Statistical Science
2013, Vol. 28, No. 3, 360–375
DOI: 10.1214/13-STS423

Component-Wise Markov Chain Monte Carlo: Uniform and Geometric Ergodicity under Mixing and Composition

Alicia A. Johnson, Galin L. Jones and Ronald C. Neath

Abstract. It is common practice in Markov chain Monte Carlo to update the simulation one variable (or sub-block of variables) at a time, rather than conduct a single full-dimensional update. When it is possible to draw from each full-conditional distribution associated with the target this is just a Gibbs sampler. Often at least one of the Gibbs updates is replaced with a Metropolis–Hastings step, yielding a Metropolis–Hastings-within-Gibbs algorithm. Strategies for combining component-wise updates include composition, random sequence and random scans. While these strategies can ease MCMC implementation and produce superior empirical performance compared to full-dimensional updates, the theoretical convergence properties of the associated Markov chains have received limited attention. We present conditions under which some component-wise Markov chains converge to the stationary distribution at a geometric rate. We pay particular attention to the connections between the convergence rates of the various component-wise strategies. This is important since it ensures the existence of tools that an MCMC practitioner can use to be as confident in the simulation results as if they were based on independent and identically distributed samples. We illustrate our results in two examples including a hierarchical linear mixed model and one involving maximum likelihood estimation for mixed models.

Key words and phrases: Geometric ergodicity, uniform ergodicity, Markov chain, Monte Carlo, Gibbs sampler, Metropolis-within-Gibbs, random scan, convergence rate.

1. INTRODUCTION

Let ϖ be a probability distribution having support $\mathsf{X} \subseteq \mathbb{R}^q$, $q \geq 1$. The fundamental Markov chain Monte Carlo (MCMC) method for making draws from ϖ is the Metropolis–Hastings algorithm, described here. Let $X^{(k)} = x$ denote the current state, and suppose ϖ has a density function π. Let $p(\cdot, \cdot)$ denote the user-defined proposal density. The updated state $X^{(k+1)}$ is obtained via the following:

1. Simulate x^* from proposal density $p(x, \cdot)$.
2. Calculate acceptance probability $\alpha(x, x^*)$, where
$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{p(y, x)}{p(x, y)}\right\}.$$
3. Set
$$X^{(k+1)} = \begin{cases} x^*, & \text{with probability } \alpha(x, x^*), \\ x, & \text{with probability } 1 - \alpha(x, x^*). \end{cases}$$

Thus, creating a Metropolis–Hastings sampler boils down to choosing a proposal density $p(\cdot, \cdot)$. If $p(x, y) = p(y, x)$, this is a *Metropolis* algorithm. If, further, $p(x, y) = p(x - y) = p(y - x)$ for all x and y,

Alicia A. Johnson is Assistant Professor, Department of Mathematics, Statistics, and Computer Science, Macalester College, Saint Paul, Minnesota 55105, USA (e-mail: ajohns24@macalester.edu). Galin L. Jones is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, Minnesota 55405, USA (e-mail: galin@stat.umn.edu). Ronald C. Neath is Assistant Professor, Department of Mathematics and Statistics, Hunter College, City University of New York, New York, New York 10065, USA (e-mail: rneath@hunter.cuny.edu).

it is a *Metropolis random walk*. When the proposal $p(\cdot)$ does not depend on the current state the chain is a *Metropolis–Hastings independence sampler* (MHIS).

The selection of the proposal density can be challenging, particularly in problems where q is large or the support of ϖ is complicated. This has led to investigation of optimal scaling of Metropolis algorithms and so-called adaptive algorithms which allow the proposal kernel to change over the course of the simulation (see, e.g., Bédard and Rosenthal, 2008; Rosenthal, 2011). An alternative to full-dimensional updates is a *component-wise* approach where we update one variable (or sub-block of variables) at a time.

The choice between full-dimensional and component-wise updates is frequently unclear (see, e.g., Roberts and Sahu, 1997), although a general guideline seems to be that updating as a single block may not be advantageous if the components of ϖ are only weakly correlated. For example, Neal and Roberts (2006) considered target distributions having independent and identically distributed components and showed that for an idealized version of a component-wise Metropolis algorithm it is optimal to update only one variable at a time. On the other hand, these authors also showed that when using Metropolis adjusted Langevin algorithms, block updating is more efficient.

Whatever MCMC method is used, an important consideration is the rate of convergence of the chain to its stationary distribution. Let $\mathcal{B}$ be the Borel σ-algebra on X and let $P^n(x, dy)$ denote the n-step Markov transition kernel, that is, for any $x \in \mathsf{X}$, $A \in \mathcal{B}$, and $n \in \mathbb{Z}^+$, $P^n(x, A) = \Pr(X^{(n+j)} \in A | X^{(j)} = x)$ for the Markov chain $\Phi = \{X^{(0)}, X^{(1)}, X^{(2)}, \ldots\}$. Let $\|\cdot\|$ denote the total variation norm. If the chain is Harris ergodic, then for all $x \in \mathsf{X}$ we have $\|P^n(x, \cdot) - \varpi(\cdot)\| \to 0$ as $n \to \infty$. Now suppose there exist a real-valued function $M(x)$ on X and $0 < t < 1$ such that

$$\|P^n(x, \cdot) - \varpi(\cdot)\| \le M(x)t^n. \tag{1}$$

If M is bounded, then Φ is *uniformly ergodic* and otherwise it is *geometrically ergodic*.

A common goal of an MCMC experiment is to evaluate the quantity $E_\varpi g = \int_{\mathsf{X}} g(x)\varpi(dx)$, where g is a real-valued function on X whose expectation exists. Upon simulation of the Markov chain, $E_\varpi g$ is approximated by the sample average $\bar{g}_n = n^{-1}\sum_{i=0}^{n-1} g(X^{(i)})$. This approximation is usually justified through Birkhoff's ergodic theorem. Now, along with a moment condition on g, the existence of M and t in (1) ensures the existence of a central limit theorem for the Monte Carlo error, that is, there exists $0 < \sigma_g^2 < \infty$ such that, as $n \to \infty$,

$$\sqrt{n}(\bar{g}_n - E_\varpi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2). \tag{2}$$

Along with various moment conditions, the existence of M and t in (1) is also a key sufficient condition for using a variety of methods such as batch means, spectral methods or regenerative simulation to construct a strongly consistent estimator of σ_g^2, ensuring asymptotically valid Monte Carlo standard errors (Atchadé, 2011; Flegal and Jones, 2010; Hobert et al., 2002; Jones et al., 2006). Thus, when Φ is at least geometrically ergodic, a practitioner has the tools to be as confident in their simulations as if it were possible to make independent and identically distributed draws from ϖ (Flegal, Haran and Jones, 2008; Flegal and Jones, 2011).

Much work has been done on establishing geometric and uniform ergodicity of various versions of Metropolis–Hastings when full-dimensional updates are used; for example, Mengersen and Tweedie (1996) and Tierney (1994) studied the MHIS while Johnson and Geyer (2012), Mengersen and Tweedie (1996), Roberts and Tweedie (1996), Jarner and Hansen (2000), Christensen, Møller and Waagepetersen (2001) have established conditions under which Metropolis yields a geometrically ergodic chain. Other research on establishing convergence rates of Metropolis–Hastings chains includes Geyer (1999), Jarner and Hansen (2000) and Meyn and Tweedie (1994). Note well that none of these convergence rate results apply to component-wise implementations of Metropolis–Hastings. Also, a full-dimensional updating algorithm may fail to be geometrically ergodic while many component-wise updating samplers are. Consider the following simple example.

EXAMPLE 1. Suppose $\log\pi(x, y) = -(x^2 + x^2y^2 + y^2)$. Roberts and Tweedie (1996) showed that a Metropolis random walk having target π cannot be geometrically ergodic. Thus, one might consider component-wise methods where the target of each update is the relevant conditional distribution, $X|Y = y \sim \mathrm{N}(0, \frac{1}{2}(1 + y^2)^{-1})$ or $Y|X = x \sim \mathrm{N}(0, \frac{1}{2}(1 + x^2)^{-1})$. Fort et al. (2003) established geometric ergodicity of the uniform random scan Metropolis random walk. That is, at each step one of the components is selected with probability $1/2$ to remain fixed while a Metropolis random walk step is performed for the other. We will show that the Gibbs sampler is geometrically ergodic

as are the random scan Gibbs and random sequence Gibbs samplers for any selection probabilities.

We study conditions for ensuring geometric or uniform ergodicity for several component-wise strategies. Despite the near ubiquity of component-wise methods in MCMC practice, there has been very little work on this problem. In particular, there has been almost none in the case where the component-wise updates are done with Metropolis–Hastings (but see Fort et al., 2003; Jones, Roberts and Rosenthal, 2013; Roberts and Rosenthal, 1998). The one component-wise method that has received some attention in the literature is the Gibbs sampler, especially the two-variable deterministically updated Gibbs sampler; see, for example, the work in Doss and Hobert (2010), Hobert and Geyer (1998), Hobert et al. (2002), Johnson and Jones (2010), Jones and Hobert (2004), Marchev and Hobert (2004), Papaspiliopoulos and Roberts (2008), Roberts and Polson (1994), Roberts and Rosenthal (1999), Román and Hobert (2012), Rosenthal (1995, 1996), Roy and Hobert (2007), Tan and Hobert (2009) and Tierney (1994).

