Convergence of slice sampler Markov chains

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

(July 1997; last revised September 10, 1998.)

In this paper, we analyse theoretical properties of the slice sampler. We find that the algorithm has extremely robust geometric ergodicity properties. For the case of just one auxiliary variable, we demonstrate that the algorithm is stochastically monotone, and deduce analytic bounds on the total variation distance from stationarity of the method using Foster-Lyapunov drift condition methodology.

1. Introduction.

This paper considers the use of *slice samplers* to sample from a complicated d-dimensional probability distribution. Slice samplers are a form of *auxiliary variable* technique, which introduces auxiliary random variables $Y_1, \ldots, Y_k$ to facilitate the design of an improved Markov chain Monte Carlo (MCMC) sampling algorithm.

The idea of using auxiliary variables for improving MCMC was introduced for the Ising model by Swendsen and Wang (1987). Edwards and Sokal (1988) generalised the Swendsen-Wang technique, and since then, their use in statistical problems has gradually increased, partly as a result of Besag and Green (1993). In recent years there has been a large amount of activity on this topic, including a very clear discussion of auxiliary variable techniques by Higdon (1996), a variety of examples of uses of auxiliary variable techniques in statistical problems by Damien *et. al.* (1997), and some theoretical progress by Mira and Tierney (1997) and Fishman (1996). The slice sampler is a particularly interesting

* Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K. Internet: `G.O.Roberts@statslab.cam.ac.uk`. Supported in part by EPSRC of the U.K.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: `jeff@utstat.toronto.edu`. Supported in part by NSERC of Canada.

algorithm from a practical point of view, since it frequently allows very straightforward implementation (see for example Damien *et. al.,* 1997, and Neal, 1997).

However, except for the original Swendsen-Wang method (which has been shown to be superior to more naive Gibbs methods for sub-critical Ising models), rather little is known about the theoretical properties of auxiliary variable algorithms. In this paper, we concentrate on the slice sampler, an important special case of an auxiliary variable method. We give a number of results that demonstrate that the algorithms have extremely good theoretical properties.

In Section 2, we introduce the algorithm, demonstrate that apparently more general versions of the algorithm can be reduced to the problem of sampling from a uniform density on a particular region. For the simplest case where the number of auxiliary variables is one (the *simple slice sampler*), this can be seen as sampling from the uniform density in the region bounded above by the density of interest (see Figure 2.1). In Section 3 we prove that the Markov chain induced by the algorithm has other useful invariance properties, and also that the simple slice sampler is stochastically monotone under an appropriate ordering on its state space.

In Section 4, we shall show that the simple slice sampler is nearly always geometrically ergodic using Foster-Lyapunov drift condition techniques. This result is interesting though rather surprising considering the fact that very few other MCMC algorithms exhibit comparably robust properties (see especially Roberts and Rosenthal, 1997 in this context).

Moreover, the stochastic monotonicity properties of the algorithm allow us to give useful rigorous quantitative bounds on the total variation distance from stationarity after a given number of iterations. Results of this type are described in Section 5, including a rather general statement that all distributions satisfying (5.4) (this condition is similar to requiring log-concavity of the density) converge to stationarity in less than 530 iterations (Theorem 12). The techniques used in Section 5 involve quantitative bounds recently developed by Roberts and Tweedie (1998).

In Section 6, we consider further properties of the so-called *product slice sampler*, corresponding to $k \geq 2$ above. We give conditions which ensure geometric ergodicity of the algorithm in this case. The conditions given here are sufficient, but we suspect far from necessary. Certainly, further work is required here to understand further the combined effect of a collection of auxiliary variables.

Finally in Section 7, a special case of the product slice sampler (the *opposite slice sampler*) is analysed, and conditions given for its geometric ergodicity.

2. Slice samplers: definitions and preliminaries.

Suppose that $\pi : \mathbf{R}^d \to [0, \infty)$ is a density (i.e., a non-negative measurable function which is not a.e. 0) with respect to d-dimensional Lebesgue measure. Such a density gives rise to a probability measure $\nu_\pi(\cdot)$, by

$$\nu_\pi(A) \;=\; \frac{\int\limits_A \pi(\mathbf{x}) d\mathbf{x}}{\int\limits_{\mathbf{R}^d} \pi(\mathbf{x}) d\mathbf{x}}\,, \qquad A \subseteq \mathbf{R}^d\,.$$

Typically, π is a complicated function, and d is reasonably large. The *slice sampler* then provides a Markov chain algorithm which can be used to sample from $\nu_\pi(\cdot)$.

Specifically, suppose π can be written as $\pi(\mathbf{x}) = \prod_{i=0}^k f_i(\mathbf{x})$, for some functions $f_i : \mathbf{R}^d \to [0, \infty)$. The f_0-*slice sampler*, P_{f_0}, proceeds as follows. Given $\mathbf{X}_n$, we sample k independent uniform random variables $Y_{n+1,1}, Y_{n+1,2}, \ldots, Y_{n+1,k}$, with $Y_{n+1,i} \sim \mathcal{U}(0, f_i(\mathbf{X}_n))$. We then sample $\mathbf{X}_{n+1}$ from the truncated probability distribution having density proportional to $f_0(\cdot)\mathbf{1}_{L(\mathbf{Y}_{n+1})}(\cdot)$, where

$$L(\mathbf{y}) \;=\; \left\{\mathbf{x} \in \mathbf{R}^d\,;\; f_i(\mathbf{x}) \geq y_i\,,\;\; i = 1, 2, \ldots, k\right\}\,.$$

As a motivating example, consider the following. Suppose $\pi(\mathbf{x}) = \exp\{-\|\mathbf{x}\|^2/2\} \times (1 + \|\mathbf{x} - \mathbf{x}_0\|^4) \times (1 + \|\mathbf{x} - \mathbf{x}_1\|^2)$ for constants $\mathbf{x}_0$ and $\mathbf{x}_1 \in \mathbf{R}^d$. The form of π suggests the following decomposition: $f_0(\mathbf{x}) = \exp\{-\|\mathbf{x}\|^2/2\}$, $f_1(\mathbf{x}) = 1 + \|\mathbf{x} - \mathbf{x}_0\|^4$ and $f_2(\mathbf{x}) = 1 + \|\mathbf{x} - \mathbf{x}_1\|^2$, that is set up for an f_0-slice sampler with $k = 2$. This is a particularly appealing factorisation since each of the functions f_1 and f_2 are invertible, so that the sets $\{L(\mathbf{y});\; \mathbf{y} \in \mathbf{R}^2\}$ are easy to identify. Implementation of the algorithm is therefore straightforward by iterating between

(1) sampling from the truncated normal distribution, $N(\mathbf{0}, I_d)$ conditioned on being in the set $L(Y_1, Y_2) = \{\mathbf{x};\ (1 + |\mathbf{x} - \mathbf{x}_0|^4) \geq Y_1,\ (1 + |\mathbf{x} - \mathbf{x}_1|^4) \geq Y_2\}$ (perhaps done by rejection sampling); and

(2) sampling new values Y_i from $\mathcal{U}\,(0, f_i(\mathbf{X}))$ independently, $i = 1, 2$.

Returning to the general case, the algorithm gives rise to a Markov chain $\{\mathbf{X}_n\}_{n=0}^{\infty}$, having transition probabilities $P_{f_0}(\mathbf{x}, A) \equiv \mathbf{P}(\mathbf{X}_{n+1} \in A \,|\, \mathbf{X}_n = \mathbf{x})$. This Markov chain has $\nu_\pi(\cdot)$ as a stationary distribution. To see this, just note that the Markov chain $(\mathbf{X}, \mathbf{Y})$ is a Gibbs sampler on the distribution with density $f_0(\mathbf{x})$ with respect to Lebesgue measure on the region $\{(\mathbf{x}, \mathbf{y});\ \mathbf{x} \in L(\mathbf{y})\}$. By conditional independence of the elements of $\mathbf{Y}$ given $\mathbf{X}$, $\mathbf{X}$ is also a Markov chain, and by integrating out $\mathbf{y}$, its marginal stationary distribution is $\nu_\pi(\cdot)$. Figure 2.1 illustrates a typical sample path for the case $d = k = 1$.

