Two convergence properties of hybrid samplers

by

Gareth O. Roberts
Statistical Laboratory
University of Cambridge

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 9703, January 30, (1997)

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Two convergence properties of hybrid samplers

by

Gareth O. Roberts and Jeffrey S. Rosenthal

(January, 1997.)

1. Introduction.

Theoretical work on Markov chain Monte Carlo (MCMC) algorithms has so far mainly concentrated on the properties of simple algorithms such as the Gibbs sampler, or the full dimensional Hastings-Metropolis algorithm. This is understandable since even these simple algorithms are difficult to analyse, and are still not fully understood. In practice, these simple algorithms are used as building blocks for more sophisticated methods, which we shall refer to as *hybrid samplers*. It is often hoped that good convergence properties of the building blocks will translate to properties of the hybrid chains, however to date, very little work has been done to try and make these arguments rigorous. This article attempts to build on the results of Roberts and Rosenthal (1996), which consider geometric ergodicity properties of hybrid chains in terms of their constituent component algorithms.

In this paper, we concentrate on two special cases, where we can make more practical geometric ergodicity statements. In the first case, we are actually able to give quantitative result for the rate of convergence of the resulting hybrid algorithm, although this is at the expense of imposing a very strong uniform type of geometric ergodicity on the constituent component algorithms. In the second case, we consider hybrid chains arising from combining various Metropolis algorithms, and adapt results of Roberts and Tweedie (1996) to establish geometric ergodicity.

2. Preliminaries.

Recall that, given a probability distribution $\pi(\cdot)$ on the state space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_k$, the usual *deterministic-scan Gibbs sampler* (DUGS) is the Markov kernel $P = Q_1 Q_2 \ldots Q_k$, where Q_i is the Markov kernel which replaces the i^{th} coordinate by a draw from $\pi(dx_i | \{x_j\}_{j \neq i})$, leaving x_j fixed for $j \neq i$. The *random-scan Gibbs sampler* (RSGS), given by $P = \frac{1}{k} \sum_i Q_i$, is sometimes used instead. These are standard Markov chain Monte Carlo techniques (see, e.g. Gelfand and Smith, 1990; Smith and Roberts, 1993; Tierney, 1994).

Often the full conditionals $\pi(dx_i | \{x_j\}_{j \neq i})$ may be easily sampled, so that DUGS or RSGS may be efficiently run on a computer. However, sometimes this is not feasible. Instead, one can define new operators P_i which are easily implemented, such that P_i^n converges to Q_i as $n \to \infty$. This is the method of "variable-at-a-time Metropolis-Hastings" or "Metropolis within Gibbs" (cf. Tierney, 1994, Section 2.4; Chan and Geyer, 1994, Theorem 1; Green, 1994). Such samplers prompt the following definition (taken from Roberts and Rosenthal, 1996).

Definition. Let $C = (P_1, P_2, \ldots, P_k)$ be a collection of Markov kernels on a state space $\mathcal{X}$. The *random-scan hybrid sampler* for C is the sampler defined by

$$P_{RS} = \frac{1}{k}(P_1 + \ldots + P_k).$$

In addition to the variable-at-a-time Metropolis-Hastings algorithms mentioned above, such hybrid samplers often arise when larger MCMC algorithms are "constructed" out of smaller ones. For example, if the P_i are themselves RSGS samplers, then the random-scan hybrid sampler would correspond to building a large Gibbs sampler out of smaller ones. Similarly, if the P_i are themselves Metropolis-Hastings algorithms, then the hybrid sampler can again be viewed as a Metropolis-Hastings algorithm, but with (in general) a singular proposal distribution (cf. Tierney, 1995); this is considered further in Section 4 below.

Theoretical properties of such hybrid samplers were considered in Roberts and Rosenthal (1996). In particular, it was shown (Theorem 6) that if for a particular model RSGS is geometrically ergodic in an appropriate sense (say, in $L^2(\pi)$), and if $(P_i)^n \to Q_i$ as

$n \to \infty$ (again, say, in $L^2(\pi)$), then the resulting random-scan hybrid sampler would again be geometrically ergodic.