In Section 2 we fix some notation, state assumptions and develop a general framework for component-wise updates. Then in Section 3 we study the convergence rates of component-wise methods. In particular, we connect the convergence rate of deterministic scan samplers with random sequence scan and random scan methods. We also develop conditions for the uniform ergodicity of component-wise versions of the Metropolis–Hastings algorithm with state-independent candidate distributions. Along the way we apply our results to two practically relevant examples, including the Gibbs sampler for a Bayesian linear mixed model and one involving maximum likelihood estimation for mixed models, and provide empirical comparisons between samplers using component-wise updates and their full-dimensional counterparts. Notably, the empirical performance of the component-wise samplers is compellingly better, thus providing further support for the use of component-wise methods in practical problems; see also Caffo, Jank and Jones (2005), Coull et al. (2001), Johnson and Jones (2010), Jones et al. (2006), Jones and Hobert (2001), Lee et al. (2013), McCulloch (1997) and Neath (2013). Proofs of our results and other technical material are given in the supplement to the current article (Johnson, Jones and Neath, 2013).

2. COMPONENT-WISE UPDATES

Two fundamental strategies for combining Markov kernels are mixing and composition. Suppose $P_1, \ldots, P_d$ are Markov kernels having common invariant distribution ϖ. The general composition kernel is given by $P_{\text{comp}}(x, \cdot) = (P_1 \cdots P_d)(x, \cdot)$. Let $\mathbb{P}^d = \{(r_1, \ldots, r_d) \in \mathbb{R}^d : \text{each } r_i > 0, \sum_{i=1}^d r_i = 1\}$. Define the general mixing kernel

$$P_{\text{mix}}(x, \cdot) = r_1 P_1(x, \cdot) + \cdots + r_d P_d(x, \cdot), \quad r \in \mathbb{P}^d.$$

Then P_{mix} and P_{comp} are Markov kernels preserving the invariance of ϖ and we say that P_{mix} *has selection probabilities* r.

We can use these two strategies to create many component-wise algorithms. Suppose π is a density of ϖ with respect to a measure $\mu = \mu_1 \times \cdots \times \mu_d$ and has support $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ with Borel σ-algebra $\mathcal{B}$. We allow each $\mathsf{X}_i \subseteq \mathbb{R}^{b_i}$ so that the total dimension is $b_1 + \cdots + b_d$. If $x \in \mathsf{X}$, set $x_{(i)} = x \setminus x_i$. For $i = 1, \ldots, d$ let $g_i(y_i|x)$ be a density satisfying

$$\pi(y_i|x_{(i)}) = \int g_i(y_i|x)\pi(x_i|x_{(i)})\mu_i(dx_i). \tag{3}$$

That is, the conditional density $\pi(x_i|x_{(i)})$ is invariant for $g_i(y_i|x)$. Note that (3) is trivially satisfied if g_i corresponds to an elementary Gibbs update, that is, $g_i(y_i|x) = \pi(y_i|x_{(i)})$. Also, condition (3) is satisfied by construction if g_i corresponds to a Metropolis–Hastings algorithm having $\pi(y_i|x_{(i)})$ as its target.

Given (3), define a Markov kernel P_i as

$$P_i(x, A) = \int_A g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dy) \quad \text{for } A \in \mathcal{B}, \tag{4}$$

where δ is Dirac's delta. Then ϖ is invariant for each P_i so that $\varpi P_i = \varpi$ since

$$\int_{\mathsf{X}} \pi(x)g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dx) = \pi(y).$$

Since component-wise updates are not ϖ-irreducible, we need to combine the P_i in order to achieve a useful algorithm. Let $r \in \mathbb{P}^d$ be the selection probabilities corresponding to the components. Then we can write the random scan Markov kernel as

$$P_{\text{RS}}(x, A) = \sum_{i=1}^d r_i P_i(x, A) \tag{5}$$

and it is obvious that $\varpi P_{\text{RS}} = \varpi$. Moreover, P_{RS} admits a Markov transition density (Mtd)

$$h_{\text{RS}}(y|x) = \sum_{i=1}^d r_i g_i(y_i|x)\delta(y_{(i)} - x_{(i)}).$$

Another way to combine the P_i is through composition, that is, deterministically cycling through the

component-wise updates one at a time, in which case the Markov kernel is

$$P_C(x, A) = (P_1 \cdots P_d)(x, A) \tag{6}$$

and it is easy to see that $\varpi P_C = \varpi$ and that the associated Mtd is

$$h_C(y|x) = g_1(y_1|x)g_2(y_2|y_1, x_{(1)}) \cdots g_d(y_d|y_{(d)}, x_d).$$

There are $d!$ orders in which composition can be used and it is natural to consider using mixing to combine some of them. If $r \in \mathbb{P}^p$ for $p \le d!$, the sequence mixing kernel is given by

$$P_{\mathrm{RQ}}(x, A) = \sum_{j=1}^{p} r_j P_{C,j}(x, A), \tag{7}$$

where the $P_{C,j}$ are kernels created via composition but in different orders. Since $\varpi P_{C,j} = \varpi$ for each j, it is easy to see that $\varpi P_{\mathrm{RQ}} = \varpi$. Clearly, the Mtd is

$$h_{\mathrm{RQ}}(y|x) = \sum_{j=1}^{p} r_j h_{C,j}(y|x).$$

Note that the kernels defined in (5), (6) and (7) are special cases of the general definitions of P_{mix} and P_{comp}. We will employ the notation P_C, P_{RS} and P_{RQ} for the special case of component-wise updates, that is, when the P_i satisfy (4).

3. CONVERGENCE RATES UNDER COMPONENT-WISE UPDATES

Most of the research on convergence rates of component-wise MCMC algorithms has focused on those formed by composition, such as deterministic scan Gibbs samplers. One of our goals in this section is to show that the convergence rates of samplers formed by mixing and composition are related in concrete ways. However, we begin with a brief description of some techniques for establishing the existence of M and t in (1); see Meyn and Tweedie [(1993), Chapter 15] and Roberts and Rosenthal (2004) for details and Jones and Hobert (2001) for an accessible introduction.

Recall that $\mathcal{B}$ is the Borel σ-algebra on X and $P^n(x, dy)$ denotes a n-step Markov transition kernel, that is, for any $x \in \mathsf{X}$, $A \in \mathcal{B}$, and $n \in \mathbb{Z}^+$, $P^n(x, A) = \Pr(X^{(n+j)} \in A | X^{(j)} = x)$ for the Markov chain $\Phi = \{X^{(0)}, X^{(1)}, X^{(2)}, \ldots\}$.

Suppose there exist a positive integer n_0, an $\epsilon > 0$, a set $C \in \mathcal{B}$, and a probability measure Q on $\mathcal{B}$ such that

$$P^{n_0}(x, A) \ge \epsilon Q(A) \quad \text{for all } x \in C, A \in \mathcal{B}. \tag{8}$$

Then a *minorization condition* holds on the set C, called a *small set*. Uniform ergodicity is equivalent to the existence of a minorization condition on X.

Let

$$PV(x) := E\big[V\big(X^{(t+1)}\big) | X^{(t)} = x\big]. \tag{9}$$

A *drift condition* holds if there exists some function $V : \mathsf{X} \to [1, \infty)$, constants $0 < \gamma < 1$ and $k < \infty$ and a set C such that

$$PV(x) \le \gamma V(x) + k I_C(x) \quad \text{for all } x \in \mathsf{X}. \tag{10}$$

If C is small, then (10) is equivalent to geometric ergodicity (Meyn and Tweedie, 1993; Roberts and Rosenthal, 1997, 2004).

3.1 Uniform Ergodicity Under Mixing and Composition

We begin with some results concerning samplers P_{mix} and P_{comp}, after which we will specialize to some component-wise samplers.

THEOREM 1. *If P_{mix} is uniformly ergodic for some selection probabilities, then it is uniformly ergodic for all selection probabilities.*

It is easy to see that if one of the component samplers is uniformly ergodic, then P_{mix} will be uniformly ergodic. Also, it is sufficient to study P_{comp} to establish uniform ergodicity of P_{mix}.

THEOREM 2. *Suppose P_{comp} is uniformly ergodic. Then the corresponding P_{mix} is uniformly ergodic for any selection probabilities.*

For component-wise updates it can be difficult to establish uniform ergodicity of P_{mix} due to the form of its Mtd. Our next result follows directly from Theorems 1 and 2 and the observation that P_{RS} and P_{RQ} are special cases of P_{mix} and P_C is a special case of P_{comp}. See Łatuszynski, Roberts and Rosenthal (2013) and Roberts and Rosenthal (1997) for related results.

THEOREM 3. *If P_C is uniformly ergodic, then P_{RS} and P_{RQ} are uniformly ergodic for any selection probabilities.*

3.1.1 *Component-wise independence samplers.* It is clear that many component-wise Markov chains will not be uniformly ergodic. For example, it is well known that Metropolis random walks on $\mathbb{R}$ are not uniformly ergodic (Mengersen and Tweedie, 1996). Hence, when using such chains as the building blocks of a component-wise algorithm one does not expect to produce a uniformly ergodic Markov chain. On the

other hand, Mengersen and Tweedie (1996) did show that the full-dimensional Metropolis–Hastings independence sampler can be uniformly ergodic. We now turn our attention to the component-wise algorithm where each component-wise update is a Metropolis–Hastings algorithm with state-independent proposals. In this case, the composition sampler, P_{CIS}, the random sequence sampler, P_{RQIS}, and the random scan sampler, P_{RSIS}, are all *component-wise independence samplers* (CWIS).

