The Polar Slice Sampler

by

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

(May, 1999; last revised November 9, 2001.)

This paper investigates the polar slice sampler, a particular type of the Markov chain Monte Carlo algorithm known as the slice sampler. This algorithm is shown to have convergence properties which under some circumstances are essentially independent of the dimension of the problem. For log-concave densities, the algorithm provably converges (from appropriate starting point) to within 0.01 of stationarity in total variation distance in a number of iterations given as a computable function of the spherical asymmetry of the density. In particular, for spherically symmetric log-concave densities, in arbitrary dimension, with appropriate starting point, we prove that the algorithm converges in at most 525 iterations. Simulations are done which confirm the polar slice sampler's excellent performance.

1. Introduction.

The slice sampler is a specialised type of Markov chain Monte Carlo (MCMC) auxiliary variable method (Swendsen and Wang, 1987; Edwards and Sokal, 1988; Besag and Green, 1993; Higdon, 1998) that has been recently popularised by Neal (1997), Fishman (1996), and Damien, Wakefield, and Walker (1999). In Roberts and Rosenthal (1999), it is demonstrated that the slice sampler has rather robust theoretical properties. In particular, the so-called *simple* slice sampler, with just a single auxiliary variable, is geometrically ergodic under very general conditions. This property contrasts strongly with other Metropolis-Hastings algorithms, for which geometric ergodicity typically depends strongly on the form of the target density (see for example Roberts and Tweedie, 1996).

* Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K. Internet: G.O.Roberts@lancaster.ac.uk.

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

The simple slice sampler's tractability to theoretical study stems from a stochastic monotonicity property. This allows good bounds on total variation distance from convergence to be obtained using the techniques developed in Roberts and Tweedie (1999, 1998), based on the work of Rosenthal (1995) and Lund and Tweedie (1996). Furthermore, convergence bounds can be given which are uniform over large classes of possible target density. Roberts and Rosenthal (1999) show that for any target distribution satisfying a particular condition on its level sets (restated as (9) below), convergence to within 0.01 in total variation distance is guaranteed in at most 530 iterations, for any starting values whose target density is at least 0.0025 of the modal density value. For other work which uses the monotonicity properties of the slice sampler, see the perfect simulation implementation in Mira, Møller, and Roberts (2001).

For applications, it is important to interpret the condition on the target density's level sets (equation (9) below) in terms of more recognisible criteria. Roberts and Rosenthal (1999) demonstrate that for one-dimensional distributions having density with respect to Lebesgue measure, the condition is implied by log-concavity of the density. This is an appealing result since target distributions produced from statistical applications often have this property. However, this result does not generalise beyond one dimension. More precisely, it is not true that the simple slice sampler on log-concave densities in higher dimensions satisfies the relevant condition (9) below.

This paper focuses closely on the relationship between the dimension of the target density and certain level set conditions. In particular, we analyse simple slice sampler factorisations on log-concave densities. The investigation leads to the introduction of the *polar slice sampler*, an algorithm that effectively preserves the robust convergence properties of the slice sampler for log-concave densities, for arbitrary dimensional distributions. Furthermore, a simple comparison of computing time shows the polar slice sampler performing well in comparison with the rejection sampler and the simple slice sampler.

A major limitation of all slice sampler algorithms is the fact that their implementation is often difficult. In particular, the $\mathbf{X}$-updating step often requires considerable additional effort to perform. However, we demonstrate here that the polar slice sampler has very good convergence properties, provided that the $\mathbf{X}$-updating step can be feasibly implemented.

Section 2 introduces the slice sampler and describes some if its basic properties, including those developed in Roberts and Rosenthal (1999). In Section 3, the performance of the *uniform* simple slice sampler (where each step samples a truncated uniform distribution) in high dimensions is investigated, and it is shown empirically that even for spherically symmetric, log-concave densities, the algorithm's performance deteriorates markedly as dimension increases. The polar slice sampler is motivated and constructed in Section 4, and studied theoretically in Section 5. Section 6 gives an empirical study of the performance of the polar slice sampler for a particular example, and gives comparisons with the uniform simple slice sampler and a corresponding rejection sampling algorithm. Section 7 offers some concluding remarks.

2. The slice sampler and its convergence properties.

Let $\pi : \mathbf{R}^d \to [0, \infty)$ be a non-negative function having positive finite integral. Our target distribution $\nu_\pi(\cdot)$ will be assumed to have density (with respect to d-dimensional Lebesgue measure dx) proportional to π, i.e.

$$\nu_\pi(A) \;=\; \frac{\int_A \pi(x)\, dx}{\int_{\mathbf{R}^d} \pi(x)\, dx}\,.$$

The simple slice sampler begins by choosing a factorisation of π, of the form

$$\pi(\mathbf{x}) = f_0(\mathbf{x})\, f_1(\mathbf{x})\,. \tag{1}$$

By renormalising π as necessary, we can (and do) assume without loss of generality that

$$\sup_{\mathbf{x}\in\mathbf{R}^d} f_1(\mathbf{x}) = 1\,. \tag{2}$$

The f_0-*simple slice sampler* proceeds as follows. Given $\mathbf{X}_n$, we sample a random variable Y_{n+1}, uniformly over the interval $(0, f_1(\mathbf{X}_n))$. We then sample $\mathbf{X}_{n+1}$ from the truncated probability distribution having density proportional to $f_0(\cdot)\, \mathbf{1}_{L(Y_{n+1})}(\cdot)$, where

$$L(y) \;=\; \left\{\mathbf{x} \in \mathbf{R}^d\,;\, f_1(\mathbf{x}) \geq y\right\}\,.$$

The key to the slice sampler is that the joint chain $(\mathbf{X}_n, Y_n)$ has stationary density proportional to $f_0(\mathbf{x})\, \mathbf{1}_{\{f_1(\mathbf{x})\geq y\}}$. Hence, the marginal stationary distribution of $\mathbf{X}_n$ has

density exactly proportional to π, the target density. In this way, the slice sampler can be used to generate an approximate sample from the target distribution. It is thus an example of a Markov chain Monte Carlo (MCMC) sampling algorithm.

If f_0 is a constant, then we shall call the sampler a *uniform simple slice sampler*. In particular, if $f_0 \equiv 1$, then the stationary distribution of $(\mathbf{X}_n, Y_n)$ is precisely the uniform distribution on the region (in $\mathbf{R}^d \times \mathbf{R}$) underneath the graph of the function π. (For related results about samplers for uniform distributions on more general regions, see Roberts and Rosenthal, 1998.)