However, such a result leads to further questions. Firstly, is it possible to provide any *quantitative bounds* for these hybrid samplers? Secondly, can geometric ergodicity be established for, say, Metropolis-Hastings algorithms (which are ergodic but do *not* converge in $L^2(\pi)$)?

The first of these questions is addressed in the next section, and the second is addressed in the final section of this paper.

3. Strong uniform ergodicity and quantitative bounds.

An important and difficult problem in the theory of MCMC algorithms is to provide quantitative bounds on their distance to stationarity after a finite number of steps. Such bounds can then be used to determine how long to run the algorithm in practice, to achieve sufficient accuracy of results. While there have been some successes with this approach (see e.g. Meyn and Tweedie, 1994; Rosenthal, 1995), the question of quantitative bounds in general remains problematic.

In this section, we provide quantitative bounds on convergence rates for hybrid samplers, under a strong hypothesis about uniform convergence of the constituent Markov chains. We recall that a Markov chain is *uniformly ergodic* if there is $N \in \mathbf{N}$ and $\rho < 1$ such that $\|P^N(x, \cdot) - \pi(\cdot)\|_{\text{var}} \leq \rho$ for all $x \in \mathcal{X}$, or equivalently (cf. Meyn and Tweedie, 1993, Theorem 16.0.2) if $\sup_{x \in \mathcal{X}} \|P^n(x, \cdot) - \pi(\cdot)\|_{\text{var}} \to 0$ as $n \to \infty$.

Definition. A Markov chain $P(\cdot, \cdot)$ on a state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$, is (N, ϵ)-*strongly uniformly ergodic* for some $N \in \mathbf{N}$ and $\epsilon > 0$ if

$$P^N(x, \cdot) \;\geq\; \epsilon\, \pi(\cdot), \qquad x \in \mathcal{X}.$$

For such a chain, it follows that for $n \geq 0$,

$$P^{N+n}(x, \cdot) = \int P^N(x, dy) P^n(y, \cdot) \geq \int \epsilon \pi(dy) P^n(y, \cdot) = \epsilon \pi(\cdot).$$

In particular, P is also (k, ϵ)-strongly uniformly ergodic for any $k \geq N$.

It also follows immediately (see e.g. Meyn and Tweedie, 1993, Theorem 16.0.2) that $\|P^{tN}(x,\cdot) - \pi(\cdot)\|_{\text{var}} \leq (1-\epsilon)^t$ for $t = 1, 2, \ldots$, for any $x \in \mathcal{X}$; thus, strong uniform ergodicity implies uniform ergodicity. The converse to this implication is considered in the following Proposition.

Proposition 1. *In general, a uniformly ergodic Markov chain need not be strongly uniformly ergodic. However, if a Markov chain is both uniformly ergodic and reversible, then it is strongly uniformly ergodic.*

Proof. For a counter-example, let $\mathcal{X}$ be the set of all non-negative integers, and set $P(n, 0) = P(n, n+1) = \frac{1}{2}$, for all $n \in \mathcal{X}$. Then this Markov chain is easily seen to be uniformly ergodic but not strongly uniformly ergodic.

Suppose now that the Markov chain is reversible. By uniform ergodicity, we have that $P^n(x,\cdot) \geq \epsilon\nu(\cdot)$ for all $x \in \mathcal{X}$, for some $n \in \mathbf{N}$, $\epsilon > 0$, and probability measure ν on $\mathcal{X}$ (cf. Meyn and Tweedie, 1993, Theorem 16.0.2). But then by reversibility,

$$\pi(dx)P^n(x, dy) = \pi(dy)P^n(y, dx) \geq \pi(dy)\epsilon\nu(dx), \qquad x, y \in \mathcal{X},$$

so that $P^n(x, dy) \geq \epsilon\frac{d\nu}{d\pi}(x)\pi(dy)$. Choose $A \subseteq \mathcal{X}$ and $\delta > 0$ such that $\frac{d\nu}{d\pi}(x) > \delta$ for all $x \in A$, and $\pi(A) > 0$ (so that $\nu(A) > 0$). Then for any $z \in \mathcal{X}$, setting $K = \epsilon^2\delta\nu(A) > 0$, we have