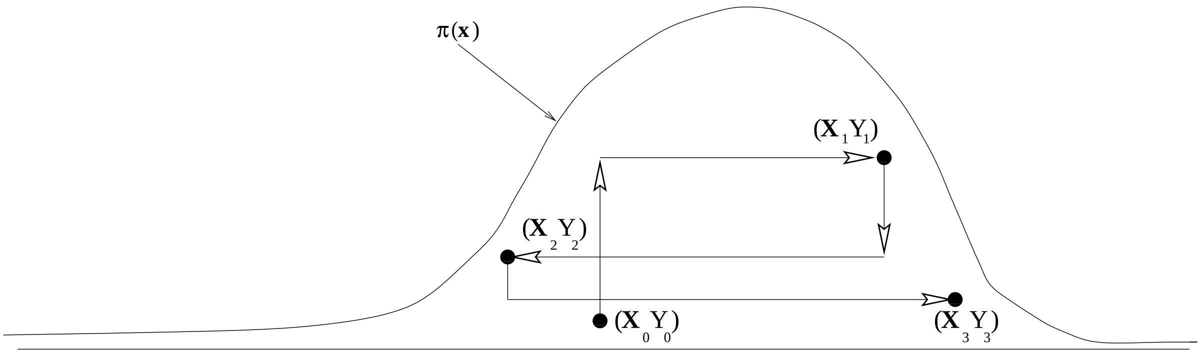

Figure 2.1 The simple slice sampler. This carries out a Gibbs sampler on the area beneath the curve of the density of π.

Furthermore, it is easily seen that the Markov chain induced by the slice sampler is ν_π-irreducible and aperiodic. Thus, from standard Markov chain theory (see e.g. Tierney, 1994) it follows that from ν_π-almost every starting point, the law of $\mathbf{X}_n$ will converge to $\nu_\pi(\cdot)$ as $n \to \infty$.

The algorithms as they have been described in this section are all constructed for densities with respect to Lebesgue measure. There are no complications in extending the algorithm to discrete distributions. This was shown by Fishman (1996), who goes on to give characterisations for the eigenvalues of the Markov chain. This can, in turn, be used to give guidelines as to the construction of slice samplers. Other work on trying to choose particularly effective slice samplers appears in Mira and Tierney (1997), where some results on the best way of choosing the factorisation $\pi(\mathbf{x}) = \prod_{i=1}^{k} f_i(\mathbf{x})$ are given.

3. Characterising the convergence properties.

It turns out that as far as analysing the Markov chains induced by these algorithms, it is sufficient to consider the case where f_0 is constant. See the appendix for a formal justification of this. All the statements we make from now on have corresponding statements for the case where f_0 is not constant.

As a result of this, we concentrate from now on on the *uniform slice sampler*, i.e. on the case when f_0 takes on only the values 0 and 1. In this case, we shall write the slice sampler Markov chain transition probabilities as P_{ssl} (for "simple slice") when $k = 1$, and as P_{psl} (for "product slice") when $k \geq 2$.

For P_{ssl} we shall write $L(y) = \{\mathbf{x} \in \mathbf{R}^d\,;\, \pi(\mathbf{x}) \geq y\}$, and shall write $Q(y)$ for $m(L(y))$, where m is d-dimensional Lebesgue measure. The algorithm then proceeds by alternately updating $Y_{n+1} \sim \mathcal{U}[0, \pi(\mathbf{X}_n)]$, and $\mathbf{X}_{n+1} \sim \mathcal{U}(L(Y_{n+1}))$. Therefore by integrating out the distribution of Y_{n+1} we can write down the transition probabilities of $\mathbf{X}$ as

$$\mathbf{P}\left(\pi(\mathbf{X}_{n+1}) < z \mid \pi(\mathbf{X}_n) = y\right) \;=\; \frac{1}{y}\int_0^y \max\left(1 - \frac{Q(z)}{Q(w)}, 0\right) dw\,. \tag{3.1}$$

Note that the behaviour of the simple slice sampler is *completely determined* by the function Q; indeed, two different densities which gave rise to the same function Q would have identical simple-slice-sampler convergence properties. This is also true for constant scaling, as the following proposition records.

Proposition 1. *Let π and $\widetilde{\pi}$ be two different densities, of dimension d and $\widetilde{d}$ respectively. Suppose there exists $a > 0$ such that their corresponding functions Q and $\widetilde{Q}$ satisfy $Q(y) = \widetilde{Q}(ay)$, for all $y > 0$. Then the convergence properties of the (uniform) simple slice sampler P_{ssl} for π and for $\widetilde{\pi}$ are identical. Specifically, we have*

$$\mathbf{P}\left(\pi(\mathbf{X}_{n+1}) < z \mid \pi(\mathbf{X}_n) = y\right) \;=\; \mathbf{P}\left(\widetilde{\pi}(\widetilde{\mathbf{X}}_{n+1}) < az \mid \widetilde{\pi}(\widetilde{\mathbf{X}}_n) = ay\right), \qquad y, z > 0\,.$$

Proof. Substituting into (3.1) y by ay, and z by az, and Q by $\widetilde{Q}$, and finally rescaling the integrated variable w by aw, the result follows. ■

Remarks.

1. This proposition shows that, for theoretical purposes, an arbitrary simple slice sampler is equivalent to the one-dimensional simple slice sampler on the density $f(x) = \inf\{w > 0; Q(w) \le x\}$ for $x > 0$ (with $f(x) = 0$ for $x \le 0$), since such a density has the appropriate value for $Q(y)$. This is often a helpful way to think about slice samplers.
2. This proposition clearly also applies if f_0 is *not* uniform, provided we use the more general definition $Q(y) = \int\limits_{L(y)} f_0(\mathbf{z})d\mathbf{z}$ instead of the uniform-specific definition $Q(y) = m(L(y))$. However, it does require that we are in the simple slice sampler case $k = 1$; in general we will need a k-dimensional function Q to completely specify the slice sampler convergence properties in this case.

To continue, we define a partial ordering on $\mathbf{R}^d$ based on values of π. That is, we say that $\mathbf{x}_1 \preceq \mathbf{x}_2$ if and only if $\pi(\mathbf{x}_1) \le \pi(\mathbf{x}_2)$, and that $\mathbf{x}_1 \prec \mathbf{x}_2$ if and only if $\pi(\mathbf{x}_1) < \pi(\mathbf{x}_2)$. Now, recall (Daley, 1968) that a Markov chain $\mathbf{X}$ on a partially ordered space is said to be *stochastically monotone* if for all fixed $\mathbf{z}$, we have that $\mathbf{P}(\mathbf{X}_1 \preceq \mathbf{z}|\mathbf{X}_0 = \mathbf{x}_1) \ge \mathbf{P}(\mathbf{X}_1 \preceq \mathbf{z}|\mathbf{X}_0 = \mathbf{x}_2)$ whenever $\mathbf{x}_1 \preceq \mathbf{x}_2$, or equivalently that $\mathbf{P}(\mathbf{X}_1 \prec \mathbf{z}|\mathbf{X}_0 = \mathbf{x}_1) \ge \mathbf{P}(\mathbf{X}_1 \prec \mathbf{z}|\mathbf{X}_0 = \mathbf{x}_2)$ whenever $\mathbf{x}_1 \preceq \mathbf{x}_2$. (Stochastically monotone chains are usually easier to analyse than more general classes of chains.) We have

Proposition 2. *With the ordering on $\mathbf{R}^d$ given above, P_{ssl} is stochastically monotone.*

Proof. We see (as in the previous proof) that for $i = 1, 2$, setting $z = \pi(\mathbf{z})$, we have from (3.1)

$$\mathbf{P}(\mathbf{X}_1 \prec \mathbf{z} \,|\, \mathbf{X}_0 = \mathbf{x}_i) \;=\; \mathbf{P}(\pi(\mathbf{X}_1) < z \,|\, \mathbf{X}_0 = \mathbf{x}_i)$$

$$= \frac{1}{\pi(\mathbf{x}_i)} \int\limits_0^{\pi(\mathbf{x}_i)} \max\left(1 - \frac{Q(z)}{Q(w)},\ 0\right) dw\,,$$

i.e. is an average of the function $f(w) = \max\left(1 - \frac{Q(z)}{Q(w)},\ 0\right)$, averaged over the interval $[0,\ \pi(\mathbf{x}_i)]$. But clearly f is non-increasing. Hence, if $\pi(\mathbf{x}_1) \leq \pi(\mathbf{x}_2)$, then $\mathbf{P}(\mathbf{X}_1 \prec \mathbf{z} \,|\, \mathbf{X}_0 = \mathbf{x}_1) \geq \mathbf{P}(\mathbf{X}_1 \prec \mathbf{z} \,|\, \mathbf{X}_0 = \mathbf{x}_2)$, as required. ∎

Although the Markov chain $\{\pi(\mathbf{X}_n),\ n \in \mathbf{N}\}$ is a non-trivial simplification of $\{\mathbf{X}_n,\ n \in \mathbf{N}\}$, the convergence properties of the two chains are identical, since by the construction of the algorithm, the conditional distribution of $\mathbf{X}_n$ given that $\pi(\mathbf{X}_n) = y$ is uniformly distributed, for all $n \geq 1$.