In general, the transition density of the simple slice sampler is given by

$$\mathbf{P}(\mathbf{X}_{n+1} \in d\mathbf{z} \,|\, \mathbf{X}_n = \mathbf{x}) = \left[\frac{1}{f_1(\mathbf{x})} \int_0^{f_1(\mathbf{x})} \frac{f_0(\mathbf{z}) \mathbf{1}_{\{f_1(\mathbf{z}) \geq y\}}}{Q(y)} dy \right] d\mathbf{z}\,, \tag{3}$$

where

$$Q(y) \equiv Q_{f_0, f_1}(y) = \int_{\mathbf{R}^d} f_0(\mathbf{z}) \mathbf{1}_{\{f_1(\mathbf{z}) \geq y\}} d\mathbf{z}\,. \tag{4}$$

Remarks.

1. We note that, in fact, $Q(y)$ is the density (with respect to Lebesgue measure) of the stationary distribution of Y_n.
2. In the uniform case (f_0 = constant), the quantity $Q(y)$ coincides with the quantity $M(y)$ used to construct the *monotone rearrangement* of a function, as in Hardy et al. (1934), p. 276.

The issue of which factorisation to choose in (1) is often decided by practical considerations (though see Mira and Tierney, 2001, for some theoretical progress on this). In this paper we shall be mostly concerned with how this choice should be affected by the dimensionality of the target density.

The following result is central to all of our work here. To state it, recall that

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{z}) - \nu_\pi(\cdot)\| \;\equiv\; \sup_{A \subseteq \mathbf{R}^d} |\mathbf{P}(\mathbf{X}_n \in A \,|\, \mathbf{X}_0 = \mathbf{z}) - \nu_\pi(A)|$$

is the *total variation distance* to the stationary distribution ν_π after n steps (when starting from the point $\mathbf{z}$).

Proposition 1. *The simple slice sampler's convergence properties depend only on the function values $Q(y)$. Specifically, given a simple slice sampler $\{\mathbf{X}_n\}_{n=0}^{\infty}$ with corresponding function f_1, we have that*

(a) The sequence $\{f_1(\mathbf{X}_n)\}_{n=0}^{\infty}$ is itself a Markov chain, with transition probabilities that depend only on the values $Q(y)$.

(b) The conditional density of $\mathbf{X}_n$, given $\{f_1(\mathbf{X}_m)\}_{m=0}^{\infty}$, is proportional to

$$f_0(\mathbf{x})\,\mathbf{1}_{\{f_1(\mathbf{x})=f_1(\mathbf{X}_n)\}}\,. \tag{5}$$

In particular, it depends only on the value of $f_1(\mathbf{X}_n)$, and not on the other values $\{f_1(\mathbf{X}_m)\}_{m\neq n}$.

(c) Suppose $\{\mathbf{X}_n\}_{n=0}^{\infty}$ and $\{\widetilde{\mathbf{X}}_n\}_{n=0}^{\infty}$ are two different simple slice samplers (perhaps in different dimensions), with corresponding Q functions $Q(y)$ and $\widetilde{Q}(y)$, with corresponding f_1 functions $f_1(\mathbf{x})$ and $\widetilde{f}_1(\mathbf{x})$ (each satisfying (2)), and with corresponding stationary distributions $\nu_\pi(\mathbf{x})$ and $\nu_{\widetilde{\pi}}(\mathbf{x})$. Then if $Q(y) = \widetilde{Q}(y)$ for all y, and if $f_1(\mathbf{z}) = \widetilde{f}_1(\widetilde{\mathbf{z}})$ for some particular choice of $\mathbf{z}$ and $\widetilde{\mathbf{z}}$, then

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot\,|\,\mathbf{X}_0 = \mathbf{z}) - \nu_\pi(\cdot)\| \;=\; \|\mathbf{P}(\widetilde{\mathbf{X}}_n \in \cdot\,|\,\widetilde{\mathbf{X}}_0 = \widetilde{\mathbf{z}}) - \nu_{\widetilde{\pi}}(\cdot)\|\,. \tag{6}$$

Proof. We compute that

$$\begin{aligned}
&\mathbf{P}\left(f_1(\mathbf{X}_{n+1}) \le w \,|\, f_1(\mathbf{X}_0) = h_0, f_1(\mathbf{X}_1) = h_1, \ldots, f_1(\mathbf{X}_n) = h_n\right) \\
&\quad = \frac{1}{h_n}\int_0^{h_n} \frac{1}{Q(y)} \left(\int f_0(\mathbf{z})\mathbf{1}_{\{f_1(\mathbf{z})\ge y\}}\mathbf{1}_{\{f_1(\mathbf{z})\le w\}} d\mathbf{z}\right) dy \\
&\quad = \frac{1}{h_n}\int_0^{h_n} \frac{1}{Q(y)} \max\left(0,\; Q(y) - Q(w)\right) dy\,.
\end{aligned}$$

This immediately proves (a), and indeed gives an explicit formula for the transition probabilities.

Part (b) follows easily from the definition of the simple slice sampler. Indeed, by Fubini's Theorem, sampling $\mathbf{X}_{n+1}$ from the density $f_0(\cdot)\mathbf{1}_{L(Y_{n+1})(\cdot)}$ is equivalent to first sampling $f(\mathbf{X}_{n+1})$ from its induced distribution, and then sampling $\mathbf{X}_{n+1}$ from the conditional distribution (5).

In the language of Roberts and Rosenthal (2001a), part (b) implies that the chains $\{\mathbf{X}_n\}$ and $\{f_1(\mathbf{X}_n)\}$ are *bisufficient*. It thus follows from Corollary 2 of Roberts and Rosenthal (2001a) that

$$\|\mathbf{P}\big(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{z}\big) \,-\, \nu_\pi(\cdot)\| \;=\; \|\mathbf{P}\big(f_1(\mathbf{X}_n) \in \cdot \,|\, f_1(\mathbf{X}_0) = f_1(\mathbf{z})\big) \,-\, (f_1 * \nu_\pi)(\cdot)\|\,, \quad (7)$$

where $(f_1 * \nu_\pi)$ is the induced stationary distribution of $f_1(\mathbf{X}_n)$.