$$P^{2n}(z, dy) \geq P^n(z, A) \inf_{x \in A} P^n(x, dy) \geq \epsilon\nu(A)\,\epsilon\delta\pi(dy) = K\pi(dy),$$

as required. ∎

Remark. It is easily seen that being strongly uniformly ergodic is equivalent to the existence of a strong stationary time (cf. Aldous and Diaconis, 1987) which is independent of the process itself.

We now use strong uniform ergodicity to establish quantitative bounds on certain hybrid samplers. We adopt the notation

$$x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k),$$

$$\mathcal{X}_{-i} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_{i-1} \times \mathcal{X}_{i+1} \times \ldots \times \mathcal{X}_k \,,$$

and

$$x^*_{-i} = \{x_1\} \times \ldots \times \{x_{i-1}\} \times \mathcal{X}_i \times \{x_{i+1}\} \times \ldots \times \{x_k\} \,.$$

Theorem 2. *Let $\pi(\cdot)$ be a probability distribution on a state space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$. For $1 \le i \le k$, let $N_i \in \mathbf{N}$ and $\epsilon_i > 0$ be given, and let P_i be a Markov kernel on $\mathcal{X}$ which fixes coordinates other than i. Assume that for each $x_{-i} \in \mathcal{X}_{-i}$, $P_i\Big|_{x^*_{-i}}$ has stationary distribution $\pi(\cdot|x_{-i})$ and is (N_i, ϵ_i)-strongly uniformly ergodic. Assume further that RSGS, with stationary distribution $\pi(\cdot)$, is (N', ϵ')-strongly uniformly ergodic. Then the random-scan hybrid sampler $P_{RS} = \frac{1}{k}(P_1 + \ldots + P_k)$ is (N_*, ϵ_*)-strongly uniformly ergodic, where*

$$N_* = N' \max_{1\le i\le k} \{N_i\} \,; \qquad \epsilon_* = \epsilon' \min_{1\le i\le k} \{\epsilon_i^{N'}\} k^{-N'\left(\max\limits_{1\le i\le k}\{N_i\}-1\right)} \,.$$

Proof. As usual, let Q_i be the Markov kernel which replaces the i^{th} coordinate by a draw from $\pi(dx_i|x_{-i})$, leaving x_{-i} fixed.

It follows from the hypotheses that

$$P_i^n(x,\cdot) \;\ge\; \epsilon_i\, Q_i(x,\cdot) \,, \qquad n \ge N_i \,, \qquad i = 1, 2, \ldots, k$$

and

$$\left[\frac{1}{k}\Big(Q_1(x,\cdot) + \ldots + Q_k(x,\cdot)\Big)\right]^{N'} \;\ge\; \epsilon'\, \pi(\cdot) \,.$$

Then

$$\begin{aligned}
(P_{RS})^{N'\max\{N_i\}}(x,\cdot) \;&=\; \left[\frac{1}{k}\Big(P_1(x,\cdot) + \ldots + P_k(x,\cdot)\Big)\right]^{N'\max\{N_i\}} \\
&\ge\; \left(k^{-\max\{N_i\}}\left[P_1^{\max\{N_i\}}(x,\cdot) + \ldots + P_k^{\max\{N_i\}}(x,\cdot)\right]\right)^{N'} \\
&\ge\; \left(k^{-(\max\{N_i\}-1)} \min\{\epsilon_i\} \frac{1}{k}\Big[Q_1(x,\cdot) + \ldots + Q_k(x,\cdot)\Big]\right)^{N'} \\
&\ge\; k^{-N'(\max\{N_i\}-1)} \min\{\epsilon_i^{N'}\}\, \epsilon'\, \pi(\cdot) \,,
\end{aligned}$$

giving the result. ∎

It follows immediately that

$$\|(P_{RS})^{tN_*}(x,\cdot)-\pi(\cdot)\|_{\text{var}} \;\leq\; (1-\epsilon_*)^t\,, \qquad t=1,2,\ldots$$

In particular, if $N' = N_1 = \ldots = N_k = 1$, then $N_* = 1$ and $\epsilon_* = \epsilon' \min_{1\leq i\leq k}\{\epsilon_i\}$, so that $\|(P_{RSh})^t(x,\cdot)-\pi(\cdot)\|_{\text{var}} \leq (1-\epsilon'\min\{\epsilon_i\})^t$.

