Adv. Appl. Prob. **46,** 422–445 (2014)
Printed in Northern Ireland

CONVERGENCE OF CONDITIONAL METROPOLIS–HASTINGS SAMPLERS

GALIN L. JONES,* *University of Minnesota*
GARETH O. ROBERTS,** *University of Warwick*
JEFFREY S. ROSENTHAL,*** *University of Toronto*

Abstract

We consider Markov chain Monte Carlo algorithms which combine Gibbs updates with Metropolis–Hastings updates, resulting in a *conditional Metropolis–Hastings sampler* (CMH sampler). We develop conditions under which the CMH sampler will be geometrically or uniformly ergodic. We illustrate our results by analysing a CMH sampler used for drawing Bayesian inferences about the entire sample path of a diffusion process, based only upon discrete observations.

Keywords: Markov chain Monte Carlo algorithm; independence sampler; Gibbs sampler; geometric ergodicity; convergence rate

2010 Mathematics Subject Classification: Primary 60J05; 60J22; 65C40; 62F15

1. Introduction

Markov chain Monte Carlo (MCMC) algorithms are an extremely popular way of approximately sampling from complicated probability distributions (see, e.g. [1], [6], [29], and [42]). In multivariate settings it is common to update the different components individually. If these updates are all drawn from full-conditional distributions then this corresponds to the *Gibbs sampler*. Conversely, if these updates are produced by drawing from a proposal distribution and then either accepting or rejecting the proposed state, then this corresponds to the *componentwise Metropolis–Hastings algorithm* (sometimes called the *Metropolis–Hastings-within-Gibbs*). We consider the mixed case in which some components are updated as in the Gibbs sampler, while other components are updated as in componentwise Metropolis–Hastings. Such chains arise when full-conditional updates are feasible for some components but not for others, which is true of the discretely observed diffusion example considered in Section 5.

For this mixed case, we shall prove various results about theoretical properties such as *geometric ergodicity*. Geometric ergodicity is an important stability property for MCMC, used, e.g. to establish central limit theorems [2], [11], [25] and to calculate asymptotically valid Monte Carlo standard errors [5], [14]. While there has been much progress in proving geometric ergodicity for many MCMC samplers (see, e.g. [7], [8], [9], [13], [17], [18], [24], [27], [31], [33], [36], [37], [41]), doing so typically requires difficult theoretical analysis.

Received 21 June 2012; revision received 25 June 2013.
* Postal address: School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.
Email address: galin@umn.edu
Partially supported by the National Institutes for Health.
** Postal address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.
Email address: gareth.o.roberts@warwick.ac.uk
*** Postal address: Department of Statistics, University of Toronto, Toronto, Ontario, M5S 3G3, Canada.
Email address: jeff@math.toronto.edu
Partially supported by the Natural Sciences and Engineering Research Council of Canada.

For ease of exposition, we begin with the two-variable case and defer consideration of extensions to more than two variables to Section 4. Let π be a probability distribution having support $\mathcal{X} \times \mathcal{Y}$, and let $\pi_{X \mid Y}$ and $\pi_{Y \mid X}$ denote the associated conditional distributions. Suppose that $\pi_{Y \mid X}$ has a density $f_{Y \mid X}$, and that $\pi_{X \mid Y}$ has density a $f_{X \mid Y}$. There are several potential componentwise MCMC algorithms, each having π as its invariant distribution. If it is possible to simulate from $\pi_{X \mid Y}$ and $\pi_{Y \mid X}$, then one can implement a deterministic-scan Gibbs sampler, which we now describe. Suppose that the current state of the chain is $(X_n, Y_n) = (x, y)$. Then the next state, (X_{n+1}, Y_{n+1}), is obtained as follows.

Algorithm 1. (*Iteration* $n+1$ *of the deterministic-scan Gibbs sampler (DUGS).*)

1. Draw $Y_{n+1} \sim \pi_{Y \mid X}(\cdot \mid x)$, and call the observed value y'.
2. Draw $X_{n+1} \sim \pi_{X \mid Y}(\cdot \mid y')$.

However, sometimes one or both of these steps will be computationally infeasible, necessitating the use of alternative algorithms. In particular, suppose that we continue to simulate directly from $\pi_{Y \mid X}$, but use a Metropolis–Hastings algorithm for $\pi_{X \mid Y}$ with proposal density $q(x' \mid x, y')$. This results in a *conditional Metropolis–Hastings sampler*, which we now describe. If the current state of the chain is $(X_n, Y_n) = (x, y)$ then the next state, (X_{n+1}, Y_{n+1}), is obtained as follows.

Algorithm 2. (*Iteration* $n+1$ *of the conditional Metropolis–Hastings (CMH) sampler.*)

1. Draw $Y_{n+1} \sim \pi_{Y \mid X}(\cdot \mid x)$, and call the observed value y'.
2. Draw $V \sim q(\cdot \mid x, y')$, and call the observed value v. Independently draw $U \sim \text{Uniform}(0, 1)$. Set $X_{n+1} = v$ if

$$U \le \frac{f_{X \mid Y}(v \mid y') q(x \mid v, y')}{f_{X \mid Y}(x \mid y') q(v \mid x, y')};$$

   otherwise, set $X_{n+1} = X_n$

As is well known, DUGS is a special case of the CMH sampler where the proposal is taken to be the conditional, that is, $q(x' \mid x, y') = f_{X \mid Y}(x' \mid y')$ [29]. Thus, it is natural to suspect that the convergence properties of DUGS and the CMH sampler may be related. On the other hand, while geometric ergodicity of the Gibbs sampler has been extensively studied [17], [21], [24], [31], the CMH sampler has received comparatively little attention [10].

If the proposal distribution for x' does not depend on the previous value of x, i.e. if $q(x' \mid x, y') = q(x' \mid y')$, then in the CMH sampler the X values are updated as in an independence sampler (see, e.g. [30] and [42]), conditional on the current value of Y. We thus refer to this special case as a *conditional independence sampler* (CIS). It is known that an independence sampler will be uniformly ergodic provided that the ratio of the target density to the proposal density is bounded [16], [19], [33], [40]. Intuitively, this suggests that the resulting CIS will have convergence properties similar to those of the corresponding DUGS; we will explore this question herein.

This paper is organised as follows. In Section 2 we present preliminary material, including a general Markov chain comparison theorem (Theorem 1). In Section 3 we derive various convergence properties of the CMH sampler, including uniform ergodicity in terms of the conditional weight function (Theorems 2 and 3) and uniform return probabilities (Theorem 4), and geometric ergodicity via a comparison to DUGS (Theorem 5). In Section 4 we extend many of our results from the two-variable setting to higher dimensions. Finally, in Section 5 we apply our results to an algorithm for drawing Bayesian inferences about the entire sample path of a diffusion process based only upon discrete observations.

Remark 1. The focus of our paper is on qualitative convergence properties such as uniform and geometric ergodicity. However, a careful look at the proofs will show that many of our results actually provide explicit quantitative bounds on spectral gaps or minorisation constants for the algorithms that we consider.

2. Preliminaries

We begin with an account of essential preliminary material.

2.1. Background about Markov chains

Let P be a Markov transition kernel on a measurable space $(\mathcal{Z}, \mathcal{F})$. Thus, $P\colon \mathcal{Z} \times \mathcal{F} \to [0, 1]$ such that, for each $A \in \mathcal{F}$, $P(\cdot, A)$ is a measurable function, and, for each $z \in \mathcal{Z}$, $P(z, \cdot)$ is a probability measure. If $\Phi = \{Z_0, Z_1, \dots\}$ is the Markov chain with transitions governed by P then, for any positive integer n, the n-step Markov transition kernel is given by $P^n(z, A) = \Pr(Z_{n+j} \in A \mid Z_j = z)$, which is assumed to be the same for all times j.

Let ν be a measure on $(\mathcal{Z}, \mathcal{F})$ and $A \in \mathcal{F}$, and define

$$\nu P(A) = \int \nu(\mathrm{d}z) P(z, A)$$

so that P acts to the left on measures. Let π be an invariant probability measure for P, that is, $\pi P = \pi$. Also, if f is a measurable function on $\mathcal{Z}$, let

$$Pf(z) = \int f(y) P(z, \mathrm{d}y)$$

and

$$\pi(f) = \int f(z) \pi(\mathrm{d}z).$$

Let $\|P^n(z, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} = \sup_{A \in \mathcal{F}} |P^n(z, A) - \pi(A)|$ be the usual total variation distance. Then P is *geometrically ergodic* if there exists a real-valued function $M(z)$ on $\mathcal{Z}$ and $0 < t < 1$ such that, for π-almost every $z \in \mathcal{Z}$,

$$\|P^n(z, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \le M(z) t^n. \tag{1}$$

Moreover, P is *uniformly ergodic* if (1) holds and $\sup_z M(z) < \infty$.

Uniform ergodicity is equivalent to a so-called minorisation condition (see, e.g. [20] and [29]). That is, P is uniformly ergodic if and only if there exists a positive integer $m \ge 1$, a constant $\varepsilon > 0$, and a probability measure Q on $\mathcal{Z}$ such that, for all $z \in \mathcal{Z}$,

$$P^m(z, A) \ge \varepsilon Q(A), \qquad A \in \mathcal{F}, \tag{2}$$

in which case we say that P is *m-minorisable*.

Establishing geometric ergodicity is most commonly done by establishing various Foster–Lyapounov criteria [12], [20], [29], but these will play no role here. Instead we will focus on another characterisation of geometric ergodicity that is appropriate for reversible Markov chains.