Now, if $\{\mathbf{X}_n\}_{n=0}^\infty$ and $\{\widetilde{\mathbf{X}}_n\}_{n=0}^\infty$ are two different distributions with identical Q functions and identical $f_1(\mathbf{X}_0)$ values, then it follows from part (a) that the chains $\{f_1(\mathbf{X}_n)\}_{n=0}^\infty$ and $\{\widetilde{f}_1(\widetilde{\mathbf{X}}_n)\}_{n=0}^\infty$ will be identically distributed. In particular, we have $\widetilde{f}_1(\widetilde{z}) = f_1(z)$, $\widetilde{f}_1 * \nu_{\widetilde{\pi}} = f_1 * \nu_\pi$, and

$$\begin{aligned}
&\|\mathbf{P}\Big(\widetilde{f}_1(\widetilde{\mathbf{X}}_n) \in \cdot \,|\, \widetilde{f}_1(\widetilde{\mathbf{X}}_0) = \widetilde{f}_1(\widetilde{\mathbf{z}})\Big) \,-\, (\widetilde{f}_1 * \nu_{\widetilde{\pi}})(\cdot)\| \\
&\qquad = \;\|\mathbf{P}\Big(f_1(\mathbf{X}_n) \in \cdot \,|\, f_1(\mathbf{X}_0) = f_1(\mathbf{z})\Big) \,-\, (f_1 * \nu_\pi)(\cdot)\|\,.
\end{aligned}$$

But then, applying (7) separately to each chain, we see that (6) follows immediately. ■

In light of Proposition 1, all conditions regarding the target density π can be stated solely in terms of the corresponding function Q. We shall at times require the following two conditions on Q:

$$yQ'(y) \text{ is non-increasing for } y \le Y \quad (8)$$

for some $0 < Y \le 1$; and

$$yQ'(y) \text{ is non-increasing for all } y\,. \quad (9)$$

The following result is proved in Roberts and Rosenthal (1999, Theorem 10).

Proposition 2. *If (9) holds, then for all* $\mathbf{x}$ *such that* $f_1(\mathbf{x})/\sup_{\mathbf{w}\in\mathbf{R}^d} f_1(\mathbf{w}) \ge 0.0025$ *and for all* $n \ge 530$,

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \le 0.01\ .$$

Remark. Note that we have assumed in (2), by renormalising as necessary, that $\sup_{\mathbf{w}\in\mathbf{R}^d} f_1(\mathbf{w}) = 1$. Hence, it is not strictly necessary to divide by this quantity in the statement of the Proposition; however we have done so to make the restriction on $\mathbf{x}$ clear, and to make the statement correct even without renormalising.

Thus 530 iterations "suffice" for convergence from a wide range of starting values, provided that (9) holds. It is shown in Roberts and Rosenthal (1999) that all one-dimensional log-concave densities satisfy (9), when f_0 is chosen to be constant. Unfortunately this property is not shared by higher dimensional log-concave densities, and the simulations below demonstrate that the performance of the simple slice sampler on log-concave densities deteriorates quickly as dimension increases.

3. The uniform simple slice sampler in higher dimensions.

To better understand the uniform simple slice sampler convergence properties, we consider on $\mathbf{R}^d$ the density (with respect to d-dimensional Lebesgue measure $d\mathbf{x}$) equal (up to a positive multiplicative constant) to $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ (where $|\mathbf{x}| = \sqrt{x_1^2 + \ldots + x_d^2}$ is the usual L^2-norm). This density is easily seen to be log-concave. However, we shall see that the uniform simple slice sampler performs quite poorly on this density, in high dimensions. In the next section, we will see that the poor performance can be overcome by a judicious choice of f_0.

To better understand the uniform simple slice sampler on the target density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ on $\mathbf{R}^d$, we note the following. By rotational symmetry, $Z = |\mathbf{X}|$ is itself a Markov chain, and indeed is precisely a slice sampler Markov chain on the one-dimensional target density

$$\eta(z) = z^{d-1}e^{-z}, \quad z > 0\,, \tag{10}$$

with corresponding slice factorisation given by $f_0(z) = z^{d-1}$, $f_1(z) = e^{-z}$. However, the further transformation $u = z^d$ makes this chain equivalent to the uniform simple slice sampler on the one-dimensional density $e^{-u^{1/d}}$. In other words, the d-dimensional uniform simple slice sampler on the density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ ($\mathbf{x} \in \mathbf{R}^d$) is equivalent to the one-dimensional uniform simple slice sampler on the density $\pi(u) = e^{-u^{1/d}}$ ($u > 0$).

Now, the mean of the distribution with density proportional to $e^{-u^{1/d}}$ is $(2d-1)!/(d-1)!$ (as can be seen by the substitution $t = u^{1/d}$), and the variance is $(3d-1)!/(d-1)! - ((2d-1)!/(d-1)!)^2$. For large d, this density is heavy tailed, and certainly not log-concave. This suggests that the uniform simple slice sampler may perform correspondingly badly. We shall see that this is indeed the case.

We ran the uniform simple slice sampler on the density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$, in $\mathbf{R}^d$ for different choices of dimension d. (In fact, we ran the equivalent one-dimensional uniform simple slice sampler on the density $\pi(u) = e^{-u^{1/d}}$ $(u > 0)$.) The logarithm of the chain value (i.e., $\log u$) was displayed. The results are summarised in Figure 1.

Figure 1. Some sample runs and some auto-correlation function (acf) plots (for the series $\log u$), for the uniform simple slice sampler on the density $e^{-u^{1/d}}$.

It is seen from Figure 1 that the auto-correlation function decreases more and more

slowly as the dimension increases. In particular, it does not appear to be bounded independent of dimension. The large auto-correlations even after a lag of 30 or 40 iterations, suggests that this sampler does not mix rapidly in higher dimensions.

Another way to explain why the uniform simple slice sampler performs worse as dimension increases is to analyse the Q function. Figure 2 shows the Q function and the derived function $yQ'(y)$. Only in dimension 1 is $yQ'(y)$ non-increasing, as would be required for the application of Proposition 2. As dimension increases, the function $yQ'(y)$ increases more rapidly.

Figure 2. The functions $Q(y)$ and $yQ'(y)$ for the spherically symmetric exponential example, in dimensions 1, 2, 5, and 10. $yQ'(y)$ is non-increasing only in dimension 1.

We thus see that the uniform simple slice sampler performs relatively poorly for this example, at least in high dimensions (i.e. with a large value of d). (For an interactive

simulation of this phenomenon over the internet, see Rosenthal, 1998.) We shall now consider a way of overcoming this difficulty.

4. The polar factorisation.

The above analysis suggests a remedy to the curse of dimensionality for the density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ ($\mathbf{x} \in \mathbf{R}^d$). Indeed, the function η given in (10) is itself log-concave, for any choice of $d \in \mathbf{N}$. Therefore, Proposition 2 can be applied to demonstrate rapid convergence of the uniform simple slice sampler on η, independently of dimension.

