eta}{\eta_0},
\end{aligned}$$

which yields $r(x) = 1 - \int \alpha(x,y)Q(p(x),y)dy \leq 1 - \eta/\eta_0$. Therefore the theorem is proved.

□

Remark 4.1. *1. With conditions (i) and (ii) of theorem (4.1), the IMH with proposal distributed as a student distribution with ν degree of freedom is also geometrically ergodic. But depending on the features of the density π, the mixing rate of the IMH chain may be slow. The restricted version of the SRWMH algorithm retains the geometric ergodicity of the IMH with the mixing rate of the RWMH algorithm on Δ.*

2. Condition (ii) of theorem (4.1) can be weakened by using a proposal kernel heavy tailed than the student density.

4.2 A simulation study.

In this simulation study, we illustrate the benefice of restricting the RWMH in terms of the existence of the Central Limit Theorem. We say that the Central Limit Theorem holds for a function h if there exists $0 < \sigma_h^2 < \infty$ such that:

$$\sqrt{n}\left(S_n(h) - \pi(h)\right) \overset{w}{\Rightarrow} N(0, \sigma_h^2),$$

where $S_n(h) = \frac{1}{n}\sum_{i=1}^n h(X_i)$. It is well know that for a geometrically ergodic Markov chain, the Central Limit Theorem holds for any $h \in L^2(\pi)$, see Roberts & Rosenthal (1997) for example.

Note $t(\nu)$ the student distribution on $\mathbb{R}$ with ν degree of freedom. Clearly, from theorem 2.2 of Jarner & Tweedie (2001), the SRWMH with target density $t(\nu)$ is not geometrically ergodic. But using a proposal distributed as $t(r)$ with $r \leq \nu$, we know from theorem 4.1 that the restricted version of the SRWMH presented above is geometrically. Thus the Central Limit Thoerem holds for any h with $\pi(h^2) < \infty$.

We use $h(x) = |x|$, $r = 0.5$ and $\nu = 3, 4$. For each of these 4 combinaisons ($\nu = 3, 4$ times ordinary and restricted SRWMH), we ran $N = 1{,}000$ independent chains starting at 0 each with length $n = 1{,}000{,}000$. For each of these 4 combinaisons, the $i-$th run is used to estimate $\pi(|x|)$ by $S^i(|X|) = \frac{1}{n}\sum_{k=1}^n \left|X_k^{(i)}\right|$. Figures 1 and 2 show the QQplot versus the normal distribution and the histogram of the normalised empirical sum $S^i(|X|) - \frac{1}{N}\sum_{i=1}^N S^i(|X|)$.

For $\nu = 3$, as expected, the Central Limit Theorem seems to hold for $h(x) = |x|$ in the case of the restricted version of the SRWMH algorithm, while it clearly fails for the ordinary SRWMH. For $\nu = 4$, the Central Limit Theorem seems to hold for both versions. This agrees with theorem 4.1 and theoretical results obtained by Jarner & Tweedie (2001) on the SRWMH algorithm.

Aknowledgements: The authors are grateful to Professor Richard Duncan for hepful discussions on Hilbert spaces operators theory.