Let $L^2(\pi)$ be the space of measurable functions that are square integrable with respect to the invariant distribution, and let

$$L^2_{0,1}(\pi) = \{f \in L^2(\pi) \colon \pi(f) = 0 \text{ and } \pi(f^2) = 1\}.$$

For $f, g \in L^2(\pi)$, define the inner product as

$$(f, g) = \int_{\mathcal{Z}} f(z)g(z)\pi(\mathrm{d}z)$$

and $\|f\|^2 = (f, f)$. The *norm* of the operator P (restricted to $L^2_{0,1}(\pi)$) is

$$\|P\| = \sup_{f \in L^2_{0,1}(\pi)} \|Pf\|.$$

If P is *reversible* with respect to π, that is, if

$$P(z, \mathrm{d}z')\pi(\mathrm{d}z) = P(z', \mathrm{d}z)\pi(\mathrm{d}z'), \tag{3}$$

then P is self-adjoint so that $(Ph_1, h_2) = (h_1, Ph_2)$. In this case,

$$\|P\| = \sup_{f \in L^2_{0,1}(\pi)} |(Pf, f)|. \tag{4}$$

Let P_0 denote the restriction of P to $L^2_{0,1}(\pi)$, and let $\sigma(P_0)$ be the spectrum of P_0. The *spectral radius* of P_0 is

$$r(P_0) = \sup\{|\lambda| \colon \lambda \in \sigma(P_0)\},$$

while the *spectral gap* of P is $\mathrm{gap}(P) = 1 - r(P_0)$. If P is reversible with respect to π and, hence, self-adjoint, then $\sigma(P_0) \subseteq [-1, 1]$, and also $r(P_0) = \|P\|$ (since we defined $\|P\|$ as being with respect to $L^2_{0,1}(\pi)$ only). Finally, if P is reversible with respect to π then P is geometrically ergodic if and only if $\mathrm{gap}(P) > 0$, or, equivalently, $\|P\| < 1$ [25].

2.2. A comparison theorem

Our goal in this section is to develop and prove a simple but powerful comparison result, similar in spirit to [3] and to Peskun orderings [22], [43], which we shall use in the sequel to help establish uniform and geometric ergodicity of the CMH sampler.

Theorem 1. *Suppose that P and Q are Markov kernels and that there exists $\delta > 0$ such that*

$$P(z, A) \geq \delta Q(z, A), \qquad A \in \mathcal{F},\ z \in \mathcal{Z}. \tag{5}$$

(i) *If P and Q have invariant distribution π and Q is uniformly ergodic, then so is P.*

(ii) *If P and Q are reversible with respect to π and Q is geometrically ergodic, then so is P.*

Proof. (i) Note that (5) implies that, for all n,

$$P^n(z, A) \geq \delta^n Q^n(z, A), \qquad A \in \mathcal{F},\ z \in \mathcal{Z}.$$

Since Q is uniformly ergodic, by (2) there exists an integer $m \geq 1$, $\epsilon > 0$, and probability measure ν such that

$$Q^m(z, A) \geq \epsilon\nu(A), \qquad A \in \mathcal{F},\ z \in \mathcal{Z}.$$

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

Putting these two observations together gives a minorisation condition for P, and, hence, yields the claim in (2).

(ii) Let $A \in \mathcal{F}$, and define

$$R(z, A) = \frac{P(z, A) - \delta Q(z, A)}{1 - \delta}.$$

Using (5) shows that R is a Markov kernel. Also,

$$P(z, A) = \delta Q(z, A) + (1 - \delta) R(z, A).$$

Let P_0, Q_0, and R_0 denote the restrictions of P, Q, and R, respectively, to $L^2_{0,1}(\pi)$. Since P is reversible with respect to π, and $\|R\| \leq 1$ so $r(R_0) \leq 1$, we have, by (4),

$$\begin{aligned}
r(P_0) &= r\big(\delta Q_0 + (1-\delta) R_0\big) \\
&= \sup_{f \in L^2_{0,1}(\pi)} |\delta (Q_0 f, f) + (1-\delta)(R_0 f, f)| \\
&\leq \delta \Big[\sup_{f \in L^2_{0,1}(\pi)} |(Q_0 f, f)| \Big] + (1-\delta) \Big[\sup_{f \in L^2_{0,1}(\pi)} |(R_0 f, f)| \Big] \\
&= \delta\, r(Q_0) + (1-\delta) r(R_0) \\
&\leq \delta\, r(Q_0) + (1-\delta).
\end{aligned}$$

Hence,

$$\operatorname{gap}(P) = 1 - r(P_0) \geq 1 - [\delta\, r(Q_0) + (1-\delta)] = \delta[1 - r(Q_0)] = \delta \operatorname{gap}(Q).$$

Since Q is geometrically ergodic, $\operatorname{gap}(Q) > 0$, and, hence, $\operatorname{gap}(P) > 0$. Therefore, P is geometrically ergodic.

2.3. The Markov chain kernels

We formally define the Markov chain kernels for the various algorithms described in Section 1. While we focus on the case of two variables here and in Section 3, in Section 4 we consider extensions to more general settings.

Let $(\mathcal{X}, \mathcal{F}_X, \mu_X)$ and $(\mathcal{Y}, \mathcal{F}_Y, \mu_Y)$ be two σ-finite measure spaces, and let $(\mathcal{Z}, \mathcal{F}, \mu)$ be their product space. Let π be a probability distribution on $(\mathcal{Z}, \mathcal{F}, \mu)$ which has a density $f(x, y)$ with respect to μ. Then the marginal distributions π_X and π_Y of π have densities given by

$$f_X(x) = \int_{\mathcal{Y}} f(x, y) \mu_Y(\mathrm{d}y) \tag{6}$$

and similarly for $f_Y(y)$. By redefining $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can (and do) assume that

$$f_X(x) > 0 \quad \text{for all } x \in \mathcal{X} \quad \text{and} \quad f_Y(y) > 0 \quad \text{for all } y \in \mathcal{Y}. \tag{7}$$

The corresponding conditional densities are then given by $f_{X \mid Y}(x \mid y) = f(x, y)/f_Y(y)$ and $f_{Y \mid X}(y \mid x) = f(x, y)/f_X(x)$.

Define a Markov kernel for a Y update by

$$P_{\mathrm{GS}:Y}(x, A) = \int_{\{y\colon (x,y) \in A\}} f_{Y \mid X}(y \mid x) \mu_Y(\mathrm{d}y),$$

and similarly an X update is described by the Markov kernel

$$P_{\mathrm{GS}:X}(y, A) = \int_{\{x:\,(x,y)\in A\}} f_{X\,|\,Y}(x \mid y)\mu_X(\mathrm{d}x).$$

We can define the Markov kernel for the DUGS by the composition of X and Y updates, i.e. $P_{\mathrm{DUGS}} = P_{\mathrm{GS}:Y} P_{\mathrm{GS}:X}$ which corresponds to doing first a Gibbs sampler Y-move and then a Gibbs sampler X-move. That is, the DUGS Markov chain updates first Y and then X: schematically, $(x, y) \to (x, y') \to (x', y')$. If $k_{\mathrm{DUGS}}(x', y' \mid x, y) = f_{Y\,|\,X}(y' \mid x) f_{X\,|\,Y}(x' \mid y')$ then we can also write this as

$$P_{\mathrm{DUGS}}((x, y), A) = \int_A k_{\mathrm{DUGS}}(x', y' \mid x, y)\mu(\mathrm{d}(x', y')), \qquad A \in \mathcal{F}.$$

Note that $\pi P_{\mathrm{DUGS}} = \pi$, i.e. π is a stationary distribution for P_{DUGS}, although P_{DUGS} is not reversible with respect to π. Also, note that DUGS depends on the current state (x, y) only through x. For DUGS, the following simple lemma is sometimes useful (and will be applied in Section 5).

Proposition 1. *If the Y-update of P_{DUGS} is 1-minorisable, in the sense that there exists a $\epsilon > 0$ and a probability measure ν such that $P_{\mathrm{GS}:Y}(x, A) \geq \epsilon\nu(A)$ for all x and A, then P_{DUGS} is 1-minorisable.*

Proof. The result follows from noting that

$$P_{\mathrm{DUGS}}((x, y), A \times B) \geq \epsilon \int_B \nu(\mathrm{d}y') P_{\mathrm{GS}:X}(y', A),$$

which is a 1-minorisation of P_{DUGS} as claimed.

Remark 2. We could have considered the alternative update order $(x, y) \to (x', y) \to (x', y')$, resulting in the Markov kernel $P^*_{\mathrm{DUGS}} = P_{\mathrm{GS}:X} P_{\mathrm{GS}:Y}$, which will play a role in Section 3.2. Note that, with essentially the same argument as in Proposition 1, if the X-update is 1-minorisable then so is P^*_{DUGS}.

A related algorithm, the *random-scan Gibbs sampler* (RSGS) with selection probability $p \in (0, 1)$ proceeds by either updating $Y \sim P_{\mathrm{GS}:Y}$ with probability p, or updating $X \sim P_{\mathrm{GS}:X}$ with probability $1 - p$. The RSGS has kernel

$$P_{\mathrm{RSGS}} = p P_{\mathrm{GS}:Y} + (1 - p) P_{\mathrm{GS}:X},$$

i.e.

$$P_{\mathrm{RSGS}}((x, y), A) = p P_{\mathrm{GS}:Y}(x, A) + (1 - p) P_{\mathrm{GS}:X}(y, A).$$

It follows that P_{RSGS} is reversible with respect to π. Furthermore, it is well known (see, e.g. [10] and [25]) that if P_{DUGS} is uniformly ergodic then so is P_{RSGS} (as follows immediately from (2), since we always have $P^{2n}_{\mathrm{RSGS}}(z, A) \geq (p(1 - p))^n P^n_{\mathrm{DUGS}}(z, A)$). We also have the following result.

Proposition 2. *If P_{RSGS} is geometrically ergodic for some selection probability p^* then it is geometrically ergodic for all selection probabilities $p \in (0, 1)$.*

Proof. For $p \in (0, 1)$, let $P_{\mathrm{RSGS},p}$ be the RSGS kernel using selection probability p, so that if $A \in \mathcal{F}$ then

$$P_{\mathrm{RSGS},p}((x, y), A) = p P_{\mathrm{GS}:Y}(x, A) + (1 - p) P_{\mathrm{GS}:X}(y, A).$$

It follows immediately that

$$P_{\mathrm{RSGS},p} \geq \left(\frac{p}{p^*} \wedge \frac{1-p}{1-p^*}\right) P_{\mathrm{RSGS},p^*}.$$

Since $P_{\mathrm{RSGS},p}$ and P_{RSGS,p^*} are each reversible with respect to π, the claim follows from Theorem 1.