To end this section, we mention a result which is presented in Mira and Tierney (1997), using a theorem of Peskun (1973; see also Tierney, 1995, Section 3). We therefore omit the proof.

Proposition 3. *Suppose π is bounded and supp(π) has finite Lebesgue measure. Then P_{ssl} is uniformly ergodic, with principal eigenvalue being bounded above by the rate of convergence of the independence sampler with uniform proposal distribution.*

4. Geometric ergodicity of slice samplers.

In this section, we consider the geometric ergodicity of slice samplers. We concentrate on the case P_{ssl}, i.e. on the case where f_0 is an indicator function and $k = 1$. For some of our results, we shall further assume that π is a *bounded* function. In that case, since the slice-sampler is scale-invariant (Proposition 1), it suffices to assume that $\pi \leq 1$, i.e. that π is bounded by 1.

Recall (see e.g. Nummelin, 1984; Meyn and Tweedie, 1993) that a Markov chain $P(x, \cdot)$ on a state space $\mathcal{X}$, having stationary distribution $\nu(\cdot)$, is *geometrically ergodic* if there is $\rho < 1$ and a ν-a.e.-finite function $V : \mathcal{X} \to [1, \infty]$, such that

$$\|\mathbf{P}(X_n \in \cdot \,|\, X_0 = x) - \nu(\cdot)\| \equiv \sup_{A \subseteq \mathcal{X}} |\mathbf{P}(X_n \in A \,|\, X_0 = x) - \nu(A)| \leq V(x)\rho^n, \qquad x \in \mathcal{X}.$$

Recall further that this is equivalent to the existence of a π-a.e. finite function $V : \mathcal{X} \to [1, \infty]$, a subset $C \subseteq \mathcal{X}$, a probability measure $\mu(\cdot)$ on $\mathcal{X}$, and constants $\epsilon > 0$, $\lambda < 1$, and $b < \infty$, such that (a) $P(x, \cdot) \geq \epsilon\mu(\cdot)$ for all $x \in C$ (i.e., the set C is *small*); and (b) $PV(x) \leq \lambda V(x) + b\mathbf{1}_C(x)$ for all $x \in \mathcal{X}$ (i.e., V satisfies a *drift condition*). We shall examine these two conditions separately.

Condition (a) above is fairly straightforward. Indeed, we have the following.

Proposition 4. *Consider the slice sampler P_{ssl} on a density π. For any fixed $y^* > y_* > 0$, define the subset $C \subseteq \mathbf{R}^d$ by*

$$C \;=\; \left\{\mathbf{x} \in \mathbf{R}^d \,;\, y_* \leq \pi(\mathbf{x}) \leq y^*\right\} .$$

Then we have

$$P_{ssl}(\mathbf{x}, \cdot) \geq \frac{y_*}{y^*}\mu(\cdot) , \qquad \mathbf{x} \in C ,$$

where

$$\mu(A) = y_*^{-1} \int_0^{y_*} \frac{m(A \cap L(y))}{Q(y)} dy \ .$$

That is, the set C is small with $\epsilon = y_*/y^*$. In particular, if π is bounded (without loss of generality by 1 say), then $L(y_*)$ is small with $\epsilon = y_*$.

Proof. If we start the slice sampler at some $\mathbf{X}_n \in C$, then we clearly have

$$\mathcal{L}(Y_{n+1} \,|\, \mathbf{X}_n) \geq \frac{y_*}{y^*}\,\mathcal{U}([0, y_*]).$$

But since $\mathcal{L}(\mathbf{X}_{n+1} \,|\, Y_{n+1}) = \mathcal{U}\left(L(Y_{n+1})\right)$, the result follows immediately. ∎

To continue, we need to establish a drift condition (i.e., condition (b) above) for P_{ssl}. This is somewhat more difficult. We shall need the following well known stochastic approximation result (the "FKG inequality"), which we state in a way relevant to our current context. Briefly, it states that if $\mathcal{M}_1$ has non-decreasing Radon-Nikodym derivative with respect to $\mathcal{M}_2$, then any non-decreasing function will have larger conditional expected value with respect to $\mathcal{M}_1$ than with respect to $\mathcal{M}_2$. For a similar application to conditional expectations, and discussion of the result, see Roberts (1991).

Lemma 5. *Suppose that $\mathcal{M}_1$ and $\mathcal{M}_2$ are two probability measures on* $\mathbf{R}$, *such that there is a version of the Radon-Nikodym derivative $R(x) = \mathcal{M}_2(dx)/\mathcal{M}_1(dx)$, which is a non-decreasing function. Suppose also that f is a non-decreasing function from* $\mathbf{R}$ *into* $\mathbf{R}^+$. *Let $\mathbf{E}_i$, $i = 1, 2$ denote expectations with respect to the two measures $\mathcal{M}_i$, $i = 1, 2$. Then for any set A for which the following conditional expectations exist,*

$$\mathbf{E}_1[f(X)|X \in A] \le \mathbf{E}_2[f(X)|X \in A] \, .$$

Using this lemma, we are now able to establish a drift condition for P_{ssl}.

Proposition 6. *Consider the slice sampler P_{ssl} on a density $\pi \le 1$. Suppose its corresponding function $Q(y) = m(L(y))$ is differentiable, and that there exists a constant $\alpha > 1$ such that $Q'(y) y^{1+\frac{1}{\alpha}}$ is non-increasing, at least for $y \le Y$. Then, for any β with $0 < \beta < \min\left(\frac{\alpha-1}{\alpha}, \frac{1}{\alpha}\right)$, and for any $y_* \in (0, Y)$, we have*

$$P_{ssl} V(\mathbf{x}) \ \le \ \lambda V(\mathbf{x}) + b \mathbf{1}_{L(y_*)}(\mathbf{x}) \, , \mathbf{x} \in \mathbf{R}^d \, ,$$

where $V(\mathbf{x}) = \pi(\mathbf{x})^{-\beta}$, and where

$$\lambda \ \equiv \ \frac{1}{(1-\beta)(1+\alpha\beta)} + \frac{\alpha\beta (y_*/Y)^\beta}{1+\alpha\beta}$$

and

$$b \ = \frac{Y^{-\beta}(1+\alpha\beta(1-\beta))}{(1-\beta)(1+\alpha\beta)} - \lambda \, .$$

(Since $(1-\beta)(1+\alpha\beta) > 1$ for $0 < \beta < \frac{\alpha-1}{\alpha}$, it follows that by choosing $y_* > 0$ sufficiently small, we can insure that $\lambda < 1$.) Furthermore, if $Y = 1$ then the formula for b may be simplified to

$$b \ = \ \frac{\alpha\beta(1-y_*^\beta)}{(1+\alpha\beta)}$$

Proof. We note that if $\mathbf{x} \in \mathbf{R}^d$ is such that $\pi(\mathbf{x}) \leq Y$, then

$$\begin{aligned}
P_{ssl}V(\mathbf{x}) &= \frac{1}{\pi(\mathbf{x})} \int_0^{\pi(\mathbf{x})} \frac{1}{Q(y)} \int_{L(y)} \pi(\mathbf{z})^{-\beta} d\mathbf{z}\ dy \\
&= \frac{1}{\pi(\mathbf{x})} \int_0^{\pi(\mathbf{x})} \frac{1}{Q(y)} \int_y^{\infty} w^{-\beta}(-Q'(w))dw\ dy \\
&= \frac{1}{\pi(\mathbf{x})} \int_0^{\pi(\mathbf{x})} \frac{\left(\int_y^Y + \int_Y^\infty\right) w^{-\beta}(-Q'(w))dw}{\left(\int_y^Y + \int_Y^\infty\right)(-Q'(w))dw} dy \\
&\leq \frac{1}{\pi(\mathbf{x})} \int_0^{\pi(\mathbf{x})} \frac{\int_y^Y w^{-\beta}(-Q'(w))dw}{\int_y^Y (-Q'(w))dw} dy \\
&\leq \frac{1}{\pi(\mathbf{x})} \int_0^{\pi(\mathbf{x})} \frac{\int_y^Y w^{-(1+\beta+\alpha^{-1})}dw}{\int_y^Y w^{-(1+\alpha^{-1})}dw} dy \\
&= \frac{1}{1+\alpha\beta} \frac{1}{\pi(\mathbf{x})} \left(\int_0^{\pi(\mathbf{x})} y^{-\beta} dy + \int_0^{\pi(\mathbf{x})} \frac{Y^{-\alpha^{-1}}(y^{-\beta} - Y^{-\beta})}{y^{-\alpha^{-1}} - Y^{-\alpha^{-1}}} dy \right) \\
&\leq \frac{V(\mathbf{x})}{(1-\beta)(1+\alpha\beta)} + \frac{\alpha\beta Y^{-\beta}}{1+\alpha\beta}
\end{aligned}$$

Here the first equality follows simply from writing out the definition of $PV(\mathbf{x})$, and the second equality then follows from rewriting the inner integral with respect to the measure $-Q'(w)dw$. The first inequality follows from the fact that $w^{-\beta}$ is a non-increasing function. The second inequality follows from Lemma 5 with $\mathcal{M}_2(dy) \propto y^{-(1+\alpha^{-1})}dy$ and $\mathcal{M}_1(dy) \propto (-Q'(y))dy$. The third inequality follows from the fact that $(y^{-\beta} - Y^{-\beta})/(y^{-\alpha^{-1}} - Y^{-\alpha^{-1}})$ is a non-decreasing function of $y \in (0, Y)$, at least when $\beta\alpha < 1$ as we've stipulated; this can be checked by differentiating with respect to y and then maximising over Y. Hence an upper bound for this function is obtained by taking the limit as $y \to Y$.