We are interested in establishing conditions under which the CWIS are uniformly ergodic. By Theorem 3 it is sufficient to consider CIS. Note that since a typical P_{CIS} update will be some combination of accepted and rejected component-wise proposals, the P_{CIS} is not truly an independence sampler at all and, thus, the results of Mengersen and Tweedie (1996) are not applicable. It is, however, tempting to think that extending Mengersen and Tweedie's (1996) work on MHIS to P_{CIS} will be straightforward. Let $p_i(\cdot)$, a density on X_i, denote the state-independent proposal density for the ith update, $i = 1, \ldots, d$. If we let $p(x) = \prod_{i=1}^{d} p_i(x_i)$, a density on X, is the existence of $\epsilon > 0$ such that $p(x) \geq \epsilon\pi(x)$ a sufficient condition for uniform ergodicity of P_{CIS}? If we attempt to directly generalize Mengersen and Tweedie's (1996) argument, we are faced with 2^d cases to consider and, hence, this approach is fruitless. However, with a different approach we are able to give a pair of conditions that together are sufficient for uniform ergodicity of the CWIS.

To describe our results, we require a new notation. We will continue to let a subscript indicate the position of a vector component and a parenthetical superscript indicate the step in a Markov chain. Additionally, for each $i = 1, \ldots, d$, let $x_{[i]} = (x_1, \ldots, x_i)$ and $x^{[i]} = (x_i, \ldots, x_d)$; let $x_{[0]}$ and $x^{[d+1]}$ be null (vectors of dimension 0). We can now state the result for CWIS.

THEOREM 4. *Consider the kernel P_{CIS} with proposal densities p_i for $i = 1, \ldots, d$. Define $p(x) := \prod_{i=1}^{d} p_i(x_i)$, a density on X. Further suppose there exists $\delta > 0$ such that $p(x) \geq \delta\pi(x)$ for all $x \in \mathsf{X}$, and $\varepsilon > 0$ such that for any $x, y \in \mathsf{X}$ with $\pi(x) > 0$ and $\pi(y) > 0$,*

$$\begin{aligned}\pi(x)\pi(y) &= \pi\big(x_{[i]}, x^{[i+1]}\big)\pi\big(y_{[i]}, y^{[i+1]}\big)\\ &\geq \varepsilon\pi\big(x_{[i]}, y^{[i+1]}\big)\pi\big(y_{[i]}, x^{[i+1]}\big) > 0\end{aligned} \tag{11}$$

for each $i = 1, \ldots, d-1$. Then, for any $x \in \mathsf{X}$ and $A \in \mathcal{B}(\mathsf{X})$,

$$P_{\mathrm{CIS}}(x, A) \geq \delta\varepsilon^{\lfloor d/2 \rfloor}\pi(A)$$

and, thus, P_{CIS} is uniformly ergodic. Hence, P_{RQIS} and P_{RSIS} are also uniformly ergodic for any selection probabilities.

The following two corollaries indicate settings where the conditions of the theorem are easily verified.

COROLLARY 1. *Consider the kernel P_{CIS} with proposal densities p_i for $i = 1, \ldots, d$. Define $p(x) := \prod_{i=1}^{d} p_i(x_i)$, a density on X. Further suppose there exists $\delta > 0$ such that $p(x) \geq \delta\pi(x)$ for all $x \in \mathsf{X}$, and pairs of positive functions g_i and h_i on X_i for $i = 1, \ldots, d$ such that*

$$\prod_{i=1}^{d} g_i(x_i) \leq \pi(x) \leq \prod_{i=1}^{d} h_i(x_i) \tag{12}$$

for any $x \in \mathsf{X}$, and $\inf_{x_i \in \mathsf{X}_i}\{g_i(x_i)/h_i(x_i)\} > 0$ for each $i = 1, \ldots, d$. Then P_{CIS}, P_{RQIS} and P_{RSIS} are all uniformly ergodic.

The conditions of Corollary 1 amount to requiring at most a weak form of dependence in the target distribution. The most obvious special case is when the components of π are jointly independent, in which case (12) holds with equality on both sides.

COROLLARY 2. *Consider the kernel P_{CIS} with proposal densities p_i for $i = 1, \ldots, d$. Define $p(x) := \prod_{i=1}^{d} p_i(x_i)$, a density on X. If there exist $0 < a \leq b < \infty$ and $c > 0$ such that $a \leq \pi(x) \leq b$ and $p(x) \geq c$ for ϖ-almost all x, then P_{CIS}, P_{RQIS} and P_{RSIS} are all uniformly ergodic.*

3.1.2 *Maximum likelihood for mixed models.* Let $Y_i = \{Y_{i1}, \ldots, Y_{im_i}\}$ denote a vector of observable data, and let U_i denote the unobservable ith random effect, for $i = 1, \ldots, k$; let $U = (U_1, \ldots, U_k)$. Assume the Y_i are independent with distribution specified conditionally on $U = u$, so that the joint density of $Y = \{Y_{ij} : j = 1, \ldots, m_i; i = 1, \ldots, k\}$ is

$$f(y|u; \theta_1) = \prod_{i=1}^{k}\prod_{j=1}^{m_i} f(y_{ij}|u_i; \theta_1),$$

where θ_1 denotes a vector of parameters. The U_i are assumed to be independent, typically but not necessarily normally distributed, so the joint density of U is $h(u; \theta_2) = \prod_{i=1}^{k} h(u_i; \theta_2)$. Then the likelihood,

$$L(\theta; y) = \int f(y|u; \theta_1)h(u; \theta_2)\,du$$

is often analytically intractable so that calculating maximum likelihood estimates and their standard errors can be challenging. However, there are several

Monte Carlo-based algorithms, such as Monte Carlo Newton–Raphson, Monte Carlo maximum likelihood and Monte Carlo EM, which are useful for finding maximum likelihood estimators of the unknown parameter $\theta = (\theta_1, \theta_2)$ (Caffo, Jank and Jones, 2005; Hobert, 2000; McCulloch, 1997). A common feature is that all three algorithms require simulation from the same target distribution, namely, the conditional distribution of the random effects given the data, that is, for a given value of θ

$$h(u|y;\theta) \propto f(y|u;\theta_1)h(u;\theta_2).$$

We consider four Markov chains having $h(u|y;\theta)$ as the invariant density; the three component-wise independence samplers, CIS, RQIS and RSIS having proposal densities $h(u_i;\theta_2)$ for $i = 1, \ldots, k$ and a full-dimensional Metropolis–Hastings Independence Sampler (MHIS) with proposals drawn from the marginal distribution $h(u;\theta_2)$. To this end, the following result holds.

THEOREM 5. *If there exists $B(y,\theta_1) < \infty$ such that $f(y|u;\theta_1) \le B(y,\theta_1)$ for all u, then the four Markov chains described above, MHIS, CIS, RQIS and RSIS having $h(u|y,\theta)$ as the invariant density, are uniformly ergodic.*

We compare the empirical performance of the CWIS, the MHIS and a geometrically ergodic full-dimensional Metropolis random walk sampler in a concrete example in Section 4.2.

3.2 Geometric Ergodicity Under Mixing and Composition

Consider P_{mix} and suppose each of the kernels P_i are geometrically ergodic in that there are nonnegative functions M_i and $t_i \in (0,1)$ such that

$$\|P_i(x,\cdot) - \varpi(\cdot)\| \le M_i(x)t_i^n.$$

Then the triangle inequality implies

$$\|P_{\text{mix}}(x,\cdot) - \varpi(\cdot)\| \le [r_1M_1(x) + \cdots + r_dM_d(x)][\max\{t_1,\ldots,t_d\}]^n$$

and, hence, we have the following observation: if each P_i is geometrically ergodic, then so is P_{mix}. This demonstrates one difference between establishing geometric and uniform ergodicity; recall that we only required one of the P_i to be uniformly ergodic for P_{mix} to be uniformly ergodic. Also, this immediately implies that P_{RQ} is geometrically ergodic if each of the composition samplers are. However, it does not apply to P_{RS} since, in this case, each of the P_i typically are not even ϖ-irreducible. On the other hand, we have an analogue of Theorem 1, albeit with an additional assumption. Recall that a Markov kernel P is *reversible with respect to* ϖ if

$$P(x,dy)\varpi(dx) = P(y,dx)\varpi(dy).$$

THEOREM 6. *Suppose P_{mix} is reversible with respect to ϖ for all selection probabilities $r \in \mathbb{P}^d$. If P_{mix} is geometrically ergodic for some selection probability, then it is geometrically ergodic for all selection probabilities.*

Note that a special case of Theorem 6 is that if P_{RS} is geometrically ergodic for some selection probability, then it is geometrically ergodic for all selection probabilities; see Jones, Roberts and Rosenthal (2013) for related results. Also, Theorem 6 does not apply to random sequence scan samplers since these are not reversible for all selection probabilities.

3.2.1 *Two-variable settings.* We consider the case where ϖ has a density $\pi(x,y)$ with respect to $\mu_1 \times \mu_2$ and has support $\mathsf{X}_1 \times \mathsf{X}_2 \subseteq \mathbb{R}^{b_1} \times \mathbb{R}^{b_2}$. Let $\pi_{X|Y}(x|y)$ and $\pi_{Y|X}(y|x)$ be the full conditional densities and π_X and π_Y be the marginal densities derived from π (ϖ_X and ϖ_Y are the marginal distributions). This setting, though less general than that of the previous section, has many practical applications. For instance, it is the foundation for data augmentation methods (Hobert, 2011; Tanner and Wong, 1987) and many MCMC methods for practically relevant statistical models (Johnson and Jones, 2010; Román and Hobert, 2012; Roy and Hobert, 2007). We will consider two settings here: specifically, we begin with the case where sampling from $\pi_{X|Y}$ and $\pi_{Y|X}$ is possible and later turn our attention to the case where one of the Gibbs updates is replaced by a Metropolis–Hastings update.