However, the one-dimensional uniform simple slice sampler on η corresponds precisely to a d-dimensional slice sampler on π with $f_0(\mathbf{x}) = |\mathbf{x}|^{-(d-1)}$ (and therefore with $f_1(\mathbf{x}) = |\mathbf{x}|^{d-1}\pi(\mathbf{x})$). This suggests that $f_0(\mathbf{x}) = |\mathbf{x}|^{-(d-1)}$ is a good choice of factorisation for this example.

We illustrate the situation schematically, as follows:

Figure 3. A schematic illustration of the uniform and polar factorings, for the target density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ in d dimensions, and their one-dimensional equivalents.

Thus, using the polar slice sampler on a d-dimensional log-concave target density is equivalent to a uniform simple slice sampler on a corresponding one-dimensional log-concave density (in our case, $z^{d-1}e^{-z}$). This allows us to apply the convergence results of Roberts and Rosenthal (1999) directly in this case.

In general, without spherical symmetry, $|\mathbf{X}|$ will not be Markov, so the above argument cannot be used directly. Moreover, the specific calculations of Roberts and Rosenthal (1999) will not be of use in this case. However, the argument still suggests the natural factorisation for the multidimensional case.

Thus, the polar factorisation amounts to setting

$$f_0(\mathbf{x}) = |\mathbf{x}|^{-(d-1)}\,, \tag{11}$$

and therefore setting $f_1(\mathbf{x}) = |\mathbf{x}|^{d-1}\pi(\mathbf{x})$, in the factorisation (1).

The polar slice sampler thus proceeds, given $\mathbf{X}_n$, by choosing $\mathbf{X}_{n+1}$ by the following two-step procedure:

(i) `sample` $Y_{n+1} \sim \mathbf{Unif}[0,\ |\mathbf{X}_n|^{d-1}\pi(\mathbf{X}_n)]$`;`

(ii) `sample` $\mathbf{X}_{n+1}$ `from the density proportional to` $|\mathbf{x}|^{-(d-1)}\mathbf{1}_{|\mathbf{x}|^{d-1}\pi(\mathbf{x})\geq Y_{n+1}}$.

We have already seen that this polar slice sampler will perform well on the example density $\pi(\mathbf{x}) = e^{-|\mathbf{x}|}$ ($\mathbf{x} \in \mathbf{R}^d$). In the next section, we shall see that this sampler has generally good theoretical convergence properties. In Section 6, we shall present numerical evidence that this sampler performs very well for particular choices of π.

Remark. Step (ii) above can often be performed by rejection sampling, working in polar coordinates. Indeed, if we write $\mathbf{X}_{n+1} = R_{n+1}\theta_{n+1}$, where $R_{n+1} \geq 0$ and where θ_{n+1} is on the unit sphere in $\mathbf{R}^d$ (i.e. $|\theta_{n+1}| = 1$), then step (ii) can be carried out as follows:

(ii_a) `sample` $R_{n+1} \sim \mathbf{Unif}[0, R^*_{n+1}]$`, for some` $R^*_{n+1} \geq \sup\{|\mathbf{x}|\,;\ |\mathbf{x}|^{d-1}\pi(\mathbf{x}) \geq Y_{n+1}\}$`;`

(ii_b) `sample` θ_{n+1} `uniformly from the unit sphere in` $\mathbf{R}^d$ `(e.g. by setting`

$$\theta_{n+1} \;=\; \frac{(Z_{n+1,1}, Z_{n+1,2}, \ldots, Z_{n+1,d})}{\sqrt{Z^2_{n+1,1} + \ldots + Z^2_{n+1,d}}}$$

`where` $Z_{n+1,1}, Z_{n+1,2}, \ldots, Z_{n+1,d}$ `are i.i.d. standard normal);`

(ii_c) `set` $\mathbf{X}_{n+1} = R_{n+1}\theta_{n+1}$.

(ii_d) `if` $|\mathbf{X}_{n+1}|^{d-1}\pi(\mathbf{X}_{n+1}) \geq Y_{n+1}$ `then accept this choice of` $\mathbf{X}_{n+1}$`; otherwise return to step` (ii_a) `and repeat.`

Under this scheme, the factor $|\mathbf{x}|^{-(d-1)}$ of step (ii) emerges naturally as the Jacobian of the polar-coordinate transformation, and does not need to be mentioned explicitly. That is, it is automatically incorporated by the requirement in (ii_a) that the choice of R_{n+1} be made *uniformly* (rather than with density $|\mathbf{x}|^{d-1}$).

5. Theoretical analysis.

In this section, we investigate theoretical bounds on the rates of convergence of the polar slice sampler algorithms. To do this, we require the following result from Roberts and Rosenthal (1999, Theorem 9). [This result uses general quantitative bound results developed in Roberts and Tweedie (1999, 1998), based on the work of Rosenthal (1995) and Lund and Tweedie (1996).] Recall also that the renormalisation condition (2), which states that $\sup_{\mathbf{x}\in\mathbf{R}^d} f(\mathbf{x}) = 1$, is assumed throughout.

Proposition 3. *Consider the simple slice sampler for the target density $\pi(\mathbf{x})$, with any factorisation $\pi(\mathbf{x}) = f_0(\mathbf{x})\, f_1(\mathbf{x})$. If (8) holds for some $0 < Y \leq 1$, then for all $\mathbf{x} \in \mathbf{R}^d$, and for all $n \geq \xi$, we have*

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \;\leq\; K(n + \eta - \xi)\rho^n \;.$$

Here

$$K = \frac{e\, y_*(1-y_*)^{-\xi/\eta}}{\eta} \qquad (\text{where } e = 2.71828\ldots)\,,$$

$$\xi = \frac{\log\left(f_1(\mathbf{x})^{-\beta} + \frac{b}{1-\lambda} - 1\right)}{\log(\lambda^{-1})}, \qquad \eta = \frac{\log\left(\frac{\lambda s + b - y_*}{\lambda(1-y_*)}\right)}{\log(\lambda^{-1})}\,,$$

$s = y^{-\beta}$, and $\rho = (1-y_*)^{\eta^{-1}}$. Furthermore,

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

and

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

The quantities α, β, and y_* may be chosen freely, provided that $\alpha > 1$, that $0 < \beta < \min\left(\frac{\alpha-1}{\alpha},\ \frac{1}{\alpha}\right)$, and that $y_* \in (0, Y)$.