For $\pi(\mathbf{x}) \geq Y$, we note that by stochastic monotonicity (Proposition 2), it follows that $P_{ssl}V(\mathbf{x})$ is non-increasing according to the ordering $\preceq$ on $\mathbf{R}^d$. Therefore, if $\mathbf{x}$ is such that $\pi(\mathbf{x}) \geq Y$, then we must have $P_{ssl}V(\mathbf{x}) \leq P_{ssl}V(\mathbf{x}')$ where $\pi(\mathbf{x}') = Y$. Hence, from the above bound on $P_{ssl}V(\mathbf{x}')$, we have that

$$P_{ssl}V(\mathbf{x}) \leq Y^{-\beta} \frac{1+\alpha\beta(1-\beta)}{(1+\alpha\beta)(1-\beta)}, \qquad \pi(\mathbf{x}) \geq Y.$$

Now let λ and b be as in the statement of the proposition. Then it is easily verified (by considering separately the cases $\pi(\mathbf{x}) < y_*$, $y_* < \pi(\mathbf{x}) < Y$, and $\pi(\mathbf{x}) > Y$) that $P_{ssl}V(\mathbf{x}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_{L(y_*)}(\mathbf{x})$, as required.

The final statement of the proposition follows because, if $Y = 1$, then there is no case $\pi(\mathbf{x}) > Y$ to consider. Hence, in this case it is easily verified that we still have $P_{ssl}V(\mathbf{x}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_{L(y_*)}(\mathbf{x})$ with the new, simpler formula for b. ∎

Putting Propositions 4 and 6 together, and using the standard Markov chain theory discussed at the beginning of this section, we obtain

Theorem 7. *Consider the slice sampler P_{ssl} on a bounded density π. Suppose its corresponding function $Q(y) = m(L(y))$ is differentiable, and that there exists a constant $\alpha > 1$ such that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing, at least on an open set containing* 0. *Then P_{ssl} is geometrically ergodic.*

Remarks.

(1) These conditions are really rather weak. For instance, for $\mathcal{X} = \mathbf{R}$ the condition on $Q'(y)y^{1+\frac{1}{\alpha}}$ can be loosely stated as saying that π has tails that are at least as light as $x^{-\alpha}$. A couple of examples illuminate this.

 (i) Suppose that $\mathcal{X} = \mathbf{R}^+$ and that π is a positive continuous density. Suppose also that $\pi \propto e^{-\gamma x}$, at least in the right hand tail. Then for small y, $L(y) = (0, \log(y^{-1})/\gamma + \text{constant})$. Therefore $Q'(y)y^{1+\frac{1}{\alpha}} = -y^{\frac{1}{\alpha}}$ which is non-increasing for all values of α (because of the minus sign).

 (ii) Again suppose $\mathcal{X} = \mathbf{R}^+$ and that π is continuous and positive. Now suppose that $\pi \propto x^{-\delta}$, at least in the right hand tail. For small y, $L(y) = (0, y^{-\delta^{-1}} \times \text{constant})$. $Q'(y)y^{1+\frac{1}{\alpha}} \propto y^{\alpha^{-1}-\delta^{-1}}$, so the condition holds for $\alpha \leq \delta$.

(2) The existence of the derivative of Q has been assumed in this theorem. This condition can certainly be weakened slightly by expressing the key condition on $Q'(y)y^{1+\frac{1}{\alpha}}$ in terms of a suitable Radon-Nikodym derivative for the measure R defined by $R((a, b]) = Q(a) - Q(b)$.

(3) The condition $\beta < 1/\alpha$ will be slightly restrictive for us in Section 5, when we consider

quantitative bounds. Indeed, for exponentially-decreasing densities π we have that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing for any $\alpha > 0$, however Proposition 6 unfortunately does not allow us to use α larger than $1/\beta$. Now, it is possible to get around this restriction in the proof of that proposition; for example, if $\alpha\beta = M \in \mathbf{N}$ and $Y = 1$, then we can instead compute the integral $\int\limits_0^{\pi(\mathbf{x})} \frac{Y^{-\alpha^{-1}}(y^{-\beta}-Y^{-\beta})}{y^{-\alpha^{-1}}-Y^{-\alpha^{-1}}}dy$ exactly, by recalling that $\frac{y^{-\beta}-1}{y^{-\alpha^{-1}}-1} = 1+y^{-\alpha^{-1}}+y^{-2\alpha^{-1}}+\ldots+y^{-(M-1)\alpha^{-1}}$. It is not difficult to carry out these calculations; however, they do not appear to substantially improve the quantitative bounds that we study in Section 5. Therefore, we do not pursue this idea further.

Finally, we consider the case where $\pi(\cdot)$ is unbounded. In this case, we have $Q(y) > 0$ for arbitrarily large values of y, and it is important how quickly $Q(y) \to 0$ as $y \to \infty$. To examine this, we consider the function $Q^{-1}(w) \equiv \inf\{y > 0; Q(y) \geq w\}$. By applying Proposition 6 twice, we obtain the following.

Theorem 8. *Consider the slice sampler P_{ssl} on a density π. Suppose π is unbounded with infinite support, but that there exists a constant $\alpha > 1$ such that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing for y in an open set containing 0, and furthermore that $(Q^{-1})'(w)w^{1+\frac{1}{\alpha}}$ is non-increasing for w in an open set containing 0. Then P_{ssl} is geometrically ergodic.*

Proof. It is no longer true that $L(y)$ is small for any y, though by Proposition 4, sets on which π is bounded above and away from zero are still small. Geometric excursions into either tail ($\pi(X)$ close to 0 or ∞) are now possible. The tail $\pi(X) \approx 0$ can be dealt with as in Proposition 6, and an identical calculation deals with the tail $\pi(X) \approx \infty$ (using drift function $(Q(\pi(X)))^{-\beta}$). Therefore, by using Proposition 6 twice, we see that P_{ssl} has geometric drift away from any fixed neighbourhood of $\mathbf{X} = \infty$ and also away from any fixed neighbourhood of $\mathbf{X} = 0$. The result now follows similarly to Theorem 7. ■

5. Quantitative convergence bounds.

In this section we consider quantitative bounds on the convergence of P_{ssl} to its stationary distribution $\nu_\pi(\cdot)$. We recall that we have verified minorisation and drift conditions in the previous section. We further recall that we have verified that P_{ssl} is stochastically monotone (Proposition 2). From these ingredients, there are well-known quantitative bounds on the distance of $\mathcal{L}(\mathbf{X}_n)$ to stationarity. For optimal results, we use the following recent result of Roberts and Tweedie (1998), which builds on the analysis in Rosenthal (1995) and Lund and Tweedie (1996). For notation, we write $\mathbf{E}_{\nu_\pi \wedge \delta_{\mathbf{x}}}(V)$ for the expected value of V under the *stochastic minorant* (with respect to the ordering $\preceq$) of the stationary distribution $\nu_\pi(\cdot)$ and the point mass $\delta_{\mathbf{x}}(\cdot)$. That is,

$$\mathbf{E}_{\nu_\pi \wedge \delta_{\mathbf{x}}}(V) \;=\; V(\mathbf{x})\nu_\pi\{\cdot \succeq \mathbf{x}\} + \mathbf{E}_{\nu_\pi}\left(V\,\mathbf{1}_{\{\cdot \preceq \mathbf{x}\}}\right) ,$$

so that using the fact that $V \geq 1$ for the first inequality and Meyn and Tweedie, 1993, Proposition 4.3 (i) for the second:

$$\begin{aligned}
\mathbf{E}_{\nu_\pi \wedge \delta_{\mathbf{x}}}(V) &= \mathbf{E}_{\nu_\pi \wedge \delta_{\mathbf{x}}}(V-1) + 1 \\
&\leq (V(\mathbf{x}) - 1) + \mathbf{E}_\pi(V-1) + 1 \\
&= V(x) + \mathbf{E}_\pi(V) - 1 \\
&\leq V(\mathbf{x}) - 1 + \frac{b}{1-\lambda} .
\end{aligned} \tag{5.1}$$