When sampling from $\pi_{X|Y}$ and $\pi_{Y|X}$ is easy the Markov kernel formed by composition, say, P_{GS}, is the usual Gibbs sampler (GS) having Mtd

$$h_{\text{GS}}(x',y'|x,y) = \pi_{X|Y}(x'|y)\pi_{Y|X}(y'|x'). \tag{13}$$

Of course, the other update order is also a Gibbs sampler, denoted $\tilde{P}_{\text{GS}}$. Also, each of the marginal sequences $\{X^{(n)}\}$ and $\{Y^{(n)}\}$ have one-step Markov kernels P_X and P_Y with Mtds

$$h_X(x'|x) = \int \pi_{X|Y}(x'|y)\pi_{Y|X}(y|x)\mu_2(dy)$$

and

$$h_Y(y'|y) = \int \pi_{Y|X}(y'|x)\pi_{X|Y}(x|y)\mu_1(dx), \tag{14}$$

respectively. Moreover, it is easy to see that P_Y^m admits an m-step Mtd $h_Y^m(y'|y)$ as do P_X^m, P_{GS}^m and $\tilde{P}_{\mathrm{GS}}^m$. Note that π_X is invariant for $\{X^{(n)}\}$ and π_Y is invariant for $\{Y^{(n)}\}$.

It is well known that P_X, P_Y, P_{GS} and $\tilde{P}_{\mathrm{GS}}$ all converge at the same qualitative rate (Diaconis, Khare and Saloff-Coste, 2008; Robert, 1995; Roberts and Rosenthal, 2001). In particular, if one is geometrically ergodic, then so are the others. This relationship has been routinely exploited in the analysis of Gibbs samplers for practically relevant statistical models, where it is often easier to analyze P_Y or P_X than P_{GS}. Putting these observations together with our above work says that if one of P_X, P_Y, P_{GS} or $\tilde{P}_{\mathrm{GS}}$ are geometrically ergodic, then so are the others and so is the random sequence Gibbs sampler P_{RQGS}.

The first result of this subsection connects the convergence rate of P_X, P_Y, P_{GS} and $\tilde{P}_{\mathrm{GS}}$ to the random scan Gibbs sampler P_{RSGS}. Note that by using (9) and (14) we have that

$$P_Y W(y) = \int_{\mathsf{X}_2} W(y') h_Y(y'|y) \mu_2(dy').$$

THEOREM 7. *Suppose there exists* $\lambda < 1$, $W: \mathsf{X}_2 \to \mathbb{R}^+$, $b < \infty$ *such that*

$$P_Y W(y) \le \lambda W(y) + b. \tag{15}$$

Let $C_d = \{y : W(y) \le d\}$ *and suppose there is a* $g: \mathsf{X}_2 \to \mathbb{R}^+$ *and a* $d_0 > 0$ *such that for some* $m \ge 1$

$$h_Y^m(y'|y) \ge g(y') \quad \text{for all } y \in C_d \text{ and } d \ge d_0. \tag{16}$$

Then P_Y, P_X, P_{GS} *and* $\tilde{P}_{\mathrm{GS}}$ *are geometrically ergodic as are* P_{RQGS} *and* P_{RSGS}.

There is a simple sufficient condition for (16); suppose there is a $l: \mathsf{X}_2 \to \mathbb{R}^+$ such that $\pi_{X|Y}(x|y) \ge l(x)$ for all $(x, y) \in \mathsf{X}_1 \times C_d$ with $d \ge d_0$. Then if $y \in C_d$,

$$\begin{aligned} h_Y(y'|y) &= \int_{\mathsf{X}_1} \pi_{Y|X}(y'|z) \pi_{X|Y}(z|y) \mu_1(dz) \\ &\ge \int_{\mathsf{X}_1} \pi_{Y|X}(y'|z) l(z) \mu_1(dz) = g(y'). \end{aligned}$$

Although we do not state it formally, it is clear from our proof that the same conclusions obtain if we were to reformulate the conditions in terms of P_X instead of P_Y.

EXAMPLE 2. Consider the Gibbs sampler defined in Example 1. If $W(y) = y^2$, it is easy to see that $P_Y W(y) \le \lambda W(y) + 0.5$ for any $0 < \lambda < 1$. Moreover, (16) holds since if $y \in C_d$ for any $d > 0$, then $\pi_{X|Y}(x|y) \ge \pi^{-0.5} e^{-(1+d)x^2}$. Hence, the claims of Example 1 hold by the theorem.

Of the Gibbs samplers for practically relevant statistical problems proved to be geometrically ergodic—see the references in Section 1—only those in Doss and Hobert (2010) and Hobert and Geyer (1998) have had more than two components. Thus, Theorem 7 can be coupled with existing results to obtain the geometric ergodicity of the random sequence and random scan versions of many Gibbs samplers which have been proved geometrically ergodic.

Now suppose we are able to draw from $\pi_{X|Y}$, but instead of sampling from $\pi_{Y|X}$ we substitute a Metropolis–Hastings step g_2 having proposal density p_2. This results in a hybrid composition sampler (often called Metropolis–Hastings-within-Gibbs) having Markov kernel P_{HC} and Mtd

$$h_{\mathrm{HC}}(x', y'|x, y) = \pi_{X|Y}(x'|y) g_2(y'|x', y).$$

Then the marginal Y-sequence is Markovian with kernel P_Y having Mtd

$$h_Y(y'|y) = \int \pi_{X|Y}(x|y) g_2(y'|x, y) \mu_1(dx) \tag{17}$$

with invariant density π_Y but the marginal X-sequence is not Markovian. Nevertheless, Robert [(1995), Theorem 4.1] showed that the X- and Y-sequences converge at the same rate in total variation norm. That is, let $\tilde{P}_X^n((x, y), \cdot)$ be the marginal distribution of $X^{(n)}$ given initial state $(X^{(0)}, Y^{(0)}) = (x, y)$, then for each $n \ge 1$

$$\begin{aligned} \|P_Y^{n+1}(y, \cdot) - \varpi_Y(\cdot)\| &\le \|\tilde{P}_X^n((x, y), \cdot) - \varpi_X(\cdot)\| \\ &\le \|P_Y^n(y, \cdot) - \varpi_Y(\cdot)\|. \end{aligned}$$

An easy calculation shows that

$$\|P_Y^{n+1}(y, \cdot) - \varpi_Y(\cdot)\| \le \|P_{\mathrm{HC}}^n((x, y), \cdot) - \varpi(\cdot)\|.$$

It is also easy to see that the Y-sequence is deinitializing for P_{HC} (Roberts and Rosenthal, 2001). Hence, P_Y is geometrically ergodic if and only if P_{HC} is geometrically ergodic. Our next result connects the convergence rate of P_Y and P_{HC} to the convergence rate of the random scan hybrid chain having kernel P_{RSH} and Mtd

$$\begin{aligned} h_{\mathrm{RSH}}(x', y'|x, y) = {} & r \pi_{X|Y}(x'|y) \delta(y' - y) \\ & + (1 - r) g_2(y'|x, y) \delta(x' - x). \end{aligned}$$

It is straightforward to show that P_{RSH} is reversible with respect to ϖ. Note that by using (9) and (17) we have that

$$P_Y W(y) = \int_{\mathsf{X}_2} W(y') h_Y(y'|y) \mu_2(dy').$$

THEOREM 8. *Suppose there exists* $W:\mathsf{X}_2 \to \mathbb{R}^+$ *and constants* $\lambda, b < \infty$ *such that*

$$P_Y W(y) \le \lambda W(y) + b. \tag{18}$$

Let $C_d = \{y : W(y) \le d\}$ *and suppose there is a* $g:\mathsf{X}_2 \to \mathbb{R}^+$ *and a* $d_0 > 0$ *such that for some* $m \ge 1$

$$h_Y^m(y'|y) \ge g(y') \quad \text{for all } y \in C_d \text{ and } d \ge d_0. \tag{19}$$

Then P_Y *and* P_{HC} *are geometrically ergodic. Further suppose the proposal density* p_2 *for the Metropolis–Hastings step* g_2 *satisfies either*:

1. $p_2(z|x,y) = p_2(y|x,z)$ *and there exists* $K < \infty$ *such that* $p_2(z|x,y)/p_2(z|x,u) \le K$, *or*
2. $p_2(z|x,y) = p_2(z|x)$.

Then P_{RSH} *is geometrically ergodic.*

As with the Gibbs sampler setting, there is a simple sufficient condition for (19); suppose there is a nonnegative function l such that for $y \in C_d$

$$\pi_{X|Y}(x|y) g_2(z|x,y) \ge l(x,z).$$

In this case, if $y \in C_d$, then

$$\begin{aligned} h_Y(y'|y) &= \int_{\mathsf{X}_1} \pi_{Y|X}(y'|x) g_2(y'|x,y) \mu_1(dx) \\ &\ge \int_{\mathsf{X}_1} l(x,y') \mu_1(dx) = g(y'). \end{aligned}$$

3.2.2 *Gibbs samplers for a Bayesian linear mixed model.* Let Y denote an $N \times 1$ response vector and let β be a $p \times 1$ vector of regression coefficients, u be a $k \times 1$ vector of random effects, X be a known $N \times p$ design matrix and Z be a known $N \times k$ matrix. Then for $r, s, t \in \{1, 2, \ldots\}$ suppose

$$\begin{aligned} Y|\beta, u, \lambda_R, \lambda_D &\sim \mathrm{N}_N\bigl(X\beta + Zu, \lambda_R^{-1} I_N\bigr), \\ \beta|u, \lambda_R, \lambda_D &\sim \sum_{i=1}^{r} \eta_i \mathrm{N}_p\bigl(b_i, B^{-1}\bigr), \\ u|\lambda_R, \lambda_D &\sim \mathrm{N}_k\bigl(0, \lambda_D^{-1} I_k\bigr), \\ \lambda_R &\sim \sum_{j=1}^{s} \phi_j \,\mathrm{Gamma}(r_{j1}, r_{j2}), \\ \lambda_D &\sim \sum_{l=1}^{t} \psi_l \,\mathrm{Gamma}(d_{l1}, d_{l2}), \end{aligned}$$

where the mixture parameters η_i, ϕ_j and ψ_l are known nonnegative constants satisfying

$$\sum_{i=1}^{r} \eta_i = \sum_{j=1}^{s} \phi_j = \sum_{l=1}^{t} \psi_l = 1.$$

Note that we say $W \sim \mathrm{Gamma}(a,b)$ if it has density proportional to $w^{a-1} e^{-bw} I(w > 0)$. Finally, we also assume $X^T Z = 0$, $b_i \in \mathbb{R}$, and the positive definite matrix B are known and the hyperparameters r_{j1}, r_{j2}, d_{l1} and d_{l2} are positive.