4. Hybrid Metropolis chains.

In this Section, we consider hybrid samplers whose consituent chains P_i each arise from a Metropolis algorithm (see Metropolis et al., 1953; Hastings, 1970; Smith and Roberts, 1993) on the i^{th} coordinate. These hybrid samplers may themselves be regarded as Metropolis algorithms, but with singular proposal distributions (cf. Tierney, 1995). We shall prove that, under appropriate conditions, the hybrid samplers will be geometrically ergodic. Our proof uses the theory of drift and minorization conditions for general Markov chains, as in Nummelin (1984) or Meyn and Tweedie (1993), and follows a similar argument to Roberts and Tweedie (1996). Specifically, we shall eventually show that all bounded sets are small for P_{RS}, and that for an appropriate function V (which will need to depend on the dimension k), we have $\limsup_{|\mathbf{x}|\to\infty} P_{RS}V(\mathbf{x})/V(\mathbf{x}) < 1$. [Recall the definition $Pf(\mathbf{x}) = \int f(\mathbf{y})P(\mathbf{x},d\mathbf{y})$, and that a set C is *small* for P if there is $n\in\mathbf{N}$, $\epsilon>0$, and a probabiilty measure $\nu(\cdot)$, such that $P^n(x,\cdot)\geq\epsilon\nu(\cdot)$ for all $x\in C$.]

Let $\pi:\mathbf{R}^k\to\mathbf{R}^{>0}$ be a positive C^1 density (with respect to k-dimensional Lebesgue measure) for a probability distribution on the state space $\mathbf{R}^k$. For $1\leq i\leq k$, let P_i be a symmetric random-walk Metropolis algorithm (with respect to $\pi(\cdot)$) on the i^{th} coordinate. Thus, started from the k-vector $\mathbf{x}$, the proposal in the i^{th} direction is given by $\mathbf{x}+\mathbf{Z}_i$ where $\mathbf{Z}_i$ is drawn from a symmetric *increment distribution* $S_i(\mathbf{x},\cdot)$; this proposal is then accepted with probability $\min(1,\,\pi(\mathbf{x}+\mathbf{Z}_i)/\pi(\mathbf{x}))$. Suppose that $S_i(\mathbf{x},\cdot)$ has density $q_i(y\mathrm{e}_i)$ with respect to one-dimensional Lebesgue measure, where $\mathbf{e}_i$ denotes the ith coordinate vector. We shall assume for simplicity that for each i, there exist positive constants ϵ_i and δ_i such that

$$q_i(y\mathbf{e}_i)\geq\epsilon_i \text{ for } |y|<\delta_i. \tag{1}$$

We shall simplify $q_i(y\mathbf{e}_i)$ to $q_i(y)$ where no confusion is possible. Finally, we let P_{RS} be as in Section 2.

We introduce the following conditions on π. We will assume that π is bounded, that for sufficiently small $d > 0$ we have

$$\int_{\mathbf{R}^k} \pi^{1-d}(\mathbf{x}) d\mathbf{x} < \infty \,, \tag{2}$$

and that we have the "asymptotically exponentially decreasing tails" condition

$$\limsup_{|\mathbf{x}| \to \infty} \mathbf{n}(\mathbf{x}) \cdot \nabla \log \pi(\mathbf{x}) < 0 \,, \tag{3}$$