Theorem 9. *Consider the slice sampler P_{ssl} on a density $\pi \leq 1$. Set $V(\mathbf{x}) = \pi(\mathbf{x})^{-\beta}$. Then for $n\log(\lambda^{-1}) > \log\left(\mathbf{E}_{\nu_\pi \wedge \delta_{\mathbf{x}}}(V)\right)$, we have*

$$\|P^n_{ssl}(x,\cdot) - \pi(\cdot)\| \;\equiv\; \sup_{A \subseteq \mathbf{R}^d} |P^n_{ssl}(x,A) - \pi(A)| \;\leq\; K(n+\eta-\xi)\rho^n .$$

Here

$$K = \frac{e\epsilon(1-\epsilon)^{-\xi/\eta}}{\eta},$$

$$\xi = \frac{\log\left(\mathbf{E}_{\nu_\pi \wedge \delta_x}(V)\right)}{\log(\lambda^{-1})}, \qquad \eta = \frac{\log\left(\frac{\lambda s + b - \epsilon}{\lambda(1-\epsilon)}\right)}{\log(\lambda^{-1})},$$

$s = y^{-\beta}$, and $\rho = (1-\epsilon)^{\eta^{-1}}$, where the values of ϵ, λ, and b are as in Propositions 4 and 6.

Proof. The result follows immediately from Roberts and Tweedie (1997), in light of Proposition 2. ■

Example 5.1. Let $\pi(x) = e^{-x}\mathbf{1}_{x>0}$ be the density of the exponential distribution $\mathbf{Exp}(1)$. We can take α as large as we like (provided that $\alpha\beta \leq 1$), and can set $Y = 1$. Now suppose for illustration that $\mathbf{E}_{\nu_\pi \wedge \delta_x}(V) \leq 3$, and that we choose $\beta = 0.1$, $\alpha = 1/\beta = 10$, and $\epsilon = y_* = 0.1$. Then from Proposition 6, we have $\lambda = 0.95272$ and $b = 0.102836$ (so that $b/(1-\lambda) = 2.17502$). The bound of Theorem 9 then applies. We compute that $K = 0.0548648$, $\eta = 6.97809$, $\xi = 22.6824$, $s = 1.25893$, and $\rho = 0.985015$. We thus obtain that, for $n \geq 23$,

$$\|P_{ssl}^n(x,\cdot) - \pi(\cdot)\| \;\leq\; 0.054865\,(0.985015)^n(n - 15.7043)\;.$$

For example, with $n = 530$, we obtain

$$\|P_{ssl}^{530}(x,\cdot) - \pi(\cdot)\| < 0.0095\;.$$

Hence, for this example, just 530 iterations suffices to make the total variation distance to stationarity *provably* less than 1% (a convergence criterion suggested in Cowles and Rosenthal, 1996). ■

Now, it follows immediately from Proposition 1 that this same bound applies when $\pi(x) = e^{-ax}$ is the (un-normalised) density of the exponential distribution $\mathbf{Exp}(a)$ for any $a > 0$, not just for $a = 1$. Specifically, writing the transition kernel of the simple slice sampler for $\mathbf{Exp}(a)$ as P_a, and letting $V_a(x) = e^{a\beta x}$, we have that

$$P_a V_a(x) \leq \lambda V_a(x) + b\mathbf{1}_{V_a(x) \leq y_*^{-\beta}} \tag{5.2}$$

with the parameters all as given in the above example. The bound from Theorem 9 follows directly therefore.

However, it is surprising that this same bound applies to *any* density π such that

$$yQ'(y) \;\text{ is non-increasing,} \tag{5.3}$$

as the following theorem shows. We give the result under the same conditions on initial conditions as in the previous example. Analogous results are clearly possible for all different initial conditions.

Theorem 10. *Let π be a bounded density such that its corresponding function $Q(y) = m(L(y))$ is differentiable, and satisfies (5.3). Assume as in the previous example that $\mathbf{E}_{\nu_\pi \wedge \delta_x}(V) \leq 3$. Then the simple slice sampler algorithm for π satisfies*

$$\|P_{ssl}^n(x,\cdot) - \pi(\cdot)\| \;\leq\; 0.054865\,(0.985015)^n(n - 15.7043)\;, \qquad n \geq 23\,.$$

Proof. By renormalising if necessary, we can (and do) assume that $\sup_{\mathbf{x}\in\mathcal{X}} \pi(\mathbf{x}) = 1$. The proof shall proceed by comparing the slice sampler for π, i.e. P_{ssl}, to the slice sampler for the $\mathbf{Exp}(1)$ distribution (as studied in the above example). To that end, let y_* and β be as given in the example, and define the function $V(\cdot) = \pi(\cdot)^{-\beta}$. By Proposition 4, the set $L(y_*)$ is small for P_{ssl}, with $\epsilon = y_*$. The proof will be complete if we can show that the drift equation $P_{ssl}V(\mathbf{x}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_{L(y_*)}(\mathbf{x})$ is satisfied by P_{ssl}, for the same values of λ and b as in the example.

Fix $\mathbf{x}$. We can write

$$P_{ssl}V(\mathbf{x}) = \frac{1}{\pi(\mathbf{x})}\int_0^{\pi(\mathbf{x})} \frac{\int_y^1 w^{-\beta}(-Q'(w))dw}{\int_y^1(-Q'(w))dw}\;dy$$

$$\leq \frac{1}{\pi(\mathbf{x})}\int_0^{\pi(\mathbf{x})} \frac{\int_y^1 w^{-\beta-1}dw}{\int_y^1 w^{-1}dw}\;dy = P_aV_a(z)\;.$$

where z and a are positive scalars related via $\pi(\mathbf{x}) = e^{-az}$. Here the inequality follows from Lemma 5 and (5.3). Hence from (5.2) it follows that $P_{ssl}V(\mathbf{x}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_{L(y_*)}(\mathbf{x})$ as required. ∎

This theorem leads to the question of what densities π give rise to functions $Q(y)$ such that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing, for some $\alpha > 1$. Note that, since $Q'(y) \leq 0$, if $yQ'(y)$ is itself non-increasing then this condition is satisfied for every $\alpha > 1$.

Observe that if $Q^{-1}(w)$ is a (one-dimensional) log-concave function, then it is easily checked that $yQ'(y)$ is in fact non-increasing. Indeed, this follows since $\frac{d}{dw} \log Q^{-1}(w)$ equals the reciprocal of $Q'(y)y$ evaluated at $y = Q^{-1}(w)$. Hence, if the former is non-increasing as a function of w, then the latter is non-decreasing as a function of w and therefore is non-increasing as a function of y (since $Q' \leq 0$).

However connections between the condition and more familiar Euclidean concepts are more complicated in higher dimensions. We give a condition which relates properties of π along one-dimensional rays from its mode, to the condition on $yQ'(y)$.

We assume without loss of generality that π has its mode at the origin. We let $S = \{\mathbf{x} \in \mathbf{R}^d; \|\mathbf{x}\| = 1\}$ be the usual L^2 unit $(d-1)$-sphere in $\mathbf{R}^d$. For $\theta \in S$ and $y > 0$, we let $D(y;\theta) = \sup\{t > 0; \pi(t\theta) \leq y\}$. Note that the condition we impose in Proposition 11 is sufficient to guarantee unimodality of π.

Proposition 11. *Let π be a d-dimensional density such that for all $\theta \in S$,*

$$yD(y;\theta)^{d-1}\frac{\partial}{\partial y}D(y;\theta) \text{ is a non-increasing function of } y > 0\ . \tag{5.4}$$

Then the corresponding Q-function satisfies that $yQ'(y)$ is non-increasing.

Proof. We can write

$$Q(y) \ = \ \int_S \frac{D(y;\theta)^d}{d} d\theta\,,$$

where $d\theta$ is $(d-1)$-dimensional Lebesgue measure on the (curved) space S.

For $d \geq 2$, we note that

$$yQ'(y) \ = \ \int_S yD(y;\theta)^{d-1}\frac{\partial}{\partial y}D(y;\theta)d\theta\,.$$

(Here the differentiation under the integral sign is justified by e.g. Folland, 1984, Theorem 2.27.) The result follows by the condition imposed on the integrand. ∎

Remarks.