Let $\xi = (u^T, \beta^T)^T$ and $\lambda = (\lambda_R, \lambda_D)^T$. Then the posterior density is characterized by

$$\pi(\xi, \lambda|y) \propto f(y|\xi,\lambda) f(\xi|\lambda) f(\lambda),$$

where y is the observed data and f denotes a generic density. It is straightforward to derive the conditional distributions of $\xi|\lambda, y$ and $\lambda|\xi, y$, which are reported here. Let

$$v_1(\xi) = (y - X\beta - Zu)^T (y - X\beta - Zu)$$

and

$$v_2(\xi) = u^T u.$$

Then the distribution of $\lambda|\xi, y$ has density

$$f(\lambda|\xi, y) = \sum_{j=1}^{s} \sum_{l=1}^{t} \phi_j \psi_l f_{1j}(\lambda_R|\xi, y) f_{2l}(\lambda_D|\xi, y),$$

where $f_{1j}(\cdot|\xi, y)$ is a $\mathrm{Gamma}(r_{j1} + N/2, r_{j2} + v_1(\xi)/2)$ density and $f_{2l}(\cdot|\xi, y)$ denotes a $\mathrm{Gamma}(d_{l1} + k/2, d_{l2} + v_2(\xi)/2)$ density. Next, $\xi|\lambda, y \sim \sum_{i=1}^{r} \eta_i \mathrm{N}(m_0, \Sigma^{-1})$, where

$$\Sigma^{-1} = \begin{pmatrix} (\lambda_R Z^T Z + \lambda_D I_k)^{-1} & 0 \\ 0 & (\lambda_R X^T X + B)^{-1} \end{pmatrix}$$

and

$$m_0 = \begin{pmatrix} \lambda_R (\lambda_R Z^T Z + \lambda_D I_k)^{-1} Z^T y \\ (\lambda_R X^T X + B)^{-1} (\lambda_R X^T y + Bb) \end{pmatrix}.$$

It is straightforward to implement any of the Gibbs sampling strategies. For example, consider the Gibbs sampler that updates ξ followed by λ having Mtd [recall (13)]

$$h_{\mathrm{GS}}(\xi', \lambda'|\xi, \lambda) = f(\xi'|\lambda, y) f(\lambda'|\xi', y).$$

We can similarly use the full conditionals and the recipes described earlier to construct Mtds for the related Markov chains, say, h_ξ, h_λ, $\tilde{h}_{\mathrm{GS}}$, h_{RSGS} and h_{RQGS}.

Johnson and Jones (2010) establish (15) and (16) for the marginal ξ-sequence having Mtd h_ξ and, hence, we can appeal to Theorem 7 to establish the geometric ergodicity of the Gibbs samplers. Let x_i and z_i be the ith rows of X and Z, respectively. Define

$$\begin{aligned} G_i(\lambda) = &\sum_{m=1}^{N} \bigl[E_i(y_m - x_m\beta - z_m u|\lambda, y)\bigr]^2 \\ &+ \sum_{m=1}^{k} \bigl[E_i(u_m|\lambda, y)\bigr]^2, \end{aligned}$$

where E_i denotes expectation with respect to the $N_{k+p}(m_i, \Sigma^{-1})$ distribution.

THEOREM 9. *Assume there exists some $K < \infty$ such that $G_i(\lambda) \leq K$. If for all $j \in \{1, \ldots, s\}$ and $l \in \{1, \ldots, t\}$*

$$r_{j1} > 0 \vee \frac{1}{2}\left[\sum_{i=1}^{N} z_i (Z^T Z)^{-1} z_i^T - N + 2\right]$$

and

$$d_{l1} > 1,$$

then the marginal ξ- and λ-chains and GS are geometrically ergodic as are RSGS and RQGS.

Johnson and Jones (2010) provide other conditions under which GS is geometrically ergodic and, hence, the theorem does not exhaust the conditions under which RQGS and RSGS are geometrically ergodic; see also Román (2012) for some improvements on the results of Johnson and Jones (2010). In Section 4.1 we consider a special case of our model and provide an empirical comparison of GS, RQGS and RSGS with a full-dimensional Metropolis sampler.

4. EXAMPLES

We consider two examples based on the settings introduced in Sections 3.1.2 and 3.2.2. In each case we consider the finite sample empirical performance of some component-wise MCMC algorithms against full-dimensional updates. This comparison is based on several measures of efficiency, which are now described.

If $E_\varpi |g(X)|^{2+\delta} < \infty$ for some $\delta > 0$ and the Markov chain is geometrically ergodic, then a central limit theorem, recall (2), holds. Therefore, $t_* \sigma_g / \sqrt{n}$ gives the half-width of an asymptotically valid confidence interval for $E_\varpi g$ where t_* is an appropriate quantile. The width of the interval can be used to determine the number of iterations required to achieve some desired level of precision (Flegal, Haran and Jones, 2008; Flegal and Jones, 2011; Jones et al., 2006). We might also measure Markov chain efficiency relative to the efficiency of a would-be random sample from ϖ. One such measure, the integrated autocorrelation time (ACT)

$$\text{ACT} = \frac{\sigma_g^2}{\text{Var}_\varpi(g(X))},$$

compares the variability of the Monte Carlo estimate to that of an estimate based on a random sample of the same size. In practice, $\text{Var}_\varpi(g(X))$ and σ_g^2 are unknown. However, a consistent estimator of $\text{Var}_\varpi(g(X))$ is given by the sample variance, $\widehat{\text{Var}}_\varpi(g(X))$, and, because the chains are geometrically ergodic, the consistent batch means estimator of Jones et al. (2006), say, $\hat{\sigma}_g^2$, provides a consistent estimator of σ_g^2.

For a given sample size, the quality of Monte Carlo estimates can be assessed using the mean squared error (MSE). To estimate the MSE, we run m independent replications of each chain, each of which is of length n, producing independent estimates $\bar{g}_n^{(1)}, \ldots, \bar{g}_n^{(m)}$ and an independent estimate based on a long run of a given chain, say, $\bar{g}^*$. The estimated MSE is

$$\widehat{\text{MSE}}_m(\bar{g}_n) = \frac{1}{m}\sum_{i=1}^{m} (\bar{g}_n^{(i)} - \bar{g}^*)^2.$$

The above quantities allow examination of efficiency only in terms of estimating $\text{E}_\varpi g(X)$. We also compare how the chains move around the state space using the expected square Euclidean jump distance (ESEJD), that is, the expected squared distance between successive draws of the Markov chain $X^{(i)}$ and $X^{(i+1)}$. If $\|\cdot\|_2$ denotes the standard Euclidean norm, then ESEJD is the expected value of the mean square Euclidean jump Distance (MSEJD) at stationarity, where for a chain of length n,

$$\text{MSEJD} := \frac{1}{n-1}\sum_{i=1}^{n-1} \|X^{(i+1)} - X^{(i)}\|_2^2.$$

Given m independent replications of each chain, each of which is length n, we can estimate ESEJD with

$$\widehat{\text{ESEJD}}_m = \frac{1}{m}\sum_{i=1}^{m} \text{MSEJD}^{(i)}.$$

In addition to the above numerical summaries, we include standard graphical summaries such as trace plots. Taken together, these measures give us a reasonable picture of the empirical performance of the various algorithms examined below.

4.1 A Bayesian Linear Mixed Model

Consider a Bayesian version of the balanced random intercept model for k subjects and $m \geq 2$ observations per subject. Let $y_i = (y_{i1}, \ldots, y_{im})^T$ be the data for subject i and $Y = (y_1^T, \ldots, y_k^T)^T$ denote the overall $N \times 1$ response vector where $N = km$. Further, let $u = (u_1, \ldots, u_k)^T$ be a vector of subject effects and X

be a full column rank $N \times p$ design matrix corresponding to β, a $p \times 1$ vector of regression coefficients. Then the first level of the hierarchy is

$$Y|\beta, u, \lambda_R, \lambda_D \sim \mathrm{N}_N(X\beta + Zu, \lambda_R^{-1} I_N)$$

for $Z = I_k \otimes 1_m$, where $\otimes$ denotes the Kronecker product and 1_m is an $m \times 1$ vector of ones. At the next stage,

$$\beta|\lambda_R, \lambda_D \sim \mathrm{N}_p(b, B^{-1})$$

and

$$u|\lambda_R, \lambda_D \sim \mathrm{N}_k(0, \lambda_D^{-1} I_k)$$

for known $b \in \mathbb{R}^p$ and positive definite matrix B. Finally,

$$\lambda_R \sim \mathrm{Gamma}(r_1, r_2) \quad \text{and} \quad \lambda_D \sim \mathrm{Gamma}(d_1, d_2),$$

where r_1, r_2, d_1, d_2 are positive. This hierarchy is a special case of the Bayesian general linear model of Section 3.2.2 and it follows from Theorem 9 that if $d_1 > 1$, then GS, RQGS and RSGS are geometrically ergodic.