This result has many parameters in it, making it difficult to understand. To simplify it, we shall normally chose $y_* = Y/10$, and shall choose $\beta = 0.1$ and $\alpha = 10$. Furthermore, we shall generally assume that $f_1(\mathbf{x}) \geq 0.01$, i.e. that we begin the sampler at a point whose f_1 density value is at least 1% of the modal value. (Note that *some* restriction on the initial value $f_1(\mathbf{x})$ is unavoidable, since the slice sampler will not in general be *uniformly* ergodic so that any upper bound on its time to stationarity will necessarily depend upon the starting point.)

Using these choices, and given any value of Y with $0 < Y \leq 1$, we can define an integer $n_*(Y)$ to be a value such that $\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \leq 0.01$ whenever $n \geq n_*(Y)$ (assuming that $f_1(\mathbf{x}) \geq 0.01$ and that (8) is satisfied for this value of Y). That is, $n_*(Y)$ represents a number of iterations that is sufficient for the sampler to have converged to within 0.01 of its stationary distribution in total variation distance (a convergence criterion suggested in Cowles and Rosenthal, 1998).

Thus, using Proposition 3, we compute convergence times $n_*(Y)$, as a function of Y from (8), to be as follows:

Y	$n_*(Y)$
1.0	525
0.9	615
0.8	728
0.5	1,400
0.33	2,395
0.25	3,475
0.1	10,850
0.01	160,000
0.001	2,075,000

Table 1. Convergence times $n_*(Y)$ as a function of Y.

To make the meaning of $n_*(Y)$ completely clear, we record:

Corollary 4. *If (8) holds for some value of Y with $0 < Y \leq 1$ (i.e., if $yQ'(y)$ is non-increasing for $y \leq Y$), then for all $\mathbf{x}$ such that $f_1(\mathbf{x})/\sup_{\mathbf{w}\in\mathbf{R}^d} f_1(\mathbf{w}) \geq 0.01$ and for all $n \geq n_*(Y)$ with $n_*(Y)$ as in Table 1, we have that*

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \leq 0.01 \ .$$

Remark. Since $f_1(\mathbf{0}) = 0$, we see that it is *not* permitted to use a starting value $\mathbf{x}$ which is at or very close to the origin.

As can be seen from Table 1, the convergence times $n_*(Y)$ are quite "reasonable" (in the sense of generally being quite feasible to run on a computer) for values of Y which are close to 1. However, they become rather unreasonable as the value of Y decreases towards 0 (which is not surprising, since condition (8) gives less information for smaller values of Y). In particular, in the case $Y = 1$, we obtain:

Corollary 5. *If (9) holds (i.e., if $yQ'(y)$ is non-increasing for all values of y), then for all $\mathbf{x}$ such that $f_1(\mathbf{x})/\sup_{\mathbf{w}\in\mathbf{R}^d} f_1(\mathbf{w}) \geq 0.01$ and for all $n \geq 525$,*

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \leq 0.01 \ .$$

Thus 525 iterations "suffice" for convergence from a wide range of starting values, provided that (9) holds. (Note that the slight numerical difference between 525 iterations here, and 530 iterations in Proposition 2, is due to a slightly different restriction on initial point $\mathbf{x}$.)

Now, it is shown in Roberts and Rosenthal (1999) that, for the uniform simple slice sampler, all one-dimensional log-concave densities satisfy (9), when f_0 is chosen to be constant. Unfortunately this property is not shared by higher dimensional log-concave densities, and the simulations above demonstrate that the performance of the simple slice sampler on log-concave densities deteriorates quickly as dimension increases.

However, the polar slice sampler (i.e. choosing $f_0(\mathbf{x}) = |\mathbf{x}|^{-(d-1)}$ instead of choosing $f_0(\mathbf{x}) = \text{constant}$) shall still satisfy (8) for suitable choice of Y. Specifically, given a density π on $\mathbf{R}^d$, set $f_0(\mathbf{x}) = |x|^{-(d-1)}$ and $f_1(\mathbf{x}) = \pi(\mathbf{x})/f_0(\mathbf{x}) = |\mathbf{x}|^{d-1}\pi(\mathbf{x})$ as in (11), and define

$$M(f_1, \theta) \ = \ \sup\{f_1(t\theta)\,;\ t \geq 0\}$$

for each θ on the unit sphere in $\mathbf{R}^d$. Then define the *asymmetry parameter*

$$A(f_1) \;=\; \inf_\theta M(f_1,\theta) \;/\; \sup_\theta M(f_1,\theta)$$

(As in the remark following Proposition 2, we will have by (2) that $\sup_\theta M(f_1,\theta) = 1$, so it is not strictly necessary to divide by this quantity, but we do so for clarity.)

Lemma 6. *With the factorisation (11), if $\pi(\mathbf{x})$ is log-concave (at least along rays emanating from the origin), then (8) holds for $Y = A(f_1)$, i.e. $yQ'(y)$ is non-increasing for $y \le A(f_1)$.*

Proof. The Q-function in this case is given by:

$$Q(y) = \int_{\mathbf{R}^d} |\mathbf{z}|^{-(d-1)} \mathbf{1}_{\{f_1(\mathbf{z}) \ge y\}} d\mathbf{z}\,. \tag{12}$$

Let us do the polar coordinate transformation $\mathbf{z} \mapsto (r,\theta)$ (where $r = |\mathbf{z}| = \sqrt{z_1^2 + \ldots + z_d^2}$, and where $|\theta| = 1$). Then the Jacobian is given by $d\mathbf{z} = r^{d-1} dr d\theta$, where $d\theta$ is Lebesgue measure on the surface of the unit sphere in $\mathbf{R}^d$. Thus, from (12), we have

$$Q(y) = \int_{\mathbf{R}^d} \mathbf{1}_{\{f_1(r,\theta) \ge y\}} dr\, d\theta\,. \tag{13}$$

Now, let us define $R^+(y,\theta) = \sup\{r; f_1(r,\theta) \ge y\}$ and $R^-(y,\theta) = \inf\{r; f_1(r,\theta) \ge y\}$ to be the two radii, along the ray from the origin having direction vector θ, at which the function f_1 crosses the value y. Also, let us write $f_1'(r,\theta) = \frac{\partial}{\partial r} f_1(r,\theta)$; by log-concavity, this derivative exists for almost all (r,θ). In terms of these quantities, we compute using the inverse function theorem that, for $y \le A(f_1)$,

$$Q'(y) = \int \left[\frac{1}{f_1'(R^+(y,\theta),\theta)} - \frac{1}{f_1'(R^-(y,\theta),\theta)} \right] d\theta \tag{14}$$