1. In one dimension, (5.4) is weaker than log-concavity. However this is not the case when $d \geq 2$. On the other hand, it can be shown that for all log-concave densities, and for all choices of $\alpha > 0$, there exists a compact (and therefore small) set outside of which π satisfies that $y^{1+\frac{1}{\alpha}} Q'(y)$ is a non-decreasing function. The results of Proposition 6 and Theorem 7 therefore apply, and moreover quantitative results analogous to Theorem 10 are available.
2. Since the function Q completely specifies the slice sampler, and since Q is unaffected by isometries, it suffices that π be *isometric* to a function satisfying (5.4). That is, it suffices that there exists a mapping $T : \mathbf{R}^d \to \mathbf{R}^d$, which preserves d-dimensional Lebesgue measure, such that $\pi \circ T$ satisfies (5.4).

Putting the previous two results together (and allowing for isometries as in the previous remark), we obtain finally the following.

Theorem 12. *Suppose π is a d-dimensional density which is (isometric to) a function satisfying condition (5.4) above. Let P_{ssl} be the corresponding simple slice sampler for π. Then P_{ssl} is geometrically ergodic, and in fact*

$$\|P_{ssl}^n(x,\cdot) - \pi(\cdot)\| \leq 0.054865\,(0.985015)^n (n - 15.7043)\ , \qquad n \geq 23\,,$$

at least for all x such that $\mathbf{E}_{\nu_\pi \wedge \delta_\mathbf{x}}(V) \leq 3$.

In particular, this Theorem shows that for any density π satisfying (5.4), we have that $\|P_{ssl}^{530}(x,\cdot) - \pi(\cdot)\| < 0.0095$, i.e. that the simple slice sampler converges after 530 iterations, at least for starting points near the mode.

It is natural to ask how condition $\mathbf{E}_{\nu_\pi \wedge \delta_\mathbf{x}}(V) \leq 3$ translates to a more direct condition on x itself. From (5.1), and using the values for λ and b from Example 5.1 (which are also valid in the more general setting of Theorem 12), a straightforward calculation gives that we require

$$\frac{\pi(x)}{\pi(x_{\max})} \geq 0.0025$$

where $x_{\max}$ denotes the mode of π. In fact more careful explicit calculations using the exponential Example 5.1 together with a stochastic comparison argument can considerably reduce this restriction still further.

6. Product slice samplers.

In this section, we shall investigate the geometric ergodicity of product slice samplers. Suppose $\pi(\mathbf{x}) = f_1(\mathbf{x})f_2(\mathbf{x})\dots f_k(\mathbf{x})$. Recall that the product slice sampler P_{psl} on $(\mathbf{X}, Y_1, Y_2, \dots, Y_k) \in \mathbf{R}^d \times \mathbf{R} \times \dots \times \mathbf{R}$ proceeds, given $\mathbf{X}_n$, by updating $Y_{n+1,i} \sim \mathcal{U}[0, f_i(\mathbf{X}_n)]$ for $1 \leq i \leq k$ conditionally independently, and then updating $\mathbf{X}_{n+1} \sim \mathcal{U}(L(\mathbf{Y}))$, where $L(\mathbf{Y}) = L(Y_1; f_1) \cap \dots \cap L(Y_k; f_k)$ (here $L(y; f) = \{\mathbf{x} \in \mathbf{R}^d; f(\mathbf{x}) \geq y\}$). We let $Q(\mathbf{y})$ denote $m(L(\mathbf{y}))$, where m is d-dimensional Lebesgue measure.

Before we give our first result about geometric ergodicity of the product slice sampler, we need the following lemma. The hypothesis of this lemma states, roughly, that all of the functions f_i are decreasing in the same direction.

Lemma 13. *Suppose there exists $Y > 0$ such that for all $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $f_1(\mathbf{x}_1) \leq f_1(\mathbf{x}_2) \leq Y$, we have*

$$f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2)\,, \qquad 2 \leq i \leq k\,. \tag{6.1}$$

Then there exists a function $c: \mathbf{R}^d \to \mathbf{R}^+$ such that $c(\mathbf{y}) \geq y_1$ for all $\mathbf{y}$ and

$$L(\mathbf{y}) \cap L(Y, f_1)^c = \{\mathbf{z};\ c(\mathbf{y}) \leq f_1(\mathbf{z}) < Y\}\ . \tag{6.2}$$

Proof. Denote the set on the left hand side of (6.2) by S. Now suppose there exists $\mathbf{z}_1 \in S$ and $\mathbf{z}_2 \in S^c$ with $f_1(\mathbf{z}_1) \leq f_1(\mathbf{z}_2) < Y$. Then it follows from (6.1) that $f_i(\mathbf{z}_1) \leq f_i(\mathbf{z}_2)$, $2 \leq i \leq k$. Since $\mathbf{z}_1 \in S$, $f_i(\mathbf{z}_1) \geq y_i$, $1 \leq i \leq k$ so that $f_i(\mathbf{z}_2) \geq y_i$, $1 \leq i \leq k$ and so $\mathbf{z}_2 \in S$ for a contradiction. Therefore S can be expressed as an interval such as the right hand side of (6.2). The constraint on c follows from $L(\mathbf{y}) \subseteq L(y_1, f_1)$. The result is therefore proved. ∎

We now prove a result about the geometric ergodicity of product slice samplers. Like the lemma, it requires that the functions f_i all be decreasing in the same direction.

Theorem 14. *Suppose that for each i, f_i is bounded. Set $Q_1(y) = m(L(y; f_1))$, and suppose that Q_1 is differentiable with $Q_1'(y)y^{1+\alpha^{-1}}$ non-increasing, at least in some open set containing 0. Suppose that for all $\epsilon > 0$, the set $\{\mathbf{z} : \ f_1(\mathbf{z}) \geq \epsilon\}$ is compact, and for each $1 \leq i \leq k$, the function f_i is bounded away from zero on compact intervals. Finally, suppose that (6.1) holds for the functions $\{f_i\}$. Then the product slice sampler P_{psl} is geometrically ergodic.*

Proof. We shall assume without loss of generality that we take Y small enough so that $Q_1'(y)y^{1+\alpha^{-1}}$ is non-increasing on $(0, Y)$. Set $V(\mathbf{x}) = f_1(\mathbf{x})^{-\beta}$. Choose $\mathbf{x}$ such that $f_1(\mathbf{x}) < Y$. Then

$$PV(\mathbf{x}) = \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \int_0^{f_1(\mathbf{x})} \cdots \int_0^{f_k(\mathbf{x})} \frac{1}{Q(\mathbf{y})} \int_{L(\mathbf{y})} f_1(\mathbf{z})^{-\beta} d\mathbf{z} \; d\mathbf{y} \ .$$

Now partition $L(\mathbf{y}) = A(\mathbf{y}) \cup B(\mathbf{y})$, where $A(\mathbf{y}) = L(\mathbf{y}) \cap L(Y, f_1)$, and $B(\mathbf{y}) = L(\mathbf{y}) \cap L(Y, f_1)^c$. Now $V(\mathbf{z})$ is greater than or less than or equal to $Y^{-\beta}$ according as $\mathbf{z}$ is in $B(\mathbf{y})$ or $A(\mathbf{y})$ respectively. Therefore we can write

$$\begin{aligned}
PV(x) &= \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_0^{f_1(\mathbf{x})} \cdots \int_0^{f_k(\mathbf{x})} \frac{1}{Q(\mathbf{y})} \left(\int_{A(\mathbf{y})} + \int_{B(\mathbf{y})} \right) f_1(\mathbf{z})^{-\beta} d\mathbf{z} \; d\mathbf{y} \ , \\
&\leq \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_0^{f_1(\mathbf{x})} \cdots \int_0^{f_k(\mathbf{x})} \frac{1}{m(B(\mathbf{y}))} \int_{B(\mathbf{y})} f_1(\mathbf{z})^{-\beta} d\mathbf{z} \; d\mathbf{y} \ , \\
&\leq \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_0^{f_1(\mathbf{x})} \cdots \int_0^{f_k(\mathbf{x})} \frac{1}{m(L(y_1, f_1))} \int_{L(y_1, f_1)} f_1(\mathbf{z})^{-\beta} d\mathbf{z} \; d\mathbf{y} \ , \\
&= \frac{1}{f_1(\mathbf{x})} \int_0^{f_1(\mathbf{x})} \frac{1}{m(L(y_1, f_1))} \int_{L(y_1, f_1)} f_1(\mathbf{z})^{-\beta} d\mathbf{z} \; dy_1 \ . \qquad (6.3)
\end{aligned}$$

The first inequality in the above is a straightforward application of the FKG inequality (Lemma 5); and the equality and the second inequality both follow from (6.2) and Lemma 5 again. The expression in the right hand side of (6.3) has therefore been reduced to the form of the expressions manipulated in Proposition 6. It follows therefore that

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

(at least for appropriate choices of β).