We present an empirical comparison of the GS, uniform RQGS and uniform RSGS algorithms. We also compare the three Gibbs samplers to a full-dimensional Metropolis random walk. In our comparison we focus on estimating the posterior expectation of β, that is, $\mathrm{E}(\beta|y)$.

Our Metropolis random walk (RW) uses a multivariate Normal proposal distribution centered at the current value of the chain and with a diagonal covariance matrix. We set the diagonal elements equal to those of $\hat{\Sigma}^2$, where $\hat{\Sigma}$ is an estimate of the posterior covariance matrix obtained from an independent run of 10^5 iterations of the GS. For the settings described below, our RW has a proposal acceptance rate of approximately 0.30. We do not know if this RW Markov chain is geometrically ergodic.

We simulated data (values of y) under the following settings. Set $k = 10$, $m = 5$, and $p = 1$, and $X = (x_1^T, \ldots, x_{10}^T)^T$, where for all i, $x_i^T = (-0.50, -0.25, 0, 0.25, 0.50)$ with $b = 0$, $B^{-1} = 0.1$, and $r_1 = r_2 = d_1 = d_2 = 2$. Assuming the true nature of this data is unknown, we simulate the four Markov chains under the hyperparameter setting with $b = 0$, $B^{-1} = 0.1$, and $r_1 = r_2 = d_1 = d_2 = 3$. Finally, all chains are started from the prior means, $(\beta^{(0)}, u^{(0)}, \lambda_R^{(0)}, \lambda_D^{(0)}) = (0, 0_k, 1, 1)$ where 0_k is a $k \times 1$ vector of zeroes.

Since $E[\beta^4|y] < \infty$, the geometric ergodicity of GS, RQGS and RSGS guarantees a central limit theorem for the Monte Carlo error $\bar{\beta}_n - \mathrm{E}(\beta|y)$ with the variance of the asymptotic distribution denoted σ_β^2.

We ran each algorithm (RW, RSGS, RQGS and GS) independently for 10^5 iterations. Trace plots of the final 1000 β iterations are shown in Figure 1. Mixing appears to be substantially quicker for the Gibbs samplers than for the RW, while RQGS and GS appear to be more efficient than the RSGS.

The differences in the trace plots between the four simulations are reflected in the interval half-width and ACT estimates given in Table 1. For equivalent sample sizes, the RW half-width is nearly two times that of the RSGS and approximately three times as large as those of the GS and RQGS. In addition, the ACTs indicate that nearly eleven RW samples and more than three RSGS samples are required for each random draw from π in order to achieve the same level of precision for estimates of $\mathrm{E}(\beta|y)$. On the other hand, each RQGS sample and GS sample is approximately as effective as a random draw.

In order to estimate the MSE of the Monte Carlo estimates, we simulated $m = 10^3$ independent replications of RW, RSGS, RQGS and GS for $n = 10^4$ iterations each and took $\bar{\beta}^*$ to be an estimate of $\mathrm{E}(\beta|y)$ obtained from 10^5 iterations of the RW chain. The estimated MSE ratios relative to the GS,

$$\frac{\widehat{\mathrm{MSE}}(\bar{\beta}_{n,*})}{\widehat{\mathrm{MSE}}(\bar{\beta}_{n,\mathrm{GS}})}$$

are also given in Table 1 along with standard errors. Notice that ratios greater than one favor GS. Hence, these results are consistent with those above, which suggest that the single block update RW is less efficient than the Gibbs samplers with respect to estimation of $\mathrm{E}(\beta|y)$.

Estimation of the ESEJD is based on the same $m = 1000$ independent replications of RW, RSGS, RQGS and GS for $n = 10^4$ iterations each. The estimates are reported along with standard errors in Table 1. The message here is consistent with the above discussions. The RW appears to be less efficient than the Gibbs samplers in exploring the support of the posterior. Among the Gibbs samplers, there is little difference in the performance quality of the RQGS and GS, whereas both are more efficient than the RSGS.

4.2 A Logit-Normal Mixed Model

Consider the following special case of the mixed model defined in Section 3.1.2. Suppose, conditional on $U = u$, the observations Y_{ij} are independently distributed as Bernoulli(p_{ij}), where $\mathrm{logit}(p_{ij}) = \beta x_{ij} + u_i$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, k$, where the x_{ij} are covariates. Let the random effects $U_1, \ldots, U_k$ be

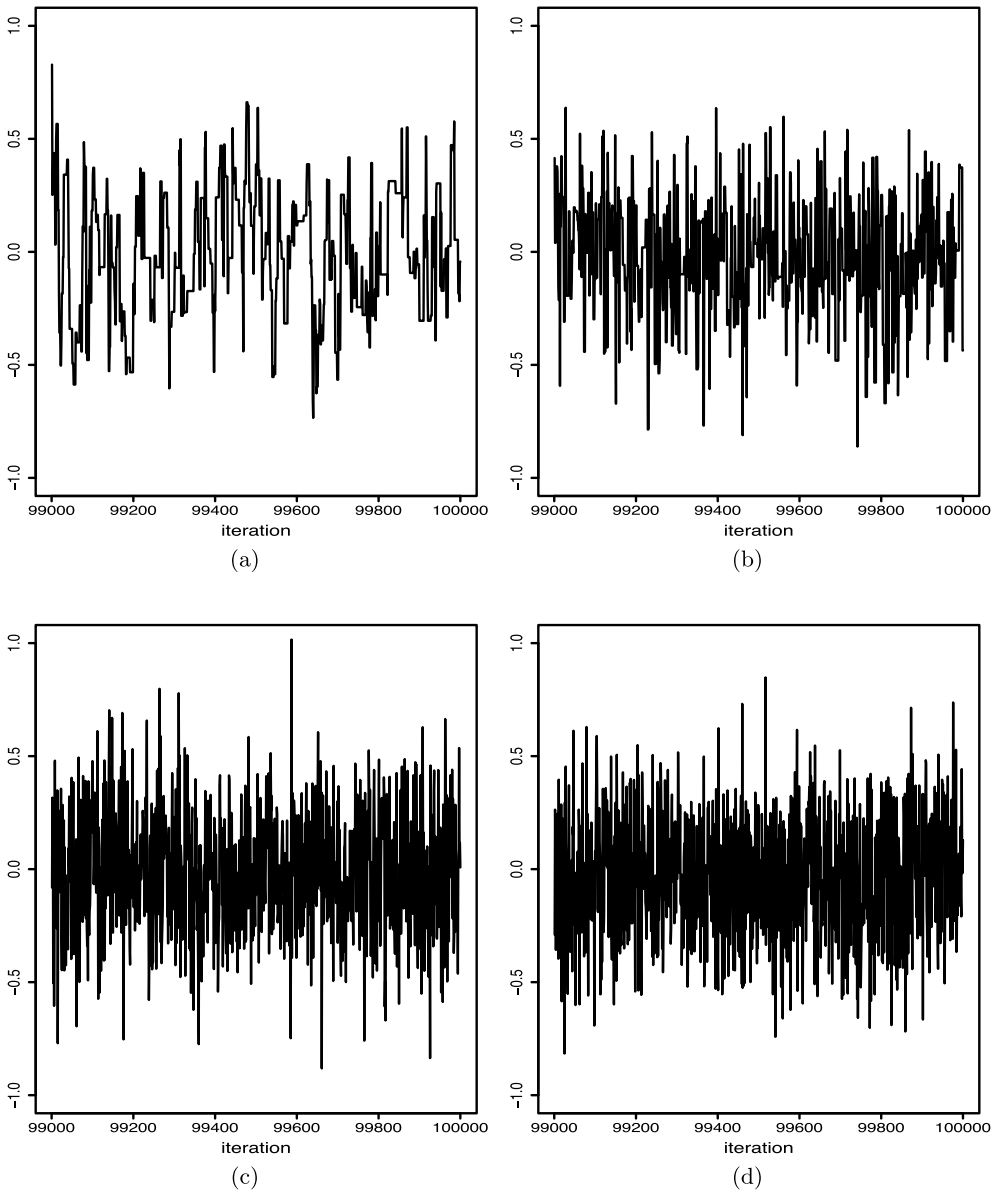

FIG. 1. *Trace plots of* β *for iterations* 9.9*e*4 *through* 1*e*5 *of the* (a) *RW*, (b) *RSGS*, (c) *RQGS and* (d) *GS in the Bayesian linear mixed model example of Section* 4.1.