Now, note that (by continuity, which holds by log-concavity), we have $f_1(R^+(y,\theta),\theta) = y$ and $f_1(R^-(y,\theta),\theta) = y$. Hence, we obtain that

$$yQ'(y) = \int \left[\frac{f_1(R^+(y,\theta),\theta)}{f_1'(R^+(y,\theta),\theta)} - \frac{f_1(R^-(y,\theta),\theta)}{f_1'(R^-(y,\theta),\theta)} \right] d\theta$$

$$= \int \left(1/\frac{\partial}{\partial r} (\log f_1)(R^+(y,\theta),\theta) \right) d\theta - \int \left(1/\frac{\partial}{\partial r} (\log f_1)(R^-(y,\theta),\theta) \right) d\theta \tag{15}$$

However, since π is log-concave, and since $|\mathbf{x}|^{d-1}$ is log-concave (at least along rays from the origin), it follows that $f_1(\mathbf{x}) = \pi(\mathbf{x})|\mathbf{x}|^{d-1}$ is also log-concave along rays from the origin. Hence, $\frac{\partial}{\partial r}(\log f_1)(r, \theta)$ is non-increasing as a function of r. Now, if y increases, then $R^+(y, \theta)$ decreases, so that $\frac{\partial}{\partial r}(\log f_1)(R^+(y, \theta), \theta)$ is non-decreasing as a function of y. Moreover, $R^-(y, \theta)$ increases as a function of y, so that $-\frac{\partial}{\partial r}(\log f_1)(R^-(y, \theta), \theta)$ is also non-decreasing as a function of y. Hence, taking reciprocals, we see that each of the integrals in (15) is non-increasing. We therefore conclude that $yQ'(y)$ is non-increasing in this case, as claimed. ∎

Remark. Intuitively, the polar factorisation (11) helps because it precisely cancels out the Jacobian term r^{d-1} when switching to polar coordinates. That is why the resulting formula (13) does not have any powers of r in it, thus simplifying the analysis and allowing $yQ'(y)$ to more easily be non-increasing.

Remark. It is quite subtle why the proof of Proposition 6 does not work for $y > A(f_1)$. In fact, the problem here is that $R^+(y, \theta)$ and $R^-(y, \theta)$ will be undefined (or, ∞ and $-\infty$, respectively) for some θ when $y > A(f_1)$. The integral in (14) should then be taken only over those θ such that $M(f_1, \theta) \geq y$, with the other θ not contributing. Since the integrand in (14) is negative, it follows that omitting certain θ values will actually *increase* the value of the integral. Thus, it may not be true that $yQ'(y)$ is non-increasing in this case.

This result allows us to state the following result.

Theorem 7. *Suppose that π is a d-dimensional density which is log-concave (at least along rays emanating from the origin). Set $f_1(\mathbf{x}) = |\mathbf{x}|^{d-1}\pi(\mathbf{x})$, and let $Y = A(f_1)$ be its asymmetry parameter as above. Then for any initial value $\mathbf{x}$ such that $f_1(\mathbf{x})/\sup_{\mathbf{w}\in\mathbf{R}^d} f_1(\mathbf{w}) \geq 0.01$, and for all $n \geq n_*(Y)$ (with $n_*(Y)$ as in Table 1), the polar slice sampler satisfies that*

$$\|\mathbf{P}(\mathbf{X}_n \in \cdot \,|\, \mathbf{X}_0 = \mathbf{x}) - \nu_\pi(\cdot)\| \leq 0.01 \ .$$

i.e. it will be within 1% of its target distribution after at most $n_*(Y)$ iterations.

Proof. Lemma 6 implies that condition (8) holds with $Y = A(f_1)$. Hence, the result follows from Corollary 4. ∎

Note in particular that the values $n_*(Y)$ depend only on the asymmetry parameter $A(f_1)$, and are otherwise completely independent of dimension.

Remark 8. *If $\pi(\mathbf{x})$ is spherically symmetric, then so is $f_1(\mathbf{x}) = |\mathbf{x}|^{d-1}\pi(\mathbf{x})$. Hence, $A(f_1) = 1$ in this case, so that Theorem 7 applies with $Y = 1$ and with $n_*(Y) = n_*(1) = 525$.*

6. A numerical example.

To illustrate the performance of the various sampling schemes that we have discussed, we consider the following example:

$$\pi(\mathbf{x}) = (x_1^2 + \ldots + x_d^2)\, e^{-\sum_{i=1}^d \alpha_i x_i^2/2}, \qquad \mathbf{x} \in \mathbf{R}^d.$$

We have considered this example for various values of the parameters $\alpha_1, \ldots, \alpha_d$:

[A] $\alpha_i = 1 + (i-1)/d, \quad 1 \le i \le d;$

[B] $\alpha_i = 1 + 4(i-1)/(d-1), \quad 1 \le i \le d;$

[C] $\alpha_i = 1 + 10(i-1)/(d-1), \quad 1 \le i \le d.$

(In the special case $d = 1$, for all three choices we simply take $\alpha_1 = 1$.) Note in particular that we always have $\alpha_i \ge 1$.

Note that these examples are *not* spherically symmetric. In fact, it is easily computed that $M(f_1, \theta)$ will be largest when θ is along the x_1-axis, and smallest when θ is along the x_d-axis. Since $\sup_{t>0} t^4 e^{-at^2/2} = 16e^{-2}/a^2$, we compute that we will have $A(f_1) = (1 + (d-1)/d)^{-2} \approx 0.25$ for case [A], and $A(f_1) = 4^{-2} = 0.0625$ for case [B], and $A(f_1) = 10^{-2} = 0.01$ for case [C]. Hence, in light of Corollary 4, we might expect to find that the polar slice sampler will converge quickest in case [A], and slowest in case [C].