Geometric ergodicity will follow if we can demonstrate that all compact sets are small (cf. arguments in Roberts and Tweedie, 1996). To see this, note that the transition density of the Markov chain $\{\mathbf{X}_n,\ n \in \mathbf{Z}^+\}$, $p(\mathbf{x}, \mathbf{z})$ say, can be written

$$p(\mathbf{x}, \mathbf{z}) = \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_0^{f_1(\mathbf{x})} \cdots \int_0^{f_k(\mathbf{x})} \frac{1}{Q(\mathbf{y})} \mathbf{1}_{L(\mathbf{y})}(\mathbf{z}) d\mathbf{y} \ .$$

Now suppose that C is a compact set and ϵ and M are positive constants with $\epsilon \le f_i(\mathbf{w}) \le M$ for $1 \le i \le k$, $\mathbf{w} \in C$. (The existence of these constants is guaranteed by hypothesis.) Then for $\mathbf{x}, \mathbf{z} \in C$

$$\begin{aligned} p(\mathbf{x}, \mathbf{z}) &\ge \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_{\epsilon/2}^{f_1(\mathbf{x})} \cdots \int_{\epsilon/2}^{f_k(\mathbf{x})} \frac{1}{Q(\mathbf{y})} \mathbf{1}_{L(\mathbf{y})}(\mathbf{z}) d\mathbf{y} \\ &= \frac{1}{\prod_{i=1}^k f_i(\mathbf{x})} \times \int_{\epsilon/2}^{f_1(\mathbf{x})} \cdots \int_{\epsilon/2}^{f_k(\mathbf{x})} \frac{1}{Q(\mathbf{y})} d\mathbf{y} \\ &\ge \left(\frac{\epsilon}{2M}\right)^k \frac{1}{m(L(\epsilon/2, f_1))} \ > 0 \ . \end{aligned}$$

Therefore all compact sets are small and geometric ergodicity follows. ■

7. Opposite slice samplers

Finally, we consider product slice samplers whose component functions f_i are *not* all decreasing in the same direction. For simplicity, we restrict ourselves to dimension $d = 1$, and to a number of component functions $k = 2$ which are decreasing in *opposite* directions.

Specifically, let $\mathcal{X} \subset \mathbf{R}$, and $\pi(x) = f_1(x) f_2(x)$, where f_1 is a non-decreasing function and f_2 is a non-increasing function. Then we shall call this special form of the product slice sampler the opposite monotone sampler with transitions P_{oms}. We shall assume that f_1 and f_2 are invertible, so that we can write P_{oms} as follows. Given X_n, sample $Y_{n+1,i}$ from $\mathcal{U}(0, f_i(X_n))$ conditionally independently for $i = 1, 2$. X_{n+1} is then sampled from $\mathcal{U}(f_1^{-1}(Y_{n+1,1}), f_2^{-1}(Y_{n+1,2}))$.

Although in general the product slice sampler is not stochastically monotone, P_{oms} regains monotonicity properties from the total-orderedness of $\mathbf{R}$. Specifically we have

Proposition 15. *P_{oms} is stochastically monotone with respect to the usual ordering on* $\mathbf{R}$.

Proof. Given arbitrary $x_1 \leq x_2$, it is enough to show that there is a joint probability construction of two processes, one started at each of x_1 and x_2, which almost surely preserves their order. However, given U_1, U_2 and U_3, all independently $\mathcal{U}(0,1)$, we can produce the construction as follows. Start the two processes off at $X_0^j = x_j$, $j = 1,2$. Let $Y_i^j = f_i(x_j)\,U_i$, $i,j = 1,2$ (so that j indexes the two processes, and i continues to index the auxiliary variables). Now set $X_1^j = f_1^{-1}(Y_1^j) + (f_2^{-1}(Y_2^j) - f_1^{-1}(Y_1^j))U_3$. Now by the respective monotonicity of f_1 and f_2 it follows that $f_i^{-1}(Y_i^1) \leq f_i^{-1}(Y_i^2)$, $i = 1,2$. Therefore $X_1^1 \leq X_1^2$ and so the result follows. ∎

We turn now to the problem of proving the geometric ergodicity of P_{oms}. The interesting case for P_{oms} is the case where one (or both) the functions f_1, f_2 are unbounded, though π is still bounded. The case of bounded f_i is virtually identical to the case of P_{ssl} and we omit any formal statement of the result except to note that a very weak decay condition on the f_i's will be needed as in Proposition 7. Instead we shall deal with the case where both f_1 and f_2 are unbounded and non-zero.

Theorem 16. *Suppose $\mathcal{X}$ is a (possibly infinite) interval, $(\mathcal{X}_-, \mathcal{X}_+) \subset \mathbf{R}$, and that f_1 and f_2 are unbounded and non-zero on $\mathcal{X}$, with f_1 increasing and f_2 decreasing. Let β be a positive constant such that f_1^β and f_2^β are convex functions. Suppose there exists $0 < \gamma < (1+2\beta)^{-1}$ such that $u^\gamma f_1 f_2^{-1}(u)$ and $u^\gamma f_2 f_1^{-1}(u)$ are both non-decreasing functions for u in some neighbourhood of 0. Then P_{oms} is geometrically ergodic.*

Proof. Let $V_1(x) = f_1(x)^\beta$. Suppose k_1 is such that $f_1 f_2^{-1}(u)$ is non-decreasing for

$u \le f_2(k_1)$. Then for $x \ge k_1$,

$$
\begin{aligned}
P_{oms}V_1(x) &= \frac{1}{\pi(x)} \int_0^{f_1(x)} \int_0^{f_2(x)} \frac{1}{f_2^{-1}(y_2) - f_1^{-1}(y_1)} \int_{f_1^{-1}(y_1)}^{f_2^{-1}(y_2)} V_1(z)\ dz\ dy_1\ dy_2 \\
&= \int_0^1 \int_0^1 \frac{1}{f_2^{-1}(f_2(x)u_2) - f_1^{-1}(f_1(x)u_1)} \int_{f_1^{-1}(f_1(x)u_1)}^{f_2^{-1}(f_2(x)u_2)} f_1(z)^\beta\ dz\ du_1\ du_2 \\
&\le \frac{1}{2} \int_0^1 \int_0^1 \left[f_1(x)^\beta u_1^\beta + f_1(f_2^{-1}(u_2 f_2(x)))^\beta du_1\ du_2 \right] \\
&= \frac{f_1(x)^\beta}{2(1+\beta)} + \frac{1}{2} \int_0^1 f_1(f_2^{-1}(u_2 f_2(x)))^\beta du_2\ ,
\end{aligned}
\tag{7.1}
$$

the inequality following from the convexity condition on f_1^β. Now,

$$
f_1(f_2^{-1}(v)) \le (f_2(x))^\gamma f_1(x)/v^\gamma
$$

for $v \le f_2(x)$, so that the second term in (7.1) may be bounded by

$$
\frac{1}{2} \int_0^1 \frac{f_1(x)^\beta}{u^{\beta\gamma}} du \ = \ \frac{f_1(x)^\beta}{2(1-\beta\gamma)}
$$

since $\gamma < (1+2\beta)^{-1}$ implies that $\beta\gamma < 1$. Hence, for $x \ge k_1$,

$$
P_{oms}V_1(x) \le \frac{V_1(x)}{2} \left(\frac{1}{1+\beta} + \frac{1}{1-\beta\gamma} \right) \equiv \lambda V_1(x)
$$

say, where $\lambda < 1$, because $\gamma < (1+2\beta)^{-1}$. Furthermore, by stochastic monotonicity, $P_{oms}V_1(x) \le \lambda V_1(k_1)$ for $x \le k_1$.

Similarly we can prove that if $V_2(x) = f_2(x)^\beta$, there exists k_2 such that $P_{oms}V_2(x) \le \lambda V_2(x)$ for $x \le k_2$ with $P_{oms}V_2(x) \le \lambda V_2(k_2)$ for $x \ge k_2$.

Geometric drift now follows with drift function $V(x) = V_1(x) + V_2(x)$. Indeed, from the above bounds on $P_{oms}V_1(x)$ and $P_{oms}V_2(x)$, it follows that for large enough $M > 0$, we will have $P_{oms}V(x) \le \lambda' V(x)$ whenever $|x| > M$, for some $\lambda' < 1$. Furthermore the set $[-M, M]$ is easily seen to be small for P_{oms}. Hence, the result follows just as in Theorem 7. ∎

Unfortunately, in general, although P_{oms} is stochastically monotone, it is not possible to calculate bounds on convergence using Theorem 9 if $(\mathcal{X}_-, \mathcal{X}_+) = \mathbf{R}$ since it is not true that either $(-\infty, x)$ or (x, ∞) are small for any x. Computable bounds are still possible from the calculations in the proof of Theorem 16 (see Roberts and Tweedie, 1997) but will not be as tight as those in Theorem 10. However if either $\mathcal{X}_-$ or $\mathcal{X}_+$ are finite, then it will be possible to use the techniques of Theorem 9. The example below is an illustration of this.