where $\mathbf{n}(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$ is the unit vector in the direction of $\mathbf{x}$. For each $\mathbf{x}$, let $\kappa(\mathbf{x})$ denote the Gauss-Bonnet curvature of the contour $\{\mathbf{y};\ \pi(\mathbf{y}) = \pi(\mathbf{x})\}$ at the point $\mathbf{x}$. We assume that $\kappa(\mathbf{x})$ is well-defined, at least for sufficiently large $|\mathbf{x}|$. We further assume that

$$\lim_{|\mathbf{x}| \to \infty} \kappa(\mathbf{x}) = 0 \,, \tag{4}$$

that

$$\limsup_{|\mathbf{x}| \to \infty} |(\nabla \log |\nabla \log \pi(\mathbf{x})|)| < \infty \,, \tag{5}$$

and that

$$\pi \text{ is log-concave, at least outside some compact set.} \tag{6}$$

Given $\mathbf{x} \in \mathbf{R}^k$, let $A_i(\mathbf{x}) = \{\mathbf{z};\ \mathbf{z} = y\mathbf{e}_i \text{ and } \pi(\mathbf{x}+\mathbf{z}) \geq \pi(\mathbf{x})\}$ and let $R_i(\mathbf{x}) = \{\mathbf{z};\ \mathbf{z} = y\mathbf{e}_i \text{ and } \pi(\mathbf{x}+\mathbf{z}) < \pi(\mathbf{x})\}$. In other words, $A_i(\mathbf{x})$ represents the set of points which if proposed under $S_i(\mathbf{x}, \cdot)$, would always be accepted, whereas $R_i(\mathbf{x})$ represents those which are rejected with positive probability. We will also need the *reflected set*, $A^i(\mathbf{x})^r = \{\mathbf{x};\ -\mathbf{x} \in A^i(\mathbf{x})\}$.

We introduce the drift function $V(\mathbf{x}) = \pi(\mathbf{x})^{-d}$, with $d > 0$ sufficiently small (as described later). The following calculation will be useful.

Proposition 3. *For all P_i,*

$$P_i V(\mathbf{x}) \leq r(d) V(\mathbf{x})$$

where $r(d) = 1 + (1-d)^{(1-d)/d} d$, for all $\mathbf{x} \in \mathcal{X}$. Hence, for all $\epsilon > 0$, there is d with $0 < d < \epsilon$, such that $1 < r(d) < 1 + \epsilon$.

Proof. Considering separately the cases where the proposal is to $R_i(\mathbf{x})$ and is rejected (so the value of V is unchanged), where the proposal is to $R_i(\mathbf{x})$ and is accepted, and where the proposal is to $A_i(\mathbf{x})$ (and is necessarily accepted), we have that

$$\frac{P_i V(\mathbf{x})}{V(\mathbf{x})} = \int_{R_i(\mathbf{x})} q_i(y) \left(1 - \frac{\pi(\mathbf{x}+y\mathbf{e}_i)}{\pi(\mathbf{x})}\right) dy +$$

$$\int_{R_i(\mathbf{x})} q_i(y) \frac{\pi(\mathbf{x}+y\mathbf{e}_i)^{-d}}{\pi(\mathbf{x})^{-d}} \left(\frac{\pi(\mathbf{x}+y\mathbf{e}_i)}{\pi(\mathbf{x})}\right) dy \ + \ \int_{A_i(\mathbf{x})} q_i(y) \frac{\pi(\mathbf{x}+y\mathbf{e}_i)^{-d}}{\pi(\mathbf{x})^{-d}} dy$$

$$= \int_{\mathbf{R}} q_i(y) I(\mathbf{x}+y\mathbf{e}_i) dy \,,$$

where

$$I(\mathbf{z}) = \begin{cases} 1 - \pi(\mathbf{z})/\pi(\mathbf{x}) + (\pi(\mathbf{z})/\pi(\mathbf{x}))^{1-d}, & \mathbf{z} \in R_i(\mathbf{x}) \\ (\pi(\mathbf{x})/\pi(\mathbf{z}))^d, & \mathbf{z} \in A_i(\mathbf{x}) \,. \end{cases}$$

We claim that $I(\mathbf{z}) \le r(d)$ for all $\mathbf{z} \in R_i(\mathbf{x}) \cup A_i(\mathbf{x})$. Indeed, $I(\mathbf{z}) \le 1$ on $A_i(\mathbf{x})$ by definition. Furthermore, setting $w = \pi(\mathbf{z})/\pi(\mathbf{x})$, we have that for $\mathbf{z} \in R_i(\mathbf{x})$, $I(\mathbf{z}) = 1 - w + w^{1-d}$ with $0 \le w \le 1$. This is maximised at $w = (1-d)^{1/d}$ with maximising value $r(d)$ above. The inequality follows.