Next, consider the deterministically updated CMH sampler which first updates Y with a Gibbs update, and then updates X with a Metropolis–Hastings update: schematically, $(x, y) \to (x, y') \to (x', y')$. In this case, the Y update follows precisely the same kernel $P_{\mathrm{GS}:Y}$ as above. To define the X update, let $q(x' \mid x, y')$ be a proposal density, and set

$$\alpha(x', x, y') = \left[1 \wedge \frac{f_{X \mid Y}(x' \mid y') q(x \mid x', y')}{f_{X \mid Y}(x \mid y') q(x' \mid x, y')}\right]$$

and

$$r(x, y') = 1 - \int q(x' \mid x, y') \alpha(x', x, y') \mu_X(\mathrm{d}x').$$

Then the X update follows the Markov kernel defined by

$$P_{\mathrm{MH}:X}((x, y'), A) = \int_{\{x' :\, (x', y') \in A\}} q(x' \mid x, y') \alpha(x', x, y') \mu_X(\mathrm{d}x') + r(x, y') \mathbf{1}_{\{(x, y') \in A\}}.$$

By construction, $P_{\mathrm{MH}:X}$ is reversible with respect to π (though it only updates the x coordinate, while leaving the y coordinate fixed).

In terms of these individual kernels, we can define the Markov kernel for the CMH sampler by their composition, corresponding to doing first a Gibbs sampler Y-move and then a Metropolis–Hastings X-move:

$$P_{\mathrm{CMH}} = P_{\mathrm{GS}:Y} P_{\mathrm{MH}:X}.$$

It then follows that $\pi P_{\mathrm{CMH}} = \pi$, but P_{CMH} is not reversible with respect to π. It is also important to note that, because of the update order we are using, P_{CMH} depends on the current state (x, y) only through x. Finally, if

$$k_{\mathrm{CMH}}(x', y' \mid x, y) = f_{Y \mid X}(y' \mid x) q(x' \mid x, y') \alpha(x', x, y')$$

then by construction we have

$$P_{\mathrm{CMH}}((x, y), A) \geq \int_A k_{\mathrm{CMH}}(x', y' \mid x, y) \mu(\mathrm{d}(x', y')), \qquad A \in \mathcal{F}.$$

We will also consider the random-scan CMH (RCMH) sampler. For any fixed selection probability $p \in (0, 1)$, the RCMH sampler is the algorithm which selects the Y coordinate with probability p, or selects the X coordinate with probability $1 - p$, and then updates the selected coordinate as in the CMH algorithm (i.e. from a full-conditional distribution for Y, or from a

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

conditional Metropolis–Hastings step for X), while leaving the other coordinate unchanged. Hence, its kernel is given by

$$P_{\mathrm{RCMH}} = pP_{\mathrm{GS}:Y} + (1-p)P_{\mathrm{MH}:X}.$$

Then P_{RCMH} is reversible with respect to π. A similar argument to that given above relating the uniform ergodicity of P_{DUGS} to that of P_{RSGS} shows that, if P_{CMH} is uniformly ergodic then so is P_{RCMH} for any selection probabilities [10, Theorem 2].

If the proposal distribution for x' does not depend on the previous value of x, i.e. if $q(x' \mid x, y') = q(x' \mid y')$, then the CMH algorithm becomes the CIS. In this case, we will continue to use all the same notation as for the CMH sampler above, except omitting the unnecessary x arguments.

2.4. Embedded X-chains

When studying geometric ergodicity, Theorem 1(ii) does not apply directly to P_{DUGS} and P_{CMH} since they are not reversible with respect to π. However, each of these samplers does produce marginal X-sequences which are reversible with respect to the marginal distribution π_X (with density as in (6)). Moreover, as we discuss below, if either of these X-sequences is geometrically ergodic then so is the corresponding parent sampler. For this reason, it is sometimes useful to study the marginal X-sequences embedded within these Markov chains.

Consider the DUGS Markov chain. Define

$$k_X(x' \mid x) = \int_{\mathcal{Y}} f_{X\mid Y}(x' \mid y) f_{Y\mid X}(y \mid x)\mu_Y(\mathrm{d}y),$$

and note that the marginal sequence $\{X_0, X_1, \dots\}$ is a Markov chain having kernel

$$P^X_{\mathrm{DUGS}}(x, A) = \int_A k_X(x' \mid x)\mu_X(\mathrm{d}x'), \qquad A \in \mathcal{F}_X.$$

Now P_{DUGS} has π as its invariant distribution while P^X_{DUGS} has the marginal distribution π_X as its invariant distribution and, in fact, P^X_{DUGS} is reversible with respect to π_X. Moreover, it is well known that P_{DUGS} and P^X_{DUGS} converge to their respective invariant distributions at the same rate [17], [23], [28]. This has been routinely exploited in the analysis of two-variable Gibbs samplers where P^X_{DUGS} may be much easier to analyze than P_{DUGS}.

Now consider the CMH algorithm, and let its resulting values be $Y_0, X_0, Y_1, X_1, Y_2, X_2, \dots$. This sequence in turn provides a marginal sequence, $X_0, X_1, \dots$, which is itself a Markov chain on $\mathcal{X}$, since the $P_{\mathrm{GS}:Y}$ update within the CMH algorithm depends only on the previous X value, not on the previous Y value, and, hence, the future chain values depend only on the current value of X, not the current value of Y. (This is a somewhat subtle point which would *not* be true if the CMH algorithm were instead defined to update first X and then Y.) Thus, this marginal X-sequence has its own Markov transition kernel on $(\mathcal{X}, \mathcal{F}_X)$, say $P^X_{\mathrm{CMH}}(x, A)$, and if

$$h_X(x' \mid x) = \int_{\mathcal{Y}} f_{Y\mid X}(y' \mid x) q(x' \mid x, y')\alpha(x', x, y')\mu_Y(\mathrm{d}y'),$$

it follows by construction that

$$P^X_{\mathrm{CMH}}(x, A) \ge \int_A h_X(x' \mid x)\mu_X(\mathrm{d}x'), \qquad A \in \mathcal{F}_X.$$

Note that P_{CMH} and P_{CMH}^X have invariant distributions π and π_X, respectively. Now P_{CMH} is not reversible with respect to π, but we shall show that P_{CMH}^X is reversible with respect to π_X. Indeed, first note that, by construction,

$$P_{\text{MH}:X}((x, y), (\mathrm{d}x', y))\pi_{X \mid Y}(\mathrm{d}x \mid y) = P_{\text{MH}:X}((x', y), (\mathrm{d}x, y))\pi_{X \mid Y}(\mathrm{d}x' \mid y).$$

Now we compute

$$\begin{aligned} P_{\text{CMH}}^X(x, \mathrm{d}x')\pi_X(\mathrm{d}x) &= \pi_X(\mathrm{d}x) \int_{\mathcal{Y}} P_{\text{MH}:X}((x, y), (\mathrm{d}x', y))\pi_{Y \mid X}(\mathrm{d}y \mid x) \\ &= \int_{\mathcal{Y}} P_{\text{MH}:X}((x, y), (\mathrm{d}x', y))\pi(\mathrm{d}x, \mathrm{d}y) \\ &= \int_{\mathcal{Y}} P_{\text{MH}:X}((x, y), (\mathrm{d}x', y))\pi_{X \mid Y}(\mathrm{d}x \mid y)\pi_Y(\mathrm{d}y) \\ &= \int_{\mathcal{Y}} P_{\text{MH}:X}((x', y), (\mathrm{d}x, y))\pi_{X \mid Y}(\mathrm{d}x' \mid y)\pi_Y(\mathrm{d}y) \\ &= \int_{\mathcal{Y}} P_{\text{MH}:X}((x', y), (\mathrm{d}x, y))\pi(\mathrm{d}x', \mathrm{d}y) \\ &= \pi_X(\mathrm{d}x') \int_{\mathcal{Y}} P_{\text{MH}:X}((x', y), (\mathrm{d}x, y))\pi_{Y \mid X}(\mathrm{d}y \mid x') \\ &= P_{\text{CMH}}^X(x', \mathrm{d}x)\pi_X(\mathrm{d}x'), \end{aligned}$$

and conclude that P_{CMH}^X is reversible with respect to π_X.

It is straightforward to see that, in the language of [28], the embedded chain P_{CMH}^X is *deinitialising* for P_{CMH}. This implies that if P_{CMH}^X is geometrically (or uniformly) ergodic then P_{CMH} is geometrically (or uniformly) ergodic [28, Theorem 1]. In fact, it is not too hard to show the converse [10] and conclude that P_{CMH}^X is geometrically (or uniformly) ergodic if and only if P_{CMH} is geometrically (or uniformly) ergodic.

3. Ergodicity properties of the CMH sampler

Our goal in this section is to derive ergodicity properties of the CMH sampler in terms of those of the corresponding Gibbs sampler. We focus on the case of two variables; this is done mainly for ease of exposition, and we will see in Section 4 that many of the results carry over to a more general setting.

3.1. Uniform ergodicity of the CMH sampler via the weight function

Analogous to previous studies of the usual full-dimensional independence sampler [16], [19], [33], [40], we define the *(conditional) weight function* by

$$w(x', x, y') := \frac{f_{X \mid Y}(x' \mid y')}{q(x' \mid x, y')}, \qquad x', x \in \mathcal{X},\ y' \in \mathcal{Y}.$$

(In the case of CIS, the weight function reduces to $w(x', y') = f_{X \mid Y}(x' \mid y')/q(x' \mid y')$.) We shall see that these weight functions are key to understanding the ergodicity properties of the CMH sampler. We begin with a simple lemma.

Lemma 1. *It holds that*

$$k_{\mathrm{CMH}}(x', y' \mid x, y) = k_{\mathrm{DUGS}}(x', y' \mid x, y)\left[\frac{1}{w(x', x, y')} \wedge \frac{1}{w(x, x', y')}\right].$$

Proof. Note that

$$\begin{aligned} k_{\mathrm{CMH}}(x', y' \mid x, y) &= f_{Y \mid X}(y' \mid x) q(x' \mid x, y') \alpha(x', x, y') \\ &= f_{Y \mid X}(y' \mid x) f_{X \mid Y}(x' \mid y')\left[\frac{q(x' \mid x, y')}{f_{X \mid Y}(x' \mid y')} \wedge \frac{q(x \mid x', y')}{f_{X \mid Y}(x \mid y')}\right] \\ &= k_{\mathrm{DUGS}}(x', y' \mid x, y)\left[\frac{1}{w(x', x, y')} \wedge \frac{1}{w(x, x', y')}\right]. \end{aligned}$$

Say that w is *bounded* if

$$\sup_{x', x, y'} w(x', x, y') < \infty$$

and is X*-bounded* if there exists $C\colon \mathcal{Y} \to (0, \infty)$ such that

$$\sup_{x', x} w(x', x, y') \le C(y'), \qquad y' \in \mathcal{Y}.$$

We then have the following result.