Note that some of the estimates in the proof of Theorem 16 are fairly crude. Various refinements are possible and the conditions imposed on f_1 and f_2 can be significantly weakened, especially in more specific contexts. We do not pursue this in here, but content ourselves with a simple example to illustrate its application.

Example 7.1. Suppose we consider the Gamma density where $f_1(x) = x^\delta$ and $f_2(x) = e^{-x}$, both densities on $(0, \infty)$. We'll assume that $\delta > 0$. Now $f_1(f_2^{-1}(u)) = (\log u^{-1})^\delta$ and $f_2(f_1^{-1}(u)) = \exp\{-u^{1/\delta}\}$. It is easy to check that for all $\gamma > 0$, and for small enough u, both functions $u^\gamma f_1(f_2^{-1}(u))$ and $u^\gamma f_2(f_1^{-1}(u))$ are non-decreasing. Moreover, we can just take $\beta \geq \delta^{-1}$ to ensure convexity of $f_1(x)^\beta$ and $f_2(x)^\beta$. Therefore by Theorem 16 the algorithm is geometrically ergodic.

8. Discussion and conclusions.

In this paper, we have studied theoretical properties of slice samplers. We have shown that under some rather general hypotheses, these samplers have some very nice convergence properties.

In particular, we have proved geometric ergodicity for all simple slice samplers on densities with asymptotically polynomial tails. This covers virtually all distributions of interest. We have also extended this result to product slice samplers, albeit under more restrictive conditions.

We have also proved quantitative bounds on the convergence of these samplers, for certain classes of densities. In particular, for all multi-dimensional densities satisfying our condition (5.4) herein, which includes all one-dimensional log-concave densities, we have established a uniform bound of 530 iterations required to achieve 1% accuracy in total variation distance. Previous rigorous quantitative bounds for MCMC samplers have generally been established only for very specific models (Meyn and Tweedie, 1994; Rosenthal,

1995) or have involved large undetermined constants (Polson, 1996). Indeed, we know of no comparable result which gives a reasonable uniform bound on the convergence rate of a realistic sampling algorithm, over such a broad class of distributions.

Of course, it may not always be easy to implement a slice sampler for a particular problem. For example, the sets $L(\mathbf{y})$ and the measures $Q(\mathbf{y})$ may be difficult or impossible to compute. However, the results of this paper suggest that, if it is possible to run a slice sampler algorithm on a given density, then the sampler will probably have excellent convergence properties.

Appendix

As defined in Section 2, the f_0-slice sampler involves sampling from a density proportional to $f_0(\cdot)\mathbf{1}_{L(\mathbf{Y}_{n+1})}(\cdot)$. Throughout the paper, we have assumed that f_0 is constant thereby reducing all the simulations to uniform distributions on various shaped regions. In this appendix, we demonstrate that there is no loss of generality in doing this, since by a suitable transformation, the general f_0-slice sampler can be written in terms of the algorithms considered in the statements of our main results.

As a consequence of the following proposition therefore, all previous results in the paper have corresponding statements for the f_0-slice sampler.

Proposition 17. *Let $T:\ \mathbf{R}^d \to \mathbf{R}^d$ be a differentiable injective transformation. Let J be its Jacobian (assumed to be positive everywhere). Then $P_{f_0}(x, A) = P_{f_0/J}(T(x), T(A))$. That is, the f_0-slice sampler on $f_0(\mathbf{x})f_1(\mathbf{x})\dots f_k(\mathbf{x})$ behaves identically to the $(f_0 \circ T^{-1}/J \circ T^{-1})$-slice sampler on $\left(f_0(T^{-1}(\mathbf{x}))/J(T^{-1}(\mathbf{x}))\right) f_1(T^{-1}(\mathbf{x}))\dots f_k(T^{-1}(\mathbf{x}))$ (where we take $f_0(T^{-1}(\mathbf{x})) = 0$ if $\mathbf{x}$ is not in the range of T). Furthermore, it is always possible to find such a transformation T for which the quotient f_0/J is equal to the indicator function of a (possibly infinite) subset of $\mathbf{R}^d$.*

Proof. The first statement follows directly from the multi-dimensional change of variable formula (see e.g. Marsden, 1974, Section 9.3); specifically, sampling $T(\mathbf{x})$ from the density $\left(f_0(T^{-1}(\cdot))/J(T^{-1}(\cdot))\right) \mathbf{1}_{T(L(\mathbf{Y}))}(\cdot)$ is equivalent to sampling $\mathbf{x}$ from the density $f_0(\cdot)\mathbf{1}_{L(\mathbf{Y})}$.

For the final statement, we define $T_1(\mathbf{x}) = \int_0^{x_1} f_0(t, x_2, \dots, x_d)dt$. We further define $T_i(\mathbf{x}) = x_i$ for $i \geq 2$. We then set $T = (T_1, T_2, \dots, T_d)$. It is easily verified that this gives

$J(\mathbf{x}) = f_0(\mathbf{x})$, so that $f_0(T^{-1}(\mathbf{x}))/J(T^{-1}(\mathbf{x}))$ is equal to the indicator function of the range of T. ∎

Remark. We note that $T(\mathbf{R}^d)$ has finite Lebesgue measure if and only if $\int f_0(\mathbf{x})d\mathbf{x} < \infty$; in fact, in that case the Lebesgue measure of $T(\mathbf{R}^d)$ is precisely equal to $\int f_0(\mathbf{x})d\mathbf{x}$.

Acknowledgements. We thank Radford Neal for introducing us to slice samplers.

REFERENCES

Besag, J.E. and Green, P.J. (1993), Spatial statistics and Bayesian computation (with discussion). *J. Royal Stat. Soc. Ser. B* **55**, 25–38.

Cowles, M.K. and Rosenthal, J.S. (1996), A simulation approach to convergence rates for Markov chain Monte Carlo algorithms. Preprint.

Damien, P., Wakefield, J.C. and Walker, S. (1997), Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables. *J. Royal Stat. Soc., Series B*, to appear.

Daley, D. J. (1968), Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitsth. Verw Geb.*, **10**, 305–317.

Edwards, R.G. and Sokal, A.D. (1988), Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. Lett.* **38**, 2009–2012.

Fishman, G. (1996), An analysis of Swendsen-Wang and Related Sampling Methods. Preprint, University of North Carolina, Department of Operations Research.

Folland, G.B. (1984), Real analysis: Modern techniques and their applications. John Wiley & Sons, New York.

Higdon, D.M. (1996), Auxiliary variable methods for Markov chain Monte Carlo with applications. Preprint, Institute of Statistics and Decision Sciences, Duke University.

Lund, R.B. and Tweedie, R.L. (1996), Geometric convergence rates of stochastically ordered Markov chains. *Math. Oper. Research* **21**, 182–194.

Marsden, J.E. (1974), Elementary classical analysis. W.H. Freeman and Company, New York.

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

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

Mira, A. and Tierney, L. (1997), On the use of auxiliary variables in Markov chain Monte Carlo sampling. Preprint, School of Statistics, University of Minnesota.

Neal, R. (1997) Markov chain Monte Carlo methods based on 'slicing' the density function. Preprint.

Nummelin, E. (1984), General irreducible Markov chains and non-negative operators. Cambridge University Press.

Peskun, P.H. (1973), Optimum Monte Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.

Polson, N.G. (1996), Convergence of Markov chain Monte Carlo algorithms. In *Bayesian Statistics V*, 599-608, Clarendon press, Oxford.

Roberts, G.O. (1991), A comparison theorem for conditioned Markov processes. *J. Appl. Prob.* **28**, 74–83.

Roberts, G.O. and Rosenthal, J.S. (1997), On convergence rates of Gibbs samplers for uniform distributions. *Ann. Appl. Prob.*, to appear.

Roberts, G.O. and Tweedie, R.L. (1996), Geometric convergence and central limit theorems for multidimensional Hastings Metropolis algorithms. *Biometrika*, 83, 96–110.

Roberts, G.O. and Tweedie, R.L. (1998), Bounds on regeneration times and convergence rates for Markov chains. Preprint.

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

Swendsen, R.H. and Wang, J.S. (1987), Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58**, 86–88.

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

Tierney, L. (1995), A note on Metropolis Hastings kernels for general state spaces. *Ann. Appl. Prob.*, to appear.