The second statement is immediate since $\lim_{d \to 0+} r(d) = 1$. ∎

Lemma 4. *All bounded subsets of* $\mathbf{R}^k$ *are small for* P_{RS}.

Proof. By (1), it is easy to see that $P_{RS}^k(x, \cdot)$ has continuous component with respect to k-dimensional Lebesgue measure. Call this continuous component $s(\mathbf{x}, \cdot)$, say. Note that by (1), for suitable constants ϵ and δ, we have

$$s(\mathbf{x}, \mathbf{x}+\mathbf{y}) \ge \epsilon \,, \qquad \text{whenever } |y_i| \le \delta \text{ for } 1 \le i \le k \,.$$

Hence the set $[-\delta', \delta']^k$ is small. By taking convolutions, it follows that for any $N \in \mathbf{N}$, there is $\epsilon' > 0$ such that the continuous component $s_N(\mathbf{x}, \cdot)$ of $P_{RS}^{kN}(\mathbf{x}, \cdot)$ satisfies

$$s_N(\mathbf{x}, \mathbf{x}+\mathbf{y}) \ge \epsilon' \,, \qquad \text{whenever } |y_i| \le N\delta/2 \text{ for } 1 \le i \le k \,.$$

Hence, the set $[-N\delta/2, N\delta/2]^k$ is small. The result follows since any bounded set C is contained in $[-N\delta/2, N\delta/2]^k$ for some sufficiently large N. ∎

Theorem 5. *Suppose conditions (2) to (5) are satisfied. Then the random scan hybrid chain P_{RS} is geometrically ergodic.*

Proof. Because of Lemma 4, and by (2) which ensures that $V \in L^1(\pi)$ for sufficiently small d, it suffices (see e.g. Nummelin, 1984, Proposition 5.21; Meyn and Tweedie, 1993, Theorem 15.0.1; Roberts and Tweedie, 1996) to demonstrate that

$$\limsup_{|\mathbf{x}|\to\infty} \frac{P_{RS}V(\mathbf{x})}{V(\mathbf{x})} < 1 \ .$$

So, for contradiction, suppose that we have a sequence of points $\{\mathbf{x}_j\}$, with $|\mathbf{x}_j| \to \infty$, such that

$$\liminf_{i\to\infty} \frac{P_{RS}V(\mathbf{x}_j)}{V(\mathbf{x}_j)} \geq 1 \ .$$

By taking a subsequence if necessary, we can (and do) assume that $\nabla \log \pi(\mathbf{x}_i)/|\nabla \log \pi(\mathbf{x}_i)|$ converges to a limiting direction $\mathbf{f}$. There must be at least one coordinate direction $\mathbf{e}_i$ with $\mathbf{f}\cdot\mathbf{e}_i \neq 0$. By renumbering the coordinates as necessary, we assume that $1 \leq n \leq k$ is such that $\mathbf{e}_i \cdot \mathbf{f} \neq 0$ for $1 \leq i \leq n$ but that $\mathbf{f}$ is orthogonal to $\mathbf{e}_i$ for $n+1 \leq i \leq k$.

Now (3) and (4) imply that $A_i(\mathbf{x})^r \,\Delta R_i(\mathbf{x}) \to 0$ in q_i-measure as $|\mathbf{x}| \to \infty$, for $1 \leq i \leq n$, where Δ denotes symmetric difference of sets.