Theorem 2. *If w is bounded and P_{DUGS} is uniformly ergodic, then P_{CMH} is uniformly ergodic.*

Proof. By Lemma 1 we have

$$k_{\mathrm{CMH}}(x', y' \mid x, y) = k_{\mathrm{DUGS}}(x', y' \mid x, y)\left[\frac{1}{w(x', x, y')} \wedge \frac{1}{w(x, x', y')}\right].$$

Since w is bounded, there exists a constant $C < \infty$ such that

$$k_{\mathrm{CMH}}(x', y' \mid x, y) \ge \frac{1}{C} k_{\mathrm{DUGS}}(x', y' \mid x, y),$$

and, hence,

$$P_{\mathrm{CMH}}((x, y), A) \ge \frac{1}{C} P_{\mathrm{DUGS}}((x, y), A), \qquad A \in \mathcal{F}.$$

The result now follows from Theorem 1.

As noted above, uniform ergodicity of deterministic-scan algorithms immediately implies uniform ergodicity of the corresponding random-scan algorithm, so we immediately obtain the following result.

Corollary 1. *If w is bounded and P_{DUGS} is uniformly ergodic, then P_{RCMH} is uniformly ergodic for any selection probability $p \in (0, 1)$.*

The condition on w in Theorem 2 can be weakened if we strengthen the assumption on the Gibbs sampler.

Theorem 3. *Suppose that w is X-bounded, and that there exists a nonnegative function g on $\mathcal{Z}$, with $\mu\{(x, y)\colon g(x, y) > 0\} > 0$, such that, for all x and y,*

$$k_{\mathrm{DUGS}}(x', y' \mid x, y) \ge g(x', y'). \tag{8}$$

Then P_{CMH} is uniformly ergodic.

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

Proof. By Lemma 1 we have

$$k_{\mathrm{CMH}}(x', y' \mid x, y) = k_{\mathrm{DUGS}}(x', y' \mid x, y)\left[\frac{1}{w(x', x, y')} \wedge \frac{1}{w(x, x', y')}\right].$$

That w is X-bounded implies that there exists a $C\colon \mathcal{Y} \to (0, \infty)$ such that

$$k_{\mathrm{CMH}}(x', y' \mid x, y) \geq \frac{1}{C(y')} k_{\mathrm{DUGS}}(x', y' \mid x, y),$$

and, using (8), we obtain

$$k_{\mathrm{CMH}}(x', y' \mid x, y) \geq \frac{g(x', y')}{C(y')}.$$

Letting

$$\epsilon = \int_{\mathcal{X}\times\mathcal{Y}} \frac{g(x, y)}{C(y)} \mu(\mathrm{d}(x, y)) > 0 \quad \text{and} \quad h(x, y) = \epsilon^{-1}\frac{g(x, y)}{C(y)},$$

we have

$$P_{\mathrm{CMH}}((x, y), A) \geq \epsilon \int_A h(u, v)\mu(\mathrm{d}(u, v)), \qquad A \in \mathcal{F}.$$

That is, P_{CMH} is 1-minorisable and, hence, is uniformly ergodic.

Remark 3. Note that condition (8) implies that P_{DUGS} is 1-minorisable.

Once again, the corresponding random-scan result follows immediately.

Corollary 2. *If w is X-bounded, and condition (8) holds, then P_{RCMH} is uniformly ergodic for any selection probability $p \in (0, 1)$.*

3.2. A counterexample

In this section we show that Theorem 3 might not hold if P_{DUGS} is just 2-minorisable (as opposed to 1-minorisable). We begin with a lemma about interchanging the update orders for Gibbs samplers. Specifically, define the Markov kernel P^*_{DUGS} to represent the Gibbs sampler which updates first X and then Y: $(x, y) \to (x', y) \to (x', y')$. This kernel has transition density

$$k^*_{\mathrm{DUGS}}(x', y' \mid x, y) = f_{X \mid Y}(x' \mid y) f_{Y \mid X}(y' \mid x').$$

Lemma 2 below shows that we can convert a 1-minorisation for P^*_{DUGS} into a 2-minorisation for P_{DUGS}.

Lemma 2. *Suppose that there exists a nonnegative function g on $\mathcal{Z}$, with $\mu\{(x, y)\colon g(x, y) > 0\} > 0$, such that, for all x and y,*

$$k^*_{\mathrm{DUGS}}(x', y' \mid x, y) \geq g(x', y').$$

Then there exists $\epsilon > 0$, and a probability measure ν on $\mathcal{Z}$, such that, for all x and y,

$$P^2_{\mathrm{DUGS}}((x, y), A) \geq \epsilon\nu(A), \qquad A \in \mathcal{F}.$$

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

Proof. We compute that

$$
\begin{aligned}
&k^2_{\mathrm{DUGS}}(x', y' \mid x, y)\\
&\quad= \int_{\mathcal{X}} \int_{\mathcal{Y}} k_{\mathrm{DUGS}}(x', y' \mid u, v) k_{\mathrm{DUGS}}(u, v \mid x, y) \mu_Y(\mathrm{d}v) \mu_X(\mathrm{d}u)\\
&\quad= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{Y \mid X}(y' \mid u) f_{X \mid Y}(x' \mid y') f_{Y \mid X}(v \mid x) f_{X \mid Y}(u \mid v) \mu_Y(\mathrm{d}v) \mu_X(\mathrm{d}u)\\
&\quad= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X \mid Y}(x' \mid y') f_{Y \mid X}(v \mid x) [f_{X \mid Y}(u \mid v) f_{Y \mid X}(y' \mid u)] \mu_Y(\mathrm{d}v) \mu_X(\mathrm{d}u)\\
&\quad= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X \mid Y}(x' \mid y') f_{Y \mid X}(v \mid x) k^*_{\mathrm{DUGS}}(u, y' \mid x, v) \mu_Y(\mathrm{d}v) \mu_X(\mathrm{d}u)\\
&\quad\geq \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X \mid Y}(x' \mid y') f_{Y \mid X}(v \mid x) g(u, y') \mu_Y(\mathrm{d}v) \mu_X(\mathrm{d}u)\\
&\quad= \int_{\mathcal{X}} f_{X \mid Y}(x' \mid y') g(u, y') \bigg[\int_{\mathcal{Y}} f_{Y \mid X}(v \mid x) \mu_Y(\mathrm{d}v) \bigg] \mu_X(\mathrm{d}u)\\
&\quad= \int_{\mathcal{X}} f_{X \mid Y}(x' \mid y') g(u, y') \mu_X(\mathrm{d}u)\\
&\quad=: h(x', y').
\end{aligned}
$$

Note that our assumption on g, and assumption (7), ensures that $\mu\{(x, y) \colon h(x, y) > 0\} > 0$. It follows that $\int h(x', y') \mu(\mathrm{d}(x', y')) > 0$. The result then follows by setting $\epsilon = \int h(x', y') \mu(\mathrm{d}(x', y'))$ and $\nu(A) = \epsilon^{-1} \int_A h(x', y') \mu(\mathrm{d}(x', y'))$.

We now proceed to our counterexample.

Proposition 3. *It is possible that P_{DUGS} is uniformly ergodic and, in fact, 2-minorisable, and furthermore w is X-bounded, but P_{CMH} fails to be even geometrically ergodic.*

Proof. Let π be the distribution on $(0, \infty)^2$ with density function $f(x, y) = \frac{1}{2} \mathrm{e}^{-y} \mathbf{1}_A(x, y)$, where A is the union of the squares $(m, m+1] \times (m-1, m]$ for $m = 1, 2, 3 \dots$ together with the infinite rectangle $(0, 1] \times (0, \infty)$ (see Figure 1).

We consider the CIS version of the CMH sampler. Let $q(x' \mid y')$ be the density of the Normal$(0,\ 1/y')$ distribution. Then, for $m - 1 < y \leq m$,

$$
w(x, y) := \frac{f_{X \mid Y}(x \mid y)}{q(x \mid y)} = \frac{\mathbf{1}_{[0,1] \cup (m, m+1]}(x)/2}{\sqrt{y/2\pi}\, \mathrm{e}^{-x^2 y/2}} = \frac{1}{2} \sqrt{\frac{2\pi}{y}} \mathrm{e}^{x^2 y/2} \mathbf{1}_{[0,1] \cup (m, m+1]}(x),
$$

so

$$
\sup_x w(x, y) = w(m+1, y) = \frac{1}{2} \sqrt{\frac{2\pi}{y}} \mathrm{e}^{(m+1)^2 y/2} < \infty,
$$

i.e. w is X-bounded.

Next, let P^*_{DUGS} be the Markov kernel corresponding to a Gibbs sampler in which we update first X and then Y. Then P^*_{DUGS} is 1-minorisable. This is easy to prove with an argument similar to that used in the proof of Proposition 1. Specifically, if the X-update is 1-minorisable then so is P^*_{DUGS}. Note that if $m - 1 < y \leq m$ then

$$
f_{X \mid Y}(x' \mid y) = \tfrac{1}{2} \mathbf{1}_{[0,1] \cup (m, m+1]}(x') \geq \tfrac{1}{2} \mathbf{1}_{[0,1]}(x').
$$

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

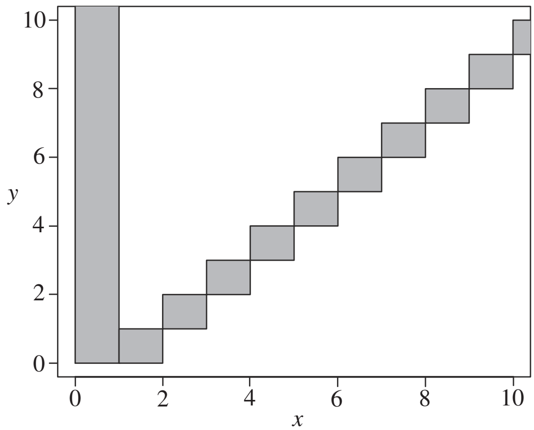

FIGURE 1: The region A used in the proof of Proposition 3.

Moreover, the right-hand side of the inequality holds for every value of $y > 0$ and, hence, we have, for all $y > 0$,

$$f_{X \mid Y}(x' \mid y) \geq \tfrac{1}{2}\mathbf{1}_{[0,1]}(x').$$

From this, it is easy to see that P^*_{DUGS} is minorised by the measure 2^{-1} Uniform$[0, 1] \times$ Exp(1). Hence, by Lemma 2, P_{DUGS} is 2-minorisable and, hence, is uniformly ergodic.