For our simulations, it is necessary to compute a bounding value R^*_{n+1} in order to use rejection sampling as in step (ii_a) above. We do this rather crudely, by simply noting

that, for any $\delta > 0$, since $\delta s \leq e^{\delta s}$, we have

$$|\mathbf{x}|^2 \pi(\mathbf{x}) = |\mathbf{x}|^4 e^{-\sum_{i=1}^d \alpha_i x_i^2/2} \leq |\mathbf{x}|^4 e^{-\sum_{i=1}^d x_i^2/2} = (|\mathbf{x}|^2)^2 e^{-|\mathbf{x}|^2/2} = \delta^{-2} (\delta\, |\mathbf{x}|^2)^2 e^{-|\mathbf{x}|^2/2}$$

$$\leq \delta^{-2} e^{2\delta |\mathbf{x}|^2} e^{-|\mathbf{x}|^2/2} = \delta^{-2} e^{(2\delta - \frac{1}{2})|\mathbf{x}|^2} .$$

Hence, in step (ii') above, for any $0 < \delta < \frac{1}{4}$ we may simply solve for

$$Y_{n+1} = \delta^{-2} e^{(2\delta - \frac{1}{2})(R^*_{n+1})^2} ,$$

to obtain that

$$R^*_{n+1} = \sqrt{\frac{1}{2\delta - \frac{1}{2}} \log\left(\delta^2 Y_{n+1}\right)} .$$

An approximately optimal choice of δ, to make the algorithm run faster and mix faster, was determined by informal experimentation. (Typically, we used a value of δ around 0.05.)

For comparison purposes, we have compared the polar slice sampler to the uniform simple slice sampler, and also to an ordinary rejection sampling algorithm, and to a Random walk Metropolis algorithm with spherically symmetric full-dimensional Gaussian proposal. We examined the performance of all 4 methods for case [A] above. The uniform simple slice and rejection algorithms also required bounds involving δ, whose value was approximately optimised similarly to the above. The random walk Metropolis algorithm was considered as a benchmark standard MCMC method, and was optimally tuned to have acceptance rate of about 0.234 (cf. Roberts et al., 1997; Roberts and Rosenthal, 2001b).

The results are summarised in Table 2. We have presented the amount of CPU time required. We have also presented the larger of two different auto-correlation times, for the functional $|\mathbf{x}|^2$ and for the functional x_1^2 respectively. (Recall that the auto-correlation time for a functional g is defined by $ACT = 1 + 2\sum_{k=1}^{\infty} \mathrm{Corr}\left(g(\mathbf{X}_n),\, g(\mathbf{X}_{n+k})\right)$, where $\mathbf{X}_n$ is taken to be in stationarity.)

	Rejection			Uniform Slice			Polar Slice			RWM		
dim	CPU	ACT	C∗A	CPU	ACT	C∗A	CPU	ACT	C∗A	CPU	ACT	C∗A
1	0.14	1	0.14	0.11	1.56	0.17	0.06	1.53	0.091	0.03	8.01	0.24
2	0.22	1	0.22	0.15	1.79	0.27	0.12	1.36	0.163	0.04	13.26	0.53
10	2.9	1	2.9	0.99	8.90	14.4	1.47	1.18	1.73	0.09	31.8	2.9
20	34.4	1	34.4	8.32	16.88	140	5.1	1.35	6.9	0.14	98	13.8
40	3660	1	3660	–	–	–	31.0	1.48	49.9	0.25	188	47

Table 2. A comparison of the rejection sampler, the uniform simple slice sampler, and the polar slice sampler, all run on the target density $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$ with α_i as in case [A] above. Here CPU the processor time used per thousand simulated iterations, in seconds. ACT is the larger of the auto-correlation times for the functional x_1^2 and the functional $|\mathbf{x}|^2$. C∗A is the product of these two quantities, describing the time needed to obtain a sample having sampling variance equivalent to that of 1000 i.i.d. draws from π.

In Table 2, C∗A is a measure of the time required to acquire the equivalent of 1000 i.i.d. samples from the target distribution, according to the criterion of variance of ergodic estimates. As can be seen from Table 2, the ordinary rejection sampler takes much more CPU time to execute, especially in higher dimensions. This more than compensates for the fact that it (of course) has an auto-correlation time of 1, i.e. produces i.i.d. samples.

The uniform simple slice sampler executes reasonably quickly in moderate dimensions, with a reasonably small auto-correlation time. However, it takes longer and longer to run as the dimension increases, and indeed appears to get "stuck" and be unusable in dimension 40. Furthermore, its auto-correlation times, while fairly small, are clearly increasing with dimension.

By contrast, the polar slice sampler executes reasonably quickly, even in dimensions as high as 40. Furthermore, its auto-correlation times are extremely small, and do not appear to be growing with dimension. This is consistent with Corollary 4, which applies here with $Y = 0.25$, and which proves that the mixing time for the polar slice sampler in this example is bounded independent of dimension. (Of course, the *running time* of the

polar slice sampler will grow in dimension, due to rejections from the substep (ii_d) above.) We conclude from this that the polar slice sampler performs very well on this example.

The (optimally tuned) random walk Metropolis method is known to have a convergence time which is approximately linear in dimension but is relatively inexpensive to implement. As a result its performance is substantially inferior to that of the polar slice sampler as measured by its auto-correlation time. When computing time is taken into account, we see that the random walk Metropolis algorithm performs very similarly to the polar slice sampler in 40 dimensions, but is considerably worse in all smaller dimensions.

Note that auto-correlation times for various other functionals (including higher moments of x_1 and $\log|\mathbf{x}|$) were also investigated, in order to avoid over-optimistic assessment of the slice sampler algorithms. These auto-correlation times were no larger than those reported here.

We next consider the entire auto-correlation functions for the uniform simple slice sampler and the polar slice sampler. We have done this for two functionals, x_1^2 and $|\mathbf{x}|^2$, respectively. (Note that, by $\pm$ symmetry, the auto-correlations would be 0 for functionals of the form x_i.) The results are presented in Figures 4, 5, 6 and 7, respectively.

Figure 4. The auto-correlation function (acf) for the uniform simple slice sampler, using the functional x_1^2, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in case [A].

Figure 5. The auto-correlation function (acf) for the uniform simple slice sampler, using the functional $|\mathbf{x}|^2$, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in case [A].

Figure 6. The auto-correlation function (acf) for the polar slice sampler, using the functional x_1^2, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in case [A].

Figure 7. The auto-correlation function (acf) for the polar slice sampler, using the functional $|\mathbf{x}|^2$, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in case [A].

We see from Figure 4 that the uniform simple slice sampler appears to be mixing very well, for this example, based on the functional x_1^2. However, the functional $|\mathbf{x}|^2$, which captures the covariance structure more accurately, shows that the uniform simple slice sampler is in fact *not* mixing well in high dimensions (similar to the example in Section 3 above). Note that this highlights the dangers of monitoring one-dimensional summaries of Markov chain output for convergence assessment – an issue which has implications for convergence diagnostic procedures for MCMC in general (see e.g. Cowles and Carlin, 1996; Robert, 1996; Brooks and Roberts, 1998).