We take d sufficiently small that $r(d) < \frac{2k-1}{2k-2}$. We compute that by (6), for large enough $|\mathbf{x}|$ we have

$$\begin{aligned}\frac{P_iV(\mathbf{x})}{V(\mathbf{x})} &= \int_{R_i(\mathbf{x})} \left(1 - \frac{\pi(\mathbf{x}+y\mathbf{e}_i)}{\pi(\mathbf{x})} + \frac{\pi(\mathbf{x}+y\mathbf{e}_i)^{1-d}}{\pi(\mathbf{x})^{1-d}}\right) q_i(y)dy + \int_{A_i(\mathbf{x})} \frac{\pi(\mathbf{x})^d}{\pi(\mathbf{x}+y\mathbf{e}_i)^d} q_i(y)dy \\ &\leq \int_{R_i(\mathbf{x})} \left(1 - \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}-y\mathbf{e}_i)} + \frac{\pi(\mathbf{x})^{1-d}}{\pi(\mathbf{x}-y\mathbf{e}_i)^{1-d}}\right) q_i(y)dy + \int_{A_i(\mathbf{x})} \frac{\pi(\mathbf{x})^d}{\pi(\mathbf{x}+y\mathbf{e}_i)^d} q_i(y)dy \ .\end{aligned} \tag{7}$$

Now, using the symmetry of q_i and the asymptotic complementary properties of $R_i(\mathbf{x})$ and $A_i(\mathbf{x})$, we see that as $|\mathbf{x}| \to \infty$,

$$\int_{A_i(\mathbf{x})} \frac{\pi(\mathbf{x})^d}{\pi(\mathbf{x}+y\mathbf{e}_i)^d} q_i(y)dy \ \approx \ \int_{R_i(\mathbf{x})} \frac{\pi(\mathbf{x})^d}{\pi(\mathbf{x}-y\mathbf{e}_i)^d} q_i(y)dy$$

and

$$\int_{R_i(\mathbf{x})} 2q_i(y)dy \ \approx \ 1\,.$$

So, taking limsups along the sequence $\{\mathbf{x}_j\}$, we have that for $1 \leq i \leq n$,

$$\limsup_{j\to\infty} \frac{P_i V(\mathbf{x}_j)}{V(\mathbf{x}_j)} \leq \limsup_{j\to\infty} \int_{R_i(\mathbf{x}_j)} \left(1 - \frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)} + \frac{\pi(\mathbf{x}_j)^{1-d}}{\pi(\mathbf{x}_j - y\mathbf{e}_i)^{1-d}} + \frac{\pi(\mathbf{x}_j)^d}{\pi(\mathbf{x}_j - y\mathbf{e}_i)^d}\right) q_i(y)dy$$

$$= \limsup_{j\to\infty} \int_{R_i(\mathbf{x}_j)} \left(2 - \left[1 + \frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)} - \frac{\pi(\mathbf{x}_j)^{1-d}}{\pi(\mathbf{x}_j - y\mathbf{e}_i)^{1-d}} - \frac{\pi(\mathbf{x}_j)^d}{\pi(\mathbf{x}_j - y\mathbf{e}_i)^d}\right]\right) q_i(y)dy$$

$$= \limsup_{j\to\infty} \int_{R_i(\mathbf{x}_j)} \left[2 - \left(1 - \left(\frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)}\right)^d\right)\left(1 - \left(\frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)}\right)^{1-d}\right)\right] q_i(y)dy$$

$$= 1 - \liminf_{j\to\infty} \int_{R_i(\mathbf{x}_j)} \left(1 - \left(\frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)}\right)^d\right)\left(1 - \left(\frac{\pi(\mathbf{x}_j)}{\pi(\mathbf{x}_j - y\mathbf{e}_i)}\right)^{1-d}\right) q_i(y)dy$$

$$< 1\,, \tag{8}$$

with the final inequality following from (3) and Fatou's Lemma. Therefore,

$$\limsup_j P_i V(\mathbf{x}_j)/V(\mathbf{x}_j) < 1\,, \qquad 1 \leq i \leq n\,.$$

To finish, consider the sequence $\{c_j\}$, where $c_j = |\nabla \log \pi(\mathbf{x}_j)|$. By (3), $\liminf_j c_j > 0$. Therefore (again by subsequencing if necessary) we can assume that $c_j \to c_\infty$ for some $c_\infty \in (0, \infty]$. We need to consider separately the cases where c_∞ is ether finite or infinite.