Finally, we use a capacitance argument (see, e.g. [15] and [39]) to show that this P_{CMH} is not uniformly ergodic (in fact, not even geometrically ergodic). However, since P^X_{CMH} is reversible with respect to π_X while P_{CMH} is not reversible with respect to π, we shall work with the former. (Recall that P^X_{CMH} and P_{CMH} have identical rates of convergence.) Before we give the capacitance argument we need a few preliminary observations.

Let $R_m = (m, m+1] \times (m-1, m]$ for some fixed $m \geq 3$, and suppose that $(x, y) \in R_m$. Then Y-moves will never leave R_m. Furthermore, X-moves will only leave R_m if a proposed value $x' \in [0, 1]$ is accepted; therefore,

$$\alpha(x', x, y) \leq \frac{w(x', y)}{w(x, y)} = \frac{\mathrm{e}^{(x')^2 y/2}}{\mathrm{e}^{x^2 y/2}} \leq \frac{\mathrm{e}^{(1)^2 m/2}}{\mathrm{e}^{m^2(m-1)/2}} = \mathrm{e}^{(-m^3+m^2+m)/2} \leq \mathrm{e}^{-m^3/4},$$

where the first inequality follows from the definition of α while the second follows since $m < x \leq m+1$, $m-1 < y \leq m$, and $0 \leq x' \leq 1$, and the third inequality follows since $m \geq 3$. Hence, for $x \in (m, m+1]$, $m \geq 3$,

$$P^X_{\mathrm{CMH}}(x, (m, m+1]^C) = P^X_{\mathrm{CMH}}(x, (0, 1]) \leq \mathrm{e}^{-m^3/4}.$$

Also, note that $\pi_X((m, m+1]) = 2^{-1}(\mathrm{e}^{-(m-1)} - \mathrm{e}^{-m})$.

Let κ be the capacitance of P^X_{CMH}. Then

$$\begin{aligned}
\kappa &:= \inf_{\{S \colon 0 < \pi_X(S) \leq 1/2\}} \frac{1}{\pi_X(S)} \int_S P^X_{\mathrm{CMH}}(x, S^C)\pi_X(\mathrm{d}x) \\
&\leq \inf_{m \geq 3} \frac{1}{\pi_X((m, m+1])} \int_{(m,m+1]} P^X_{\mathrm{CMH}}(x, ((m, m+1]^C))\pi_X(\mathrm{d}x) \\
&\leq \inf_{m \geq 3} \frac{2}{\mathrm{e}^{-(m-1)} - \mathrm{e}^{-m}} \int_{(m,m+1]} \mathrm{e}^{-m^3/4}\pi_X(\mathrm{d}x)
\end{aligned}$$

$$= \inf_{m\geq 3} \frac{2}{\mathrm{e}^{-(m-1)} - \mathrm{e}^{-m}} \mathrm{e}^{-m^3/4} \frac{1}{2} (\mathrm{e}^{-(m-1)} - \mathrm{e}^{-m})$$
$$= \inf_{m\geq 3} \mathrm{e}^{-m^3/4}$$
$$= 0.$$

Hence, P^X_{CMH} has capacitance 0, and hence has no spectral gap (see [15] and [39]), and hence fails to be geometrically ergodic [25]. Thus, P_{CMH} also fails to be geometrically ergodic.

3.3. Uniform return probabilities

To this point we have assumed that w is either bounded or X-bounded. It is natural to wonder if this is required for the uniform ergodicity of the CMH sampler. To examine this question further, we present two examples involving the CIS version of the CMH sampler. In the first example we show that in general P_{CIS} can fail to be even geometrically ergodic. In the second example we show that a slightly modified example is still uniformly ergodic even though w is neither bounded nor X-bounded.

Example 1. Let $\pi = \mathrm{Uniform}([0, 1]^2)$ so that $f_{X\mid Y}(x \mid y) = f_X(x) = \mathbf{1}(0 \leq x \leq 1)$ and $f_{Y\mid X}(y \mid x) = f_Y(y) = \mathbf{1}(0 \leq y \leq 1)$. Consider CIS with proposal density $q(x' \mid y') = 2x'$. Then the marginal chain P^X_{CIS} evolves independently of the Y values and corresponds to a usual independence sampler. This independence sampler has $f_X(x)/q(x) = (2x)^{-1}$, so $\sup_{x\in[0,1]} f_X(x)/q(x) = \infty$. It thus follows from standard independence sampler theory [16], [19], [33], [40] that P^X_{CIS} fails to be even geometrically ergodic. Hence, the joint chain P_{CIS} also fails to be geometrically ergodic.

Example 2. Again, let $\pi = \mathrm{Uniform}([0, 1]^2)$, but now let $q(x' \mid y') = 2\{y' - x'\}$, where $\{r\}$ is the fractional part of r (so $\{r\} = r$ if $0 \leq r < 1$, and $\{r\} = r + 1$ if $-1 \leq r < 0$). Then $w(x', y') = f_{X\mid Y}(x' \mid y')/q(x' \mid y') = 1/(2\{y'-x'\})$. Intuitively, the x' proposals will usually be accepted unless x is very close to y'. More precisely, let $S(x) = \{y \in [0, 1] \colon \{y - x\} \geq \frac{1}{2}\}$. If $x \in [0, 1]$ and $y' \in S(x)$, then

$$\frac{w(x', y')}{w(x, y')} = \frac{\{y' - x\}}{\{y' - x'\}} \geq \frac{1/2}{1} = \frac{1}{2}.$$

Hence, if we consider the marginal chain P^X_{CIS} then its subkernel $h_X(x' \mid x)$ satisfies

$$\begin{aligned} h_X(x' \mid x) &= \int_{y'\in\mathcal{Y}} q(x' \mid y')\alpha(x', x, y') f_{Y\mid X}(y' \mid x)\,\mathrm{d}y' \\ &\geq \int_{y'\in S(x)} q(x' \mid y') \min\biggl(1, \frac{w(x', y')}{w(x, y')}\biggr) f_{Y\mid X}(y' \mid x)\,\mathrm{d}y' \\ &\geq \int_{y'\in S(x)} (2\{y' - x'\})\biggl(\frac{1}{2}\biggr)(1)\,\mathrm{d}y' \\ &= \int_{y'\in S(x)} \{y' - x'\}\,\mathrm{d}y'. \end{aligned}$$

Now, $S(x)$ is the union of two disjoint intervals (or perhaps just one interval, if $x = 0$) within $[0, 1]$, of total length $\frac{1}{2}$. Also, the mapping $y' \mapsto \{y'-x'\}$ is some rearrangement of the identity mapping on $[0, 1]$. So, since $\int_{y'\in S(x)}\{y' - x'\}\,\mathrm{d}y'$ is an integral of some rearrangement of the

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

identity over some set of total length $\frac{1}{2}$, we must have $\int_{y'\in S(x)}\{y'-x'\}\,\mathrm{d}y' \geq \int_0^{1/2} r\,\mathrm{d}r = \frac{1}{8}$. Hence, $h_X(x' \mid x) \geq \frac{1}{8}$. Thus, for $A \in \mathcal{F}_X$,

$$P^X_{\mathrm{CIS}}(x, A) \geq \int_A h_X(x' \mid x)\mu_X(\mathrm{d}x') \geq \frac{1}{8}\mu_X(A).$$

So, P^X_{CIS} is 1-minorisable; hence, P^X_{CIS} is uniformly ergodic; therefore, P_{CIS} is also uniformly ergodic.

This last example suggests that even if w is not bounded or X-bounded, CIS will still be uniformly ergodic if the Y-move has a high probability of moving to a better subset. Generalising from the example, we have the following result.

Theorem 4. *Suppose that a CIS algorithm satisfies the following conditions:*

(i) *there is a subset $J \in \mathcal{F}_Y$ and a function $g\colon \mathcal{X} \to [0, \infty)$ with $\mu_X\{x\colon g(x) > 0\} > 0$ such that, for all $x \in \mathcal{X}$ and $y \in J$, we have $q(x \mid y) \geq g(x)$ and $f_{X\mid Y}(x \mid y) \geq g(x)$; and*

(ii) *the Y values have 'uniform return probabilities' in the sense that there exist* $0 < c < \infty$ *and $\delta > 0$ such that $\pi_{Y\mid X}(S(x) \mid x) \geq \delta$ for all $x \in \mathcal{X}$, where $S(x) = \{y' \in J\colon w(x, y') \leq c\}$.*

Then the CIS algorithm is uniformly ergodic and, furthermore, P^X_{CIS} is 1-minorisable.

Proof. We again consider the marginal chain P^X_{CIS}, whose subkernel $h_X(x' \mid x)$ now satisfies

$$\begin{aligned}
h_X(x' \mid x) &= \int_{y'\in\mathcal{Y}} q(x' \mid y')\alpha(x', x, y') f_{Y\mid X}(y' \mid x)\mu_Y(\mathrm{d}y') \\
&\geq \int_{y'\in S(x)} q(x' \mid y') \min\biggl(1, \frac{w(x', y')}{w(x, y')}\biggr) f_{Y\mid X}(y' \mid x)\mu_Y(\mathrm{d}y') \\
&\geq \int_{y'\in S(x)} q(x' \mid y') \min\biggl(1, \frac{f_{X\mid Y}(x' \mid y')}{q(x' \mid y')}\frac{1}{c}\biggr) f_{Y\mid X}(y' \mid x)\mu_Y(\mathrm{d}y') \\
&\geq \int_{y'\in S(x)} \min\biggl(q(x' \mid y'), f_{X\mid Y}(x' \mid y')\frac{1}{c}\biggr) f_{Y\mid X}(y' \mid x)\mu_Y(\mathrm{d}y') \\
&\geq \int_{y'\in S(x)} \min\biggl(1, \frac{1}{c}\biggr) g(x') f_{Y\mid X}(y' \mid x)\mu_Y(\mathrm{d}y') \\
&\geq \min\biggl(1, \frac{1}{c}\biggr) g(x')\delta.
\end{aligned}$$

Hence, for $A \in \mathcal{F}_X$,

$$P^X_{\mathrm{CIS}}(x, A) \geq \int_A h_X(x' \mid x)\mu_X(\mathrm{d}x') \geq \int_A \min\biggl(1, \frac{1}{c}\biggr) g(x')\delta\mu_X(\mathrm{d}x').$$

That is, P^X_{CIS} is 1-minorisable. Hence, P^X_{CIS} is uniformly ergodic. Therefore, P_{CIS} is also uniformly ergodic.