TABLE 1

Results for the Bayesian linear mixed model example of Section 4.1. *Estimates of* $\mathrm{E}(\beta|y)$, σ_β^2 *and* $\mathrm{Var}(\beta|y)$ *are based on* $n = 10^5$; *middle columns show half-width of* 95% *confidence interval* ($t_* = 1.960$) *and integrated autocorrelation time* (*ACT*). *MSE Ratios are relative to GS, with standard errors given in parentheses. Final column shows estimated ESEJD, with standard error in parentheses*

Algorithm	$\bar{\beta}_n$	$\hat{\sigma}_\beta^2$	$t_*\hat{\sigma}_\beta/\sqrt{n}$	**ACT**	**MSE ratio**	$\widehat{\mathbf{ESEJD}}$
RW	−0.018	0.794	0.0055	10.919	5.55 (0.32)	0.26 (0.0002)
RSGS	−0.016	0.243	0.0031	3.375	2.07 (0.13)	3.20 (0.0014)
RQGS	−0.016	0.071	0.0017	0.986	0.98 (0.05)	6.19 (0.0013)
GS	−0.015	0.083	0.0018	1.153	1.00 (0.00)	6.19 (0.0012)

i.i.d. Normal$(0, \sigma^2)$. With the parameters $\theta = (\beta, \sigma^2)$ treated as fixed, the target density is

$$h(u|y;\theta) \propto \exp\left\{\sum_{i=1}^{k}\left[u_i y_{i+} - \sum_{j=1}^{m_i}\log\left(1+e^{\beta x_{ij}+u_i}\right) - \frac{u_i^2}{2\sigma^2}\right]\right\}, \tag{20}$$

where $y_{i+} = \sum_{j=1}^{m_i} y_{ij}$ for $i = 1, \ldots, k$.

In Section 3.1.2 we introduced the MHIS, CIS, RQIS and RSIS algorithms. In the current context the conditions of Theorem 5 are satisfied and, hence, those four samplers are uniformly ergodic. In addition, we consider a full-dimensional Metropolis random walk (RW) sampler with normally distributed jump proposals, that is, the proposal density is

$$p(u, u^*) \propto \exp\left\{-\frac{1}{2\tau^2}\|u^* - u\|_2^2\right\}, \tag{21}$$

where $\|\cdot\|_2$ denotes the standard Euclidean norm and τ^2 is a tuning parameter. Then the following result holds.

THEOREM 10. *The full-dimensional Metropolis random walk sampler with invariant density* (20) *and proposal density* (21) *is geometrically ergodic.*

We compare the empirical performance of the algorithms in the context of implementing a Monte Carlo EM (MCMC) algorithm. Now at each step the MCEM requires a Monte Carlo approximation to the so-called Q-function

$$Q(\theta;\tilde{\theta}) = \int l_c(\theta; y, u) h(u|y;\tilde{\theta})\, du,$$

where

$$l_c(\theta; y, u) = \sum_{i=1}^{k}\sum_{j=1}^{m_i}\left[y_{ij}(\beta x_{ij} + u_i) - \log\left(1+e^{\beta x_{ij}+u_i}\right)\right] - \frac{k}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{k} u_i^2$$

denotes the "complete-data log-likelihood," what the log-likelihood would be if the random effects were observable. We consider implementation of MCEM in a benchmark data set given by Booth and Hobert [(1999), Table 2], assuming the true parameter value $\theta = (\beta, \sigma^2) = (5.0, 0.5)$. In this data set $x_{ij} = j/15$ for each $j = 1, \ldots, m_i \equiv 15$, for each $i = 1, \ldots, k = 10$. Let $\tilde{\theta} = (4.0, 1.5)$. We can take as an MCMC approximation of the Q-function the sample average of the chain $\{l_c(\theta; y, u^{(t)})\}$, that is,

$$Q(\theta;\tilde{\theta}) \approx \bar{l}_{Cn}(\theta;\tilde{\theta}) := \frac{1}{n}\sum_{t=1}^{n} l_c\left(\theta; y, u^{(t)}\right),$$

where $\{u^{(t)} : t = 1, 2, \ldots, n\}$ is a realization of one of our five Markov chains with stationary density $h(u|y;\tilde{\theta})$ as defined by (20). For the sake of simplicity we will consider estimating the point $Q(\tilde{\theta};\tilde{\theta})$ rather than the entire function. The mixing conditions on the Markov chain ensure the existence of a CLT for the Monte Carlo error $\bar{l}_{Cn}(\tilde{\theta};\tilde{\theta}) - Q(\tilde{\theta};\tilde{\theta})$ with the variance of the asymptotic normal distribution denoted σ_Q^2, which can be consistently estimated with the batch means estimator $\hat{\sigma}_Q^2$ (Jones et al., 2006).

We implemented MHIS, RW, CIS and RSIS as discussed above—we skip reporting our implementation of RQIS, as it is very similar to CIS in this example—in each case simulating a chain of length $n = 10^6$ and taking as our initial distribution $U^{(0)} \sim \mathrm{N}_{10}(0, \sigma^2 I)$. For the Metropolis random walk we drew our jump proposals from a $\mathrm{N}_{10}(0, \tau^2 I)$, with $\tau^2 = \sigma^2/6$ (this setting determined by trial and error, in order to minimize the autocorrelation in the resulting chain, and yielded an observed acceptance rate of 27.3%). A partial trace plot (the second 1000 updates) is shown in panel (a) of Figure 2. Analogous plots for the MHIS and component-wise algorithms appear in the remaining panels.

Consider the trace plots for the four chains. The most striking result is the dreadful performance of the MHIS, shown in panel (b). The RW chain [panel (a)] mixes much faster than the MHIS, but still shows significant autocorrelation. Now RSIS [panel (d)] appears to mix faster than MHIS but is very similar to RW. Finally, the CIS [panel (c)] chain appears to be the best of these four samplers. This suggests that when there is weak dependence between the components of the target distribution, one should use a CWIS instead of a MHIS; recall that Neal and Roberts (2006) reached a similar conclusion for the Metropolis random walk.

In general, the empirical performance of MHIS depends entirely on the "closeness" of the proposal distribution to the target and, clearly, the marginal distribution of the random effects U is not sufficiently similar to the conditional distribution of U given the data. It is worth recalling that, by Theorem 5, the MHIS depicted in panel (b) of Figure 2 is a uniformly ergodic Markov chain. Thus, this example nicely illustrates the perils

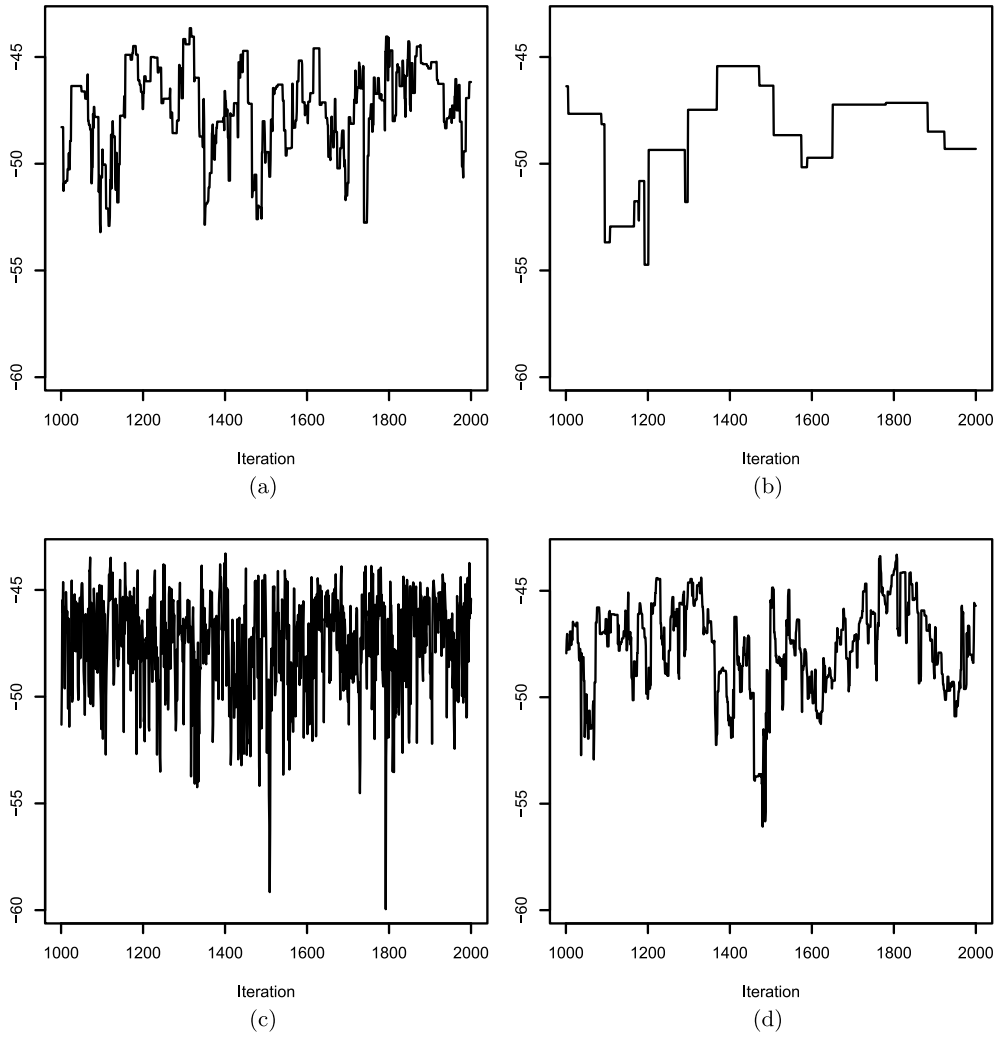

FIG. 2. *Partial trace plots for the Markov chain* $\{l_c(\theta; y, u^{(t)})\}$ *generated by* (a) *RW*, (b) *MHIS*, (c) *CIS and* (d) *RSIS in the logit-normal example of Section* 4.2.

of over-reliance on asymptotic properties of a sampler, which provide no guarantee of favorable performance in finite-sample implementations.