If $c_\infty < \infty$, then by (5) and (3), we have for $i > n$ that $\lim_{j\to\infty} \frac{\pi(\mathbf{x}_j + y\mathbf{e}_i)}{\pi(\mathbf{x}_j)} = 1$ for all $y \in \mathbf{R}$, so that $\lim_{j\to\infty} P_i V(\mathbf{x}_j)/V(\mathbf{x}_j) = 1$. Therefore by (8), $\limsup_{j\to\infty} P_{RS} V(\mathbf{x}_j)/V(\mathbf{x}_j) < 1$ for a contradiction.

If $c_\infty = \infty$, then for $i \leq n$, all proposed jumps into $R_i(\mathbf{x}_j)$ are asymptotically rejected and all jumps to $A_i(\mathbf{x}_j)$ are accepted, but $V(\mathbf{x}_j + y\mathbf{e}_i)/V(\mathbf{x}_j) \to 0$ as $j \to \infty$. More precisely, for $i \leq n$, $\lim_{j\to\infty} \frac{\pi(\mathbf{x}_j + y\mathbf{e}_i)}{\pi(\mathbf{x}_j)} = 0$ for all $y \in R_i(\mathbf{x}_j) - \{0\}$. Furthermore, for all $y \in A_i(\mathbf{x}_j)$, $\lim_{j\to\infty} \frac{\pi(\mathbf{x}_j + y\mathbf{e}_i)}{\pi(\mathbf{x}_j)} = \infty$. It follows that the integrand in (7) converges to $\mathbf{1}_{R_i(\mathbf{x})}(y)$, and since the integrand in (7) is uniformly bounded (by $r(d)$), we have by the dominated convergence theorem that

$$\lim_{j\to\infty} \frac{P_i V(\mathbf{x}_j)}{V(\mathbf{x}_j)} = \lim_{j\to\infty} \int_{\mathbf{R}} \mathbf{1}_{R_i(\mathbf{x})}(y) q_i(y)dy = 1/2\,.$$

It follows from Proposition 3 that

$$\lim_{j\to\infty} P_{RS}V(\mathbf{x}_j)/V(\mathbf{x}_j) \leq \frac{n}{2k}+\frac{r(d)(k-n)}{k} \leq \frac{1}{2k}+\frac{r(d)(k-1)}{k} < \frac{1}{2k}+\frac{(k-1)(2k-1)}{k(2k-2)} = 1\,,$$

for a contradiction in this case. ■

Remark. The nature of the proof of Theorem 5 suggests that explicit bounds on the total variation distance from stationarity (cf. Meyn and Tweedie, 1994; Rosenthal, 1995) may be obtainable in this case, though we do not pursue that here.

REFERENCES

D. Aldous and P. Diaconis (1987), Strong stopping times and finite random walks. Adv. Appl. Math. **8**, 69-97.

K.S. Chan and C.J. Geyer (1994), Discussion paper. Ann. Stat. **22**, 1747-1758.

A.E. Gelfand and A.F.M. Smith (1990), Sampling based approaches to calculating marginal densities. J. Amer. Stat. Assoc. **85**, 398-409.

W.K. Hastings (1970), Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97-109.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087-1091.

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

S.P. Meyn and R.L. Tweedie (1994), Computable bounds for convergence rates of Markov chains. Ann. Appl. Prob. **4**, 981-1011.

G.O. Roberts and J.S. Rosenthal (1996), Geometric ergodicity and hybrid Markov chains. Preprint.

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

J.S. Rosenthal (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558-566.

A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3-24.

L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701-1762.

L. Tierney (1995), A note on Metropolis Hastings kernels for general state spaces. Tech. Rep. **606**, School of Statistics, U. of Minnesota.