3.4. Geometric ergodicity of the CMH chain

Our goal in this section is to study conditions under which the geometric ergodicity of the DUGS chain implies the geometric ergodicity of the CMH chain. The key to our argument is Theorem 1(ii), which we will use to compare the convergence rates of the reversible Markov chains P^X_{CMH} and P^X_{DUGS}. The convergence rates of P^X_{CMH} and P^X_{DUGS} can then be connected to those of P_{CMH} and P_{DUGS} as described in Section 2.4. Our main result is the following.

Theorem 5. *If w is bounded and P_{DUGS} is geometrically ergodic, then P_{CMH} is geometrically ergodic.*

Proof. Let $C = \sup_{x',x,y'} w(x', x, y') < \infty$. Then

$$\begin{aligned}
h_X(x' \mid x) &= \int_{\mathcal{Y}} q(x' \mid x, y)\alpha(x', x, y) f_{Y\mid X}(y \mid x)\mu_Y(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} f_{Y\mid X}(y \mid x) f_{X\mid Y}(x' \mid y)\left[\frac{q(x' \mid x, y)}{f_{X\mid Y}(x' \mid y)} \wedge \frac{q(x \mid x', y)}{f_{X\mid Y}(x \mid y)}\right]\mu_Y(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} f_{Y\mid X}(y \mid x) f_{X\mid Y}(x' \mid y)\left[\frac{1}{w(x', x, y)} \wedge \frac{1}{w(x, x', y)}\right]\mu_Y(\mathrm{d}y) \\
&\ge \frac{1}{C}\int_{\mathcal{Y}} f_{Y\mid X}(y \mid x) f_{X\mid Y}(x' \mid y)\mu_Y(\mathrm{d}y) \\
&= \frac{1}{C} k_X(x' \mid x).
\end{aligned}$$

It follows that if $\delta = 1/C$ then

$$P^X_{\mathrm{CMH}}(x, A) \ge \delta P^X_{\mathrm{DUGS}}(x, A), \qquad x \in \mathcal{X},\ A \in \mathcal{F}_X.$$

Hence, by Theorem 1, if P^X_{DUGS} is geometrically ergodic then so is P^X_{CMH}. The result then follows by recalling that P^X_{DUGS} is geometrically ergodic if and only if P_{DUGS} is geometrically ergodic, and P^X_{CMH} is geometrically ergodic if and only if P_{CMH} is geometrically ergodic.

Example 3. Suppose that X and Y are bivariate normal with common mean 0, variances 2 and 1, respectively, and covariance 1. Then the two conditional distributions are $X \mid Y = y \sim \mathrm{N}(y, 1)$ and $Y \mid X = x \sim \mathrm{N}(\frac{1}{2}x, \frac{1}{2})$. This Gibbs sampler is known [35], [38] to be geometrically ergodic. Now consider a conditional independence sampler where we replace the Gibbs update for $X \mid Y = y$ with an independence sampler having proposal density

$$q(x \mid y) = \tfrac{1}{2}\mathrm{e}^{-|x-y|}.$$

Then it is easily seen that there exists a constant $c > 0$ such that $q(x \mid y) \ge c f_{X\mid Y}(x \mid y)$. Hence, Theorem 5 shows that the conditional independence sampler is geometrically ergodic.

Finally, we connect the geometric ergodicity of the RSGS with that of the random-scan CMH sampler.

Theorem 6. *If w is bounded and P_{RSGS} is geometrically ergodic for some selection probability, then P_{RCMH} is geometrically ergodic for any selection probability.*

Proof. Let $C = \sup_{x',x,y'} w(x', x, y') < \infty$. Then, similarly to Lemma 1,

$$\begin{aligned} P_{\mathrm{MH}:X}((x, y'), A) &\geq \int_{\{x' : (x', y') \in A\}} q(x' \mid x, y') \alpha(x', x, y') \mu_X(\mathrm{d}x') \\ &= \int_{\{x' : (x', y') \in A\}} q(x' \mid x, y') \left[1 \wedge \frac{f_{X \mid Y}(x' \mid y') q(x \mid x', y')}{f_{X \mid Y}(x \mid y') q(x' \mid x, y')} \right] \mu_X(\mathrm{d}x') \\ &= \int_{\{x' : (x', y') \in A\}} f_{X \mid Y}(x' \mid y') \left[\frac{1}{w(x', x, y')} \wedge \frac{1}{w(x, x', y')} \right] \mu_X(\mathrm{d}x') \\ &\geq \frac{1}{C} \int_{\{x' : (x', y') \in A\}} f_{X \mid Y}(x' \mid y') \mu_X(\mathrm{d}x') \\ &= \frac{1}{C} P_{\mathrm{GS}:X}((x, y'), A). \end{aligned}$$

Hence,

$$P_{\mathrm{RCMH}} = p P_{\mathrm{GS}:Y} + (1 - p) P_{\mathrm{MH}:X} \geq \frac{1}{C} [p P_{\mathrm{GS}:Y} + (1 - p) P_{\mathrm{GS}:X}] = \frac{1}{C} P_{\mathrm{RSGS}}.$$

Since both P_{RSGS} and P_{RCMH} are reversible with respect to π, the first claim now follows from Theorem 1. That the result holds for any selection probability then follows from Proposition 2.

4. Extensions to additional variables

In this section we consider the extent to which our results extend beyond the two-variable setting. Some of the above theorems (e.g. Theorem 5) make heavy use of the embedded X-chain kernels P^X_{CMH}, and such analysis appears to be specific to the case of two variables, one of which is updated using a Gibbs update. However, many of our other results extend beyond the two-variable setting without much additional difficulty aside from more general notation. Indeed, these generalisations will allow as many coordinates as desired to be updated using Metropolis–Hastings updates, so even in the two-variable case they generalise our previous theorems by no longer requiring one of the variables to be updated using a Gibbs update. In this sense the context of the results below is somewhat similar to that considered in [26], except that the results below concern 'global' rather than local/random-walk-style conditional proposal distributions.

Let $(\mathcal{X}_i, \mathcal{F}_i, \mu_i)$ be a σ-finite measure space for $i = 1, 2, \dots, d$ ($d \geq 2$), and let $(\mathcal{X}, \mathcal{F}, \mu)$ be the corresponding product space. Let π be a target probability distribution on $(\mathcal{X}, \mathcal{F}, \mu)$, having density f with respect to μ. For $x \in \mathcal{X}$ and $1 \leq i \leq d$, set $x_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$, $x_{[i]} = (x_1, \dots, x_i)$, and $x^{[i]} = (x_i, \dots, x_d)$. Also, let $x_{[0]}$ and $x^{[d+1]}$ be null. As we did in the two-variable case (recall (7)), we assume that the marginal densities satisfy $f_{X_i}(x_i) > 0$ for all $x_i \in \mathcal{X}_i$. Let f_i denote the corresponding conditional density of $X_i \mid X_{(i)}$. Then the usual DUGS has kernel

$$P_{\mathrm{DUGS}}(x, A) = \int_A k_{\mathrm{DUGS}}(x' \mid x) \mu(\mathrm{d}x'), \qquad A \in \mathcal{F},$$

where

$$k_{\mathrm{DUGS}}(x' \mid x) = f_1(x'_1 \mid x^{[2]}) f_2(x'_2 \mid x'_{[1]}, x^{[3]}) \cdots f_d(x'_d \mid x'_{[d-1]}).$$

Now consider the situation where some coordinates i are updated from the full-conditional Gibbs update $f_i(x'_i \mid x'_{[i-1]}, x^{[i+1]})$ as above, while other coordinates i are updated from a

Metropolis–Hastings update with proposal density $q_i(x_i' \mid x_{[i-1]}', x_i, x^{[i+1]})$ and corresponding acceptance probability

$$\alpha_i(x_{[i-1]}', x_i, x^{[i+1]}, x_i') = 1 \wedge \frac{f_i(x_i' \mid x_{[i-1]}', x^{[i+1]}) q_i(x_i \mid x_{[i-1]}', x_i', x^{[i+1]})}{f_i(x_i \mid x_{[i-1]}', x^{[i+1]}) q_i(x_i' \mid x_{[i-1]}', x_i, x^{[i+1]})}.$$

In fact, if $q_i(x_i' \mid x_{[i-1]}', x_i, x^{[i+1]}) = f_i(x_i' \mid x_{[i-1]}', x^{[i+1]})$ then $\alpha_i(x_{[i-1]}', x_i, x^{[i+1]}, x_i') \equiv 1$, and this is equivalent to updating coordinate i using a full-conditional Gibbs update. So, without loss of generality, we can assume that each coordinate i is updated according to a Metropolis–Hastings update as above.

To continue, let $g_i(w_i \mid z) = q_i(w_i \mid z_{[i-1]}, z_i, z^{[i+1]}) \alpha_i(z_{[i-1]}, z_i, z^{[i+1]}, w_i)$. Thus, g_i represents the absolutely continuous subkernel corresponding to the Metropolis–Hastings update of coordinate i and, in particular, g_i is a lower bound on the full update kernel for coordinate i. Of course, for those coordinates i which use a Gibbs update we have $g_i(w_i \mid z) = f_i(w_i \mid z_{[i-1]}, z^{[i+1]})$, the full-conditional density of coordinate i. Thus, if we let

$$k_{\mathrm{CMH}}(x' \mid x) = g_1(x_1' \mid x) g_2(x_2' \mid x_1', x^{[2]}) \cdots g_d(x_d' \mid x_{[d-1]}', x_d)$$

then

$$P_{\mathrm{CMH}}(x, A) \geq \int_A k_{\mathrm{CMH}}(x' \mid x) \mu(\mathrm{d}x'), \qquad A \in \mathcal{F}.$$

Correspondingly, for selection probabilities $(p_1, \dots, p_d) \in \mathbb{R}^d$ with each $p_i > 0$ and $\sum_{i=1}^d p_i = 1$, the RSGS is the algorithm which chooses coordinate i with probability p_i, and then updates that coordinate from $f_i(x_i' \mid x_{[i-1]}', x^{[i+1]})$ while leaving the other coordinates unchanged. The random-scan version of the CMH sampler, P_{RCMH}, is defined analogously.