On the other hand, we see from Figures 6 and 7 that the *polar* slice sampler is mixing very well indeed, even in high dimensions. Furthermore its low auto-correlation values are consistent across different choices of functional (x_1^2, $|\mathbf{x}|^2$, and others not shown). This is consistent with Theorem 7 above, which suggests that the convergence time of the polar slice sampler should grow only with the asymmetry parameter $A(f_1)$, but should not otherwise grow with dimensionality.

This simulation thus confirms the extremely good performance of the polar slice sampler, in this example at least.

We have also done a comparison of the polar slice sampler in the different cases [A], [B], and [C] above, all in dimension 10. The results are presented in Figures 8 and 9.

Figure 8. A polar slice run showing the values x_1^2, and auto-correlation function (acf) using the functional x_1^2, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in cases [A], [B], and [C] respectively. The dimension is 10 throughout.

Figure 9. A polar slice run showing the values $|\mathbf{x}|^2$, and auto-correlation function (acf) using the functional $|\mathbf{x}|^2$, for the example $\pi(\mathbf{x}) = |\mathbf{x}|^2 e^{-\sum \alpha_i x_i^2/2}$, with the parameters as in cases [A], [B], and [C] respectively. The dimension is 10 throughout.

We see from Figures 8 and 9 that the polar slice sampler mixes very quickly in case [A], and more slowly in cases [B] and [C]. In light of Corollary 4, this is consistent with the observation that $A(f_1) \approx 0.25$ in case [A], that $A(f_1) = 0.0625$ in case [B], and that $A(f_1) = 0.01$ in case [C].

We thus observe that, as expected, the polar slice sampler performs very well on reasonably symmetric target densities, but less well as the target density becomes more and more asymmetric.

7. Conclusion.

Our simulation results, together with our theoretical investigations, suggest that the polar slice sampler is likely to perform well in a large variety of fairly high-dimensional problems. Even in 40 dimensions in our example, it took an optimally tuned random walk Metropolis algorithm to give comparable performance taking into account computation times. It is of course true that the implementation of the slice sampler will be more difficult in more complicated examples, however the method of choosing R_{n+1} uniformly, and then obtaining θ_{n+1} by rejection sampling, would appear to be promising in some cases at least.

The theoretical results show that if implementation is possible, convergence is rapid. For suitable classes of densities (e.g. log-concave), uniform control over convergence times is possible in a way that is surprisingly unaffected by the dimension of the problem. Of course implementation time of these algorithms is very much affected by dimension, although our simulations demonstrate that in our examples the computational cost (at least in moderate dimensional problems) is not prohibitive.

Although the focus of this paper has very much been on log-concave densities (motivated by their widespread occurrence in statistical applications), the techniques of this paper could equally well be applied to other classes of densities, for instance those with $y^{1+\frac{1}{\alpha}}Q'(y)$ non-increasing for some $\alpha > 0$. Bounds for this broader class of densities will necessarily be larger, of course.

The numerical comparisons in Section 6 between the polar slice sampler and some of its competitors (the uniform simple slice sampler, the rejection sampler, and a random-walk Metropolis algorithm) are quite favourable. Of course, it should be noted that other implementations (for example, adaptive rejection sampling, see e.g. Gilks and Wilde, 1992) may well be available in other contexts. Therefore, we have to be cautious of over-extrapolating our conclusions from this limited study.

However, overall we are reasonably optimistic that the polar slice sampler can be applied successfully to many different high-dimensional target distributions π, and can thus be used as a promising MCMC algorithm for understanding difficult probability distributions.

Acknowledgements. We thank the editor and two anonymous referees for many helpful comments and suggestions.

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.

Brooks, S.P. and Roberts, G.O. (1998), Convergence assessment techniques for Markov chain Monte Carlo. Stat. and Comput. **8**, 319–335.

Cowles, M.K. and Carlin, B.P. (1996), Markov chain Monte Carlo convergence diagnostics: A comparative review. J. Amer. Stat. Assoc. **91**, 883–904.

Cowles, M.K. and Rosenthal, J.S. (1998), A simulation approach to convergence rates for Markov chain Monte Carlo algorithms. Stat. and Comput. **8**, 115–124.

Damien, P., Wakefield, J.C., and Walker, S. (1999), Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables. J. Royal Stat. Soc. Ser. B **61**, 331–344.

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. Technical Report, Department of Operations Research, University of North Carolina.

Gilks, W.R. and Wilde, P. (1992), Adaptive rejection sampling for Gibbs sampler. Applied Statistics **41**, 337–348.

G.H. Hardy, J.E. Littlewood, and G. Pólya (1934), Inequalities. Cambridge University Press.

Higdon, D.M. (1998), Auxiliary variable methods for Markov chain Monte Carlo with applications. J. Amer. Stat. Assoc. **93**, 585–595.

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

Mira, A., Møller, J., and Roberts, G.O. (2001), Perfect slice samplers. J. Royal Stat. Soc. Ser. B **63** (2001), 593–606.

Mira, A. and Tierney, L. (2001), On the use of auxiliary variables in Markov chain Monte Carlo sampling. Scandinavian J. Stat., to appear.

Neal, R. (1997), Markov chain Monte Carlo methods based on 'slicing' the density function. Technical Report, Department of Statistics, University of Toronto.

Robert, C.P. (1996), Convergence assessments for Markov chain Monte Carlo methods. Stat. Sci. **10**, 231–253.

Roberts, G.O., Gelman, A. and Gilks, W.R. (1997), Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Applied Probability **7**, 110–120.

Roberts, G.O. and Rosenthal, J.S. (1998), On convergence rates of Gibbs samplers for uniform distributions. Ann. Appl. Prob. **8**, 1291–1302.

Roberts, G.O. and Rosenthal, J.S. (1999), Convergence of slice sampler Markov chains. J. Royal Stat. Soc., Series B, **61**, 643–660.

Roberts, G.O. and Rosenthal, J.S. (2001a), Markov chains and de-initialising processes. Scandinavian Journal of Statistics, **28**, 489–505.

Roberts, G.O. and Rosenthal, J.S. (2001b), Optimal scaling of various Metropolis-Hastings algorithms. To appear in Statistical Science.

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), Rates of convergence for stochastically monotone stochastic processes. Preprint.

Roberts, G.O. and Tweedie, R.L. (1999), Bounds on regeneration times and convergence rates for Markov chains. Stoch. Proc. Appl. **80**, 211–229.

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

Rosenthal, J.S. (1998), Slice Sampler Java Applet.

`http://utstat.toronto.edu/jeff/java/slice.html`

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