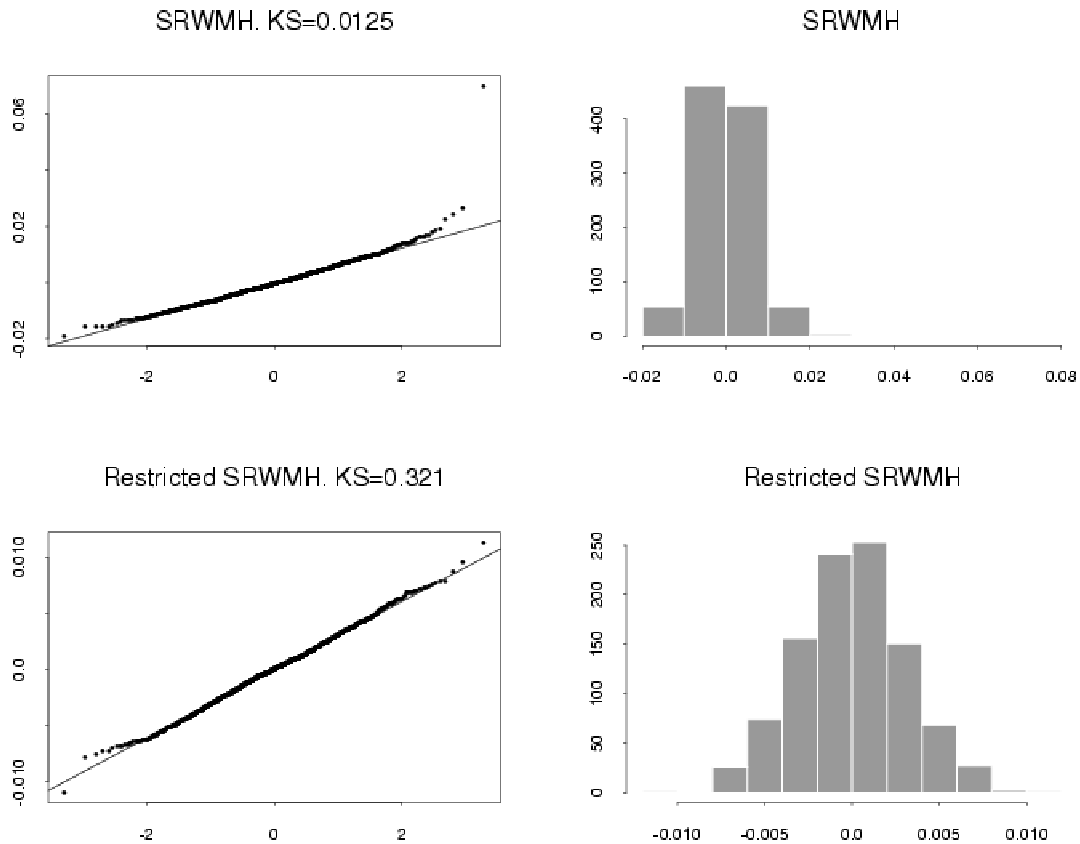

Figure 1: QQ-plot and histogram of 1,000 normalised ergodic averages of the function $h(x) = |x|$. Target density $t(3)$, proposal density $t(0.5)$. The KS value is the pvalue of the Kolmogorov-Smirnov test of normality.

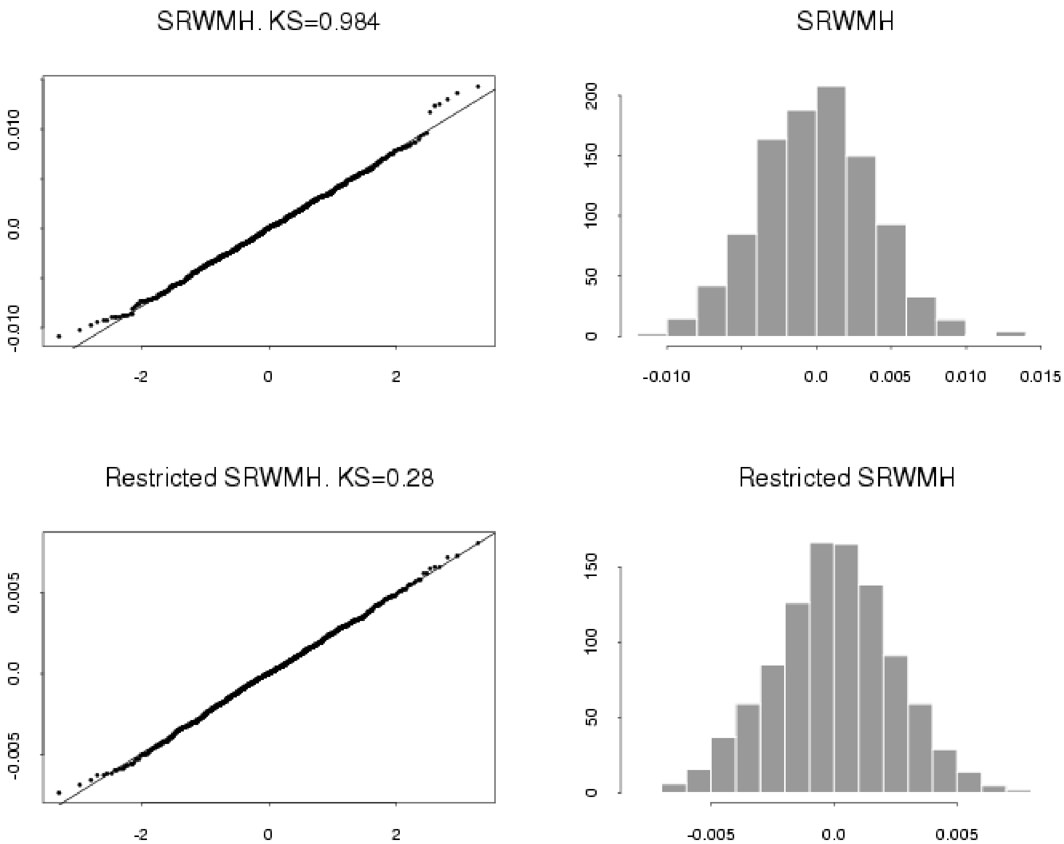

Figure 2: QQ-plot and histogram of 1,000 normalised ergodic averages of the function $h(x) = |x|$. Target density $t(4)$, proposal density $t(0.5)$. The KS value is the pvalue of the Kolmogorov-Smirnov test of normality.

References

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

Brooks, S. P. & Roberts, G. (1997). Assessing convergence of markov chain monte carlo algorithms. *MCMC Preprints* .

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

Conway, J. B. (1985). *A course in functional analysis.* New York: Springer-Verlag.

Halmos, P. R. (1957). *Introduction to Hibert space.* New York: Chelsea Publishing Company.

Jarner, S. F. & Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12**, 224–247.

Jarner, S. F. & Tweedie, R. L. (2001). Necessary conditions for geometric and polynomial ergodicity of random walk-type markov chains.. *MCMC Preprints* .

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**, 101–121.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091.

Meyn, S. P. & Tweedie, R. L. (1993). *Markov chains and stochastic stability.* London: Springer-Verlag London Ltd.

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**, 5–31. With discussion by Hemant Ishwaran and Neal Madras and a rejoinder by the authors.

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

Roberts, G. O. & Tweedie, R. L. (2000). Geometric l^2 and l^1 convergence are equivalent for reversible markov chains .

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