Note that if each g_i is a Gibbs update, i.e. $g_i(x_i' \mid x_{[i-1]}', x^{[i]}) = f_i(x_i' \mid x_{[i-1]}', x^{[i+1]})$, then P_{CMH} is just the DUGS. That is, P_{DUGS} is a special case of P_{CMH} [29], so that, as in the previous section, it is natural to seek to connect the convergence properties of the two Markov chains.

Define the (conditional) weight function by

$$w_i(x_{[i-1]}', x_i', x_i, x^{[i+1]}) = \frac{f_i(x_i' \mid x_{[i-1]}', x^{[i+1]})}{q_i(x_i' \mid x_{[i-1]}', x_i, x^{[i+1]})}.$$

Say that w_i is *bounded* if

$$\sup_{x_{[i]}', x^{[i]}} w_i(x_{[i-1]}', x_i', x_i, x^{[i+1]}) < \infty$$

and is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$*-bounded* if there exists $C\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \to (0, \infty)$ such that

$$\sup_{x_i', x^{[i]}} w_i(x_{[i-1]}', x_i', x_i, x^{[i+1]}) \leq C(x_{[i-1]}').$$

Of course, for those coordinates i which use a full-conditional Gibbs update, we have

$$w_i(x_{[i-1]}', x_i', x_i, x^{[i+1]}) \equiv 1.$$

We begin with a generalisation of Lemma 1.

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

Lemma 3. *It holds that*

$$k_{\mathrm{CMH}}(x' \mid x) = k_{\mathrm{DUGS}}(x' \mid x) \prod_{i=1}^{d} \left[\frac{1}{w_i(x'_{[i-1]}, x'_i, x_i, x^{[i+1]})} \wedge \frac{1}{w_i(x'_{[i-1]}, x_i, x'_i, x^{[i+1]})} \right].$$

Proof. Note that, for $i = 1, \dots, d$,

$$\begin{aligned} q_i(x'_i \mid x'_{[i-1]}, x^{[i]}) &\left[1 \wedge \frac{f_i(x'_i \mid x'_{[i-1]}, x^{[i+1]}) q_i(x_i \mid x'_{[i-1]}, x'_i, x^{[i+1]})}{f_i(x_i \mid x'_{[i-1]}, x^{[i+1]}) q_i(x'_i \mid x'_{[i-1]}, x^{[i]})} \right] \\ &= f_i(x'_i \mid x'_{[i-1]}, x^{[i+1]}) \left[\frac{1}{w_i(x'_{[i-1]}, x'_i, x_i, x^{[i+1]})} \wedge \frac{1}{w_i(x'_{[i-1]}, x_i, x'_i, x^{[i+1]})} \right]. \end{aligned}$$

In light of the above lemma, the proofs of the following two theorems are similar to the proofs of Theorems 2 and 3. The corollaries follow as before.

Theorem 7. *If each w_i is bounded and P_{DUGS} is uniformly ergodic, then P_{CMH} is uniformly ergodic.*

Proof. By Lemma 3 we have

$$k_{\mathrm{CMH}}(x' \mid x) = k_{\mathrm{DUGS}}(x' \mid x) \prod_{i=1}^{d} \left[\frac{1}{w_i(x'_{[i-1]}, x'_i, x_i, x^{[i+1]})} \wedge \frac{1}{w_i(x'_{[i-1]}, x_i, x'_i, x^{[i+1]})} \right].$$

Since each w_i is bounded, there exist constants C_i, $i = 1, \dots, d$, such that

$$k_{\mathrm{CMH}}(x' \mid x) \geq k_{\mathrm{DUGS}}(x' \mid x) \prod_{i=1}^{d} \frac{1}{C_i},$$

and, hence,

$$P_{\mathrm{CMH}}(x, A) \geq \left[\prod_{i=1}^{d} \frac{1}{C_i} \right] P_{\mathrm{DUGS}}(x, A), \qquad A \in \mathcal{F}.$$

The result now follows from Theorem 1.

Corollary 3. *If each w_i is bounded and P_{DUGS} is uniformly ergodic, then P_{RCMH} is uniformly ergodic for any selection probabilities.*

Theorem 8. *If each w_i is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$-bounded, and there exists a nonnegative function g on $\mathcal{X}$, with $\mu\{x \in \mathcal{X} \colon g(x) > 0\} > 0$, such that*

$$k_{\mathrm{DUGS}}(x' \mid x) \geq g(x'), \qquad x \in \mathcal{X}, \tag{9}$$

then P_{CMH} is uniformly ergodic.

Proof. By Lemma 3 we have

$$k_{\mathrm{CMH}}(x' \mid x) = k_{\mathrm{DUGS}}(x' \mid x) \prod_{i=1}^{d} \left[\frac{1}{w_i(x'_{[i-1]}, x'_i, x_i, x^{[i+1]})} \wedge \frac{1}{w_i(x'_{[i-1]}, x_i, x'_i, x^{[i+1]})} \right].$$

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

Since each w_i is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$-bounded, there exist C_i such that

$$k_{\mathrm{CMH}}(x' \mid x) \geq k_{\mathrm{DUGS}}(x' \mid x) \prod_{i=1}^{d} \frac{1}{C_i(x'_{[i-1]})}.$$

Then, using (9), we have

$$k_{\mathrm{CMH}}(x' \mid x) \geq g(x') \prod_{i=1}^{d} \frac{1}{C_i(x'_{[i-1]})}.$$

Let

$$\epsilon = \int_{\mathcal{X}} g(x) \prod_{i=1}^{d} \frac{1}{C_i(x_{[i-1]})} \mu(\mathrm{d}x) \quad \text{and} \quad h(x') = \epsilon^{-1} g(x') \prod_{i=1}^{d} \frac{1}{C_i(x'_{[i-1]})}.$$

Then, if $A \in \mathcal{F}$,

$$P_{\mathrm{CMH}}(x, A) \geq \epsilon \int_A h(x')\mu(\mathrm{d}x').$$

That is, P_{CMH} is 1-minorisable and, hence, is uniformly ergodic.

Corollary 4. *If each w_i is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$-bounded, and condition (9) holds, then P_{RCMH} is uniformly ergodic for any selection probabilities.*

Furthermore, Proposition 2 extends easily to the general case.

Proposition 4. *If P_{RSGS} is geometrically ergodic for some selection probability then it is geometrically ergodic for all selection probabilities.*

Just as with Theorem 6, we can also give sufficient conditions for geometric ergodicity of P_{RCMH} in terms of the geometric ergodicity of P_{RSGS}.

Theorem 9. *If each w_i is bounded and P_{RSGS} is geometrically ergodic, then P_{RCMH} is geometrically ergodic for any selection probabilities.*

5. Application to Bayesian inference for diffusions

An important problem, with applications to financial analysis and many other areas, involves drawing inferences about the entire path of a diffusion process based only upon discrete observations of that diffusion (see, e.g. [4] and [32]).

To fix ideas, consider a one-dimensional diffusion satisfying $\mathrm{d}X_t = \mathrm{d}B_t + \alpha(X_t)\,\mathrm{d}t$ for $0 \leq t \leq 1$, where $\alpha\colon \mathbb{R} \to \mathbb{R}$ is a C^1 function. Suppose that we observe the values X_0 and X_1, and wish to infer the entire remaining sample path $\{X_t\}_{0<t<1}$.

To proceed, let $\mathbb{P}_\theta$ be the law of the diffusion starting at X_0, conditional on θ, and let $\mathbb{W}$ be the law of Brownian motion starting at X_0. Then, by Girsanov's formula (see, e.g. [34]), the density of $\mathbb{P}_\theta$ with respect to $\mathbb{W}$ satisfies (writing $X_{[0,1]}$ for $\{X_t\}_{0\leq t\leq 1}$)

$$G_\theta(X_{[0,1]}) := \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{W}}(X_{[0,1]}) = \exp\Big[A(X_1) - A(X_0) - \int_0^1 \phi_\theta(X_s)\,\mathrm{d}s\Big], \tag{10}$$

where $A(x) = \int_0^x \alpha(u)\,\mathrm{d}u$ and $\phi_\theta(x) = [\alpha^2(x) + \alpha'(x)]/2$.

Furthermore, if $\tilde{\mathbb{P}}$ is the law of the diffusion conditional on the observed values of X_0 and X_1, and $\tilde{\mathbb{W}}$ is the law of Brownian motion conditional on the same observed values of X_0 and

https://doi.org/10.1239/aap/1401369701 Published online by Cambridge University Press

X_1 (i.e. of the corresponding Brownian bridge process), then $\mathrm{d}\tilde{\mathbb{P}}/\mathrm{d}\tilde{\mathbb{W}}$ is still proportional to the same density G from (10).

Assume now that $\alpha(x) = \sum_{i=1}^m p_i(x)\theta_i = p^\top\theta$, where $p_1, p_2, \ldots, p_m : \mathbb{R} \to \mathbb{R}$ are known C^1 functions, and $\theta_1, \theta_2, \ldots, \theta_m$ are unknown real-valued parameters to be estimated.

We consider a Bayesian analysis obtained by putting a prior $\theta \sim \mathrm{MVN}(0, \Sigma_0)$ on the vector θ for some strictly positive-definite symmetric $m \times m$ covariance matrix Σ_0. Then, conditional on X_0 and X_1, and letting $X_{\mathrm{miss}} = \{X_s : 0 < s < 1\}$ be the missing (unobserved) part of the diffusion's sample path, the joint posterior density of the pair $(\theta, X_{\mathrm{miss}})$ is proportional to

$$\mathrm{e}^{-\theta^\top\Sigma_0^{-1}\theta/2} G_\theta(X_{[0,1]}) = \exp\bigg[-\frac{1}{2}\bigg(\theta^\top\Sigma_0^{-1}\theta + \int_0^1 \sum_{i=1}^m \sum_{j=1}^m p_i(X_s)p_j(X_s)\theta_i\theta_j \,\mathrm{d}s + \int_0^1 \sum_{i=1}^m p_i'(X_s)\theta_i \,\mathrm{d}s\bigg)\bigg].$$

We can write this joint posterior density as being proportional to

$$\exp\big[-\tfrac{1}{2}\theta^\top V^{-1}\theta - r^\top\theta\big], \tag{11}$$

in terms of the column vector $r = \frac{1}{2}\int_0^1 p'(X_s)\,\mathrm{d}s$, and the positive-definite symmetric matrix

$$V^{-1} = \Sigma_0^{-1} + \int_0^1 p(X_s)(p(X_s))^\top \,\mathrm{d}s.$$

Then, since

$$-\tfrac{1}{2}(\theta + Vr)^\top V^{-1}(\theta + Vr) = -\tfrac{1}{2}\theta^\top V^{-1}\theta - r^\top\theta - \tfrac{1}{2}r^\top Vr$$

(using the facts that $V^\top = V$, and $r^\top\theta = \theta^\top r$ is a scalar), (11) in turn implies that the conditional distribution $\theta \mid X_{\mathrm{miss}}$ is given by

$$\theta \mid X_{\mathrm{miss}} \sim \mathrm{MVN}(-Vr, V). \tag{12}$$

Now, suppose that we wish to sample the pair $(\theta, X_{\mathrm{miss}})$ from its posterior density (11). We first consider using a DUGS, in which we alternately sample $\theta \mid X_{\mathrm{miss}}$ and then $X_{\mathrm{miss}} \mid \theta$.