Consider the simulation results given in Table 2. Using equivalent Monte Carlo sample sizes, the half-width of the interval estimator is roughly the same for RW and RSIS. The half-widths for RW and RSIS are more than 3 times larger than the half-width for CIS. On the other hand, the half-width for MHIS is more than 3 times larger than those of RSIS and RW and

TABLE 2

Results for the logit-normal example of Section 4.2. *Estimates of* $Q(\theta; \tilde{\theta})$, *and* σ_Q^2 *at* $\theta = \tilde{\theta} = (4.0, 1.5)$, *are based on* $n = 10^6$; *middle columns show half-width of* 95% *confidence interval* ($t_* = 1.960$) *and integrated autocorrelation time* (*ACT*). *Rightmost panel shows estimated ESEJD based on* $m = 10^3$ *replications*, *with standard errors in parentheses*

Algorithm	$\hat{Q}(\boldsymbol{\theta}\|\tilde{\boldsymbol{\theta}}; \boldsymbol{y})$	$\hat{\sigma}_Q^2$	$t_* \hat{\sigma}_Q / \sqrt{n}$	**ACT**	$\widehat{\textbf{ESEJD}}$
RW	−47.74	203.71	0.028	39.37	0.57 (0.0004)
MHIS	−47.77	2211.16	0.092	427.13	0.13 (0.0015)
CIS	−47.76	19.81	0.009	3.87	4.97 (0.0016)
RSIS	−47.76	258.85	0.032	50.08	0.50 (0.0005)

more than 10 times larger than that of CIS. The ACTs tell a similar story, RW and RSIS are comparable while MHIS is the worst and CIS is much better.

Estimation of ESEJD is based on the same $m = 10^3$ independent replications of RW, MHIS, RSIS and CIS for $n = 10^4$ iterations each. The results here are consistent with the other measures in the above discussion. The performance of MHIS is terrible, while RSIS and RW are comparable and CIS is the best of the four by a wide margin. The fact that RSIS is comparable to RW is surprising. In RSIS only one of the 10 components has a chance to be updated at each step, yet its performance is similar to a chain which updates all of its components about 30% of the time.

5. CONCLUDING REMARKS

Outside of the two-variable Gibbs sampler and the random scan Metropolis-within-Gibbs algorithms, there has been little research on convergence rates of component-wise MCMC samplers. This is unfortunate because, as outlined in Section 1, establishing geometric ergodicity is a key step in enabling a practitioner to have as much confidence in the simulation results as if the samples were independent and identically distributed.

Certainly a theme of this paper has been that studying the convergence rates of component-wise samplers formed by composition, that is, P_C, enables us to establish uniform or geometric ergodicity for other component-wise samplers such as P_{RQ} and P_{RS}. Indeed, we showed this is true for uniform ergodicity in the general setting and for geometric ergodicity in the two-variable setting. It seems that studying the convergence rates of P_{RQ} and P_{RS} should also inform us about the rate of P_C. Specifically, it is tempting to think that P_{RS} should converge no faster than P_{RQ} which should converge no faster than P_C. Especially in the two-variable setting, we suspect this is the case. Indeed, Tan, Jones and Hobert (2013) study a class of target distributions and show that either both GS and RSGS are geometrically ergodic or neither are.

Another theme has been that component-wise MCMC methods can be superior to full-dimensional updates. For example, full-dimensional MCMC methods often fail to be geometrically ergodic, but obvious component-wise implementations are. Also, the empirical investigations in Section 4 showed that the finite sample properties of component-wise methods were superior to full-dimensional methods in every case, which matches our observation in so many real data examples; see the references in Section 1. The near ubiquity of component-wise methods in the applied literature suggests that this view is widely held among MCMC practitioners.

ACKNOWLEDGMENTS

Galin L. Jones supported in part by the National Science Foundation and the National Institutes for Health. Ronald C. Neath was supported by a PSC-CUNY Award, jointly funded by The Professional Staff Congress and The City University of New York.

SUPPLEMENTARY MATERIAL

Supplementary material for "Component-Wise Markov Chain Monte Carlo: Uniform and Geometric Ergodicity Under Mixing and Composition" (DOI: 10.1214/13-STS423SUPP; .pdf). This supplemental article includes all technical details and proofs for the above results.

REFERENCES

ATCHADÉ, Y. F. (2011). Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. *Ann. Statist.* **39** 990–1011. MR2816345

BÉDARD, M. and ROSENTHAL, J. S. (2008). Optimal scaling of Metropolis algorithms: Heading toward general target distributions. *Canad. J. Statist.* **36** 483–503. MR2532248

BOOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 265–285.

CAFFO, B. S., JANK, W. and JONES, G. L. (2005). Ascent-based Monte Carlo expectation-maximization. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 235–251. MR2137323

CHRISTENSEN, O. F., MØLLER, J. and WAAGEPETERSEN, R. P. (2001). Geometric ergodicity of Metropolis–Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodol. Comput. Appl. Probab.* **3** 309–327. MR1891114

COULL, B. A., HOBERT, J. P., RYAN, L. M. and HOLMES, L. B. (2001). Crossed random effect models for multiple outcomes in a study of teratogenesis. *J. Amer. Statist. Assoc.* **96** 1194–1204. MR1946573

DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23** 151–178. MR2446500

DOSS, H. and HOBERT, J. P. (2010). Estimation of Bayes factors in a class of hierarchical random effects models using a geometrically ergodic MCMC algorithm. *J. Comput. Graph. Statist.* **19** 295–312. MR2675092

FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823

FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704

FLEGAL, J. M. and JONES, G. L. (2011). Implementing Markov chain Monte Carlo: Estimating with confidence. In *Handbook of Markov Chain Monte Carlo* (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.

FORT, G., MOULINES, E., ROBERTS, G. O. and ROSENTHAL, J. S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40** 123–146. MR1953771

GEYER, C. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry* (*Toulouse*, 1996) (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.). *Monographs on Statistics and Applied Probability* **80** 79–140. Chapman & Hall/CRC, Boca Raton, FL. MR1673118

HOBERT, J. P. (2000). Hierarchical models: A current computational perspective. *J. Amer. Statist. Assoc.* **95** 1312–1316. MR1825284

HOBERT, J. P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo* (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.

HOBERT, J. P. and GEYER, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J. Multivariate Anal.* **67** 414–430. MR1659196

HOBERT, J. P., JONES, G. L., PRESNELL, B. and ROSENTHAL, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* **89** 731–743. MR1946508

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076.

JOHNSON, A. A. and JONES, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Stat.* **4** 313–333. MR2645487

JOHNSON, A. A., JONES, G. L. and NEATH, R. C. (2013). Supplement to "Component-Wise Markov Chain Monte Carlo: Uniform and Geometric Ergodicity Under Mixing and Composition." DOI:10.1214/13-STS423SUPP.

JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447

JONES, G. L. and HOBERT, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann. Statist.* **32** 784–817. MR2060178

JONES, G. L., ROBERTS, G. O. and ROSENTHAL, J. S. (2013). Convergence of conditional Metropolis–Hastings samplers. *J. Appl. Probab.* To appear.

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478

ŁATUSZYNSKI, K., ROBERTS, G. O. and ROSENTHAL, J. S. (2013). Adaptive Gibbs samplers and related MCMC methods. *Ann. Appl. Probab.* **23** 66–98.

LEE, K. J., JONES, G. L., CAFFO, B. S. and BASSETT, S. (2013). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. Preprint.

MARCHEV, D. and HOBERT, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *J. Amer. Statist. Assoc.* **99** 228–238. MR2054301

MCCULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Amer. Statist. Assoc.* **92** 162–170. MR1436105

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

MEYN, S. P. and TWEEDIE, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770

NEAL, P. and ROBERTS, G. (2006). Optimal scaling for partially updating MCMC algorithms. *Ann. Appl. Probab.* **16** 475–515. MR2244423

NEATH, R. C. (2013). On convergence properties of the Monte Carlo EM algorithm. In *Advances in Modern Statistical Theory and Applications*: *A Festschrift in Honor of Morris L. Eaton* (G. L. Jones and X. Shen, eds.). To appear.

PAPASPILIOPOULOS, O. and ROBERTS, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Ann. Statist.* **36** 95–117. MR2387965

ROBERT, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statist. Sci.* **10** 231–253. MR1390517

ROBERTS, G. O. and POLSON, N. G. (1994). On the geometric convergence of the Gibbs sampler. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **56** 377–384. MR1281941

ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25 (electronic). MR1448322

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Two convergence properties of hybrid samplers. *Ann. Appl. Probab.* **8** 397–407. MR1624941

ROBERTS, G. O. and ROSENTHAL, J. S. (1999). Convergence of slice sampler Markov chains. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 643–660. MR1707866

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **59** 291–317. MR1440584

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

ROMÁN, J. C. (2012). Convergence analysis of block Gibbs samplers for Bayesian general linear mixed models. Ph.D. thesis, Dept. Statistics, Univ. Florida.

ROMÁN, J. C. and HOBERT, J. P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *Ann. Statist.* **40** 2823–2849.

ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

ROSENTHAL, J. S. (1996). Analysis of the Gibbs sampler for a model related to James–Stein estimators. *Statist. Comput.* **6** 269–275.

ROSENTHAL, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo* (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.

ROY, V. and HOBERT, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 607–623. MR2370071

TAN, A. and HOBERT, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. MR2598033

TAN, A., JONES, G. L. and HOBERT, J. P. (2013). On the geometric ergodicity of two-variable Gibbs samplers. In *Advances in Modern Statistical Theory and Applications*: *A Festschrift in Honor of Morris L. Eaton* (G. L. Jones and X. Shen, eds.). To appear.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528–550. MR0898357

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22** 1701–1762. MR1329166