Lemma 4. *Assume that the p_i and p_i' functions are all bounded, i.e.*

$$\max_{1\le i\le m} \sup_{x\in\mathbb{R}} \max(|p_i(x)|, |p_i'(x)|) < \infty. \tag{13}$$

Then DUGS for the pair $(\theta, X_{\mathrm{miss}})$ is 1-minorisable.

Proof. In light of Proposition 1, it suffices to show that the θ updates, as carried out through (12), are 1-minorisable.

Denote the density of $\mathrm{MVN}(\mu, \Sigma)$ by $f(\theta; \mu, \Sigma)$. We remark that this function is positive and continuous on $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{M}$ (where $\mathbb{M}$ denotes the space of positive definite $m \times m$ matrices). Therefore, by the standard compactness argument, if A is any compact set in $\mathbb{R}^m \times \mathbb{M}$ then, for all $\theta \in \mathbb{R}^m$,

$$\inf_{(\mu,\Sigma)\in A} f(\theta; \mu, \Sigma) > 0,$$

thus providing a minorisation measure. It remains therefore to show that, given all possible diffusion trajectories, the mean $(-Vr)$ and variance (V) in (12) are uniformly contained in bounded regions, with the determinant of the variance bounded away from 0. Note that (13) and the definition of V imply immediately that V is uniformly bounded, proving the first part. Moreover, showing that $\det(V)$ is uniformly bounded away from 0 is equivalent to a uniform upper bound on $\det(V^{-1})$. However, this also follows trivially from (12). Thus, it follows that the θ update is 1-minorisable.

The above lemma shows that DUGS for the pair $(\theta, X_{\text{miss}})$ is uniformly ergodic. However, in practice, it is entirely infeasible to sample the entire path X_{miss} from its correct conditional distribution given θ. Thus, to sample the pair $(\theta, X_{\text{miss}})$ from the posterior density (11), we instead consider using a CIS. Here θ plays the role of Y and X_{miss} plays the role of X. We shall alternately update θ from its full-conditional distribution conditional on the current value of X_{miss} (which is easy to implement in practice, since $\theta \mid X_{\text{miss}}$ follows a Gaussian distribution), and then update X_{miss} using a conditional Metropolis–Hastings update step with proposal distribution $q(X_{\text{miss}} \mid \theta)$ given by the corresponding Brownian bridge, i.e. with $q(X_{\text{miss}} \mid \theta) = \tilde{\mathbb{W}}$ (which can be implemented in practice by, e.g. discretising the time interval $[0, 1]$ and then using Gaussian conditional distributions of the Brownian bridge). This algorithm is thus feasible to implement in practice, thus raising the question of its ergodicity properties, which we now consider.

This CIS algorithm has conditional weight functions given by

$$w(x_{\text{miss}}, \theta) = \frac{f_{X_{\text{miss}} \mid \theta}(x_{\text{miss}} \mid \theta)}{q(x_{\text{miss}} \mid \theta)} = \frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\tilde{\mathbb{W}}}(X_{[0,1]}) = h(\theta) G_\theta(X_{[0,1]}),$$

where we explicitly include the normalisation constant $h(\theta)$ which is everywhere positive and finite. The key computation in our analysis is the following.

Lemma 5. *For the above CIS algorithm, assuming that (13) holds, the weights are X-bounded, i.e.* $\sup_x w(x, \theta) < \infty$ *for each fixed θ.*

Proof. From (10), we can write

$$\begin{aligned} w(x_{\text{miss}}, \theta) &= h(\theta) G_\theta(X_{[0,1]}) \\ &= h(\theta) \exp\biggl[A(X_1) - A(X_0) - \int_0^1 \phi_\theta(X_s)\,\mathrm{d}s\biggr] \\ &\le h(\theta) \exp[A(X_1) - A(X_0)] \exp\Bigl[-\inf_x \phi_\theta\Bigr], \end{aligned}$$

which shows that it suffices to argue that $\phi_\theta(x)$ is bounded below as a function of θ. But

$$\phi_\theta = \frac{1}{2}\biggl[\theta^\top\biggl(\int p(X_s)(p(X_s))^\top\,\mathrm{d}s\biggr)\theta + \biggl(\int (p'(X_s))^\top\,\mathrm{d}s\biggr)\theta\biggr].$$

Hence, by the boundedness of p_i and p_i' from (13), it follows that $\phi_\theta(x)$ is bounded below. This gives the result.

We can now easily prove our main result of this section.

Theorem 10. *Assuming that (13) holds, the above CIS algorithm on $(X_{\text{miss}}, \theta)$, conditional on the observed values X_0 and X_1, is uniformly ergodic.*

Proof. This follows immediately from Theorem 3, in light of Lemmas 4 and 5.

5.1. Generalisation to more data

In practice, fitting a diffusion model, we would almost certainly possess multiple data, $X_{\mathrm{obs}} = (X_{t_0}, X_{t_1}, X_{t_2}, \dots, X_{t_N})$, observed at times $t_0, t_1, t_2, \dots, t_N$, leading in turn to missing diffusion segments $X_{\mathrm{miss},i} = \{X_t : t_{i-1} < t < t_i\}$ for $1 \le i \le N$. For ease of notation, we have avoided this more general setting in this section so far. However, we now give some brief remarks to show that Theorem 10 easily generalises.

In this more general case (often called *discretely observed data*), the following algorithm was implemented in, e.g. [32] to infer the $X_{\mathrm{miss},i}$ segments and θ. To fit with earlier notation, we fix $t_0 = 0$ and $t_N = 1$.

1. Given X_{obs} and $\{X_{\mathrm{miss},i}\}_{1\le i\le N}$, simulate θ from its full conditional as given in (12).
2. Sequentially for $i = 1, 2, \dots, N$, propose an update of $X_{\mathrm{miss},i}$ conditional on X_{obs} and θ from the Brownian bridge measure between $X_{t_{i-1}}$ and time t_{i-1}, and X_{t_i} and time t_i, and accept according to the usual Metropolis–Hastings accept/reject ratio.

The key here is that, conditional on θ, the $\{X_{\mathrm{miss},i}\}_{1\le i\le N}$ segments are all conditionally independent. As a result of this, using our multidimensional theorem extensions of Section 4, we immediately obtain the following generalisation of Theorem 10.

Theorem 11. *Assuming that (13) holds, the above CIS algorithm on* $(X_{\mathrm{miss}}, \theta)$*, conditional on the observed values* $X_{t_1}, X_{t_2}, \dots, X_{t_N}$*, is uniformly ergodic.*

Acknowledgement

We thank the anonymous referee for very thorough and helpful reports which greatly improved the manuscript.

References

[1] Brooks, S., Gelman, A., Jones, G. L. and Meng, X.-L. (eds) (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL.
[2] Chan, K. S. and Geyer, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22,** 1747–1758.
[3] Diaconis, P. and Saloff-Coste, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Prob.* **3,** 696–730.
[4] Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69,** 959–993.
[5] 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.
[6] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85,** 398–409.
[7] 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.
[8] Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Process. Appl.* **85,** 341–361.
[9] Johnson, A. A. and Jones, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Statist.* **4,** 313–333.
[10] Johnson, A. A., Jones, G. L. and Neath, R. C. (2013). Component-wise Markov chain Monte Carlo: uniform and geometric ergodicity under mixing and composition. *Statist. Sci.* **28,** 360–375.
[11] Jones, G. L. (2004). On the Markov chain central limit theorem. *Prob. Surveys* **1,** 299–320.
[12] Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16,** 312–334.
[13] 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.

[14] 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.

[15] Lawler, G. F. and Sokal, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309,** 557–580.

[16] Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statist. Comput.* **6,** 113–119.

[17] Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81,** 27–40.

[18] 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.

[19] Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24,** 101–121.

[20] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

[21] Papaspiliopoulos, O. and Roberts, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Ann. Statist.* **36,** 95–117.

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

[23] Robert, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statist. Sci.* **10,** 231–253.

[24] Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *J. R. Statist. Soc. B* **56,** 377–384.

[25] Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Prob.* **2,** 13–25.

[26] Roberts, G. O. and Rosenthal, J. S. (1998). Two convergence properties of hybrid samplers. *Ann. Appl. Prob.* **8,** 397–407.

[27] Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *J. R. Statist. Soc. B* **61,** 643–660.

[28] Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Statist.* **28,** 489–504.

[29] Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Prob. Surveys* **1,** 20–71.

[30] Roberts, G. O. and Rosenthal, J. S. (2011). Quantitative non-geometric convergence bounds for independence samplers. *Methodol. Comput. Appl. Prob.* **13,** 391–403.

[31] Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parametrization for the Gibbs sampler. *J. R. Statist. Soc. B* **59,** 291–317.

[32] Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88,** 603–621.

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

[34] Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes, and Martingales,* Vol. 1, *Foundations*, 2nd edn. John Wiley, Chichester.

[35] Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90,** 558–566. (Correction: **90** (1995), 1136.)

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

[37] 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. Statist. Soc. B* **69,** 607–623.

[38] Schervish, M. J. and Carlin, B. P. (1992). On the convergence of successive substitution sampling. *J. Comput. Graph. Statist.* **1,** 111–127.

[39] Sinclair, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Prob. Comput.* **1,** 351–370.

[40] Smith, R. L. and Tierney, L. (1996). Exact transition probabilities for the independence Metropolis sampler. *Tech. Rep.*, University of North Carolina.

[41] 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.

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

[43] Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8,** 1–9.