# Convergence of Conditional Metropolis-Hastings Samplers

Galin L. Jones[*] Gareth O. Roberts[†] Jeffrey S. Rosenthal[‡]

(June, 2012; revised March 2013 and June 2013)

**Abstract**

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

## 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, 30, 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 *Metropolis-Hastings-within-Gibbs*). We consider the mixed case in which some components are updated as in the Gibbs sampler, while

[*] School of Statistics, University of Minnesota, Minneapolis, MN, USA 55455 Email: galin@umn.edu Partially supported by the National Institutes for Health.

[†] Department of Statistics, University of Warwick, CV4 7AL, Coventry, UK Email: Gareth.O.Roberts@warwick.ac.uk

[‡] Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3. Email: jeff@math.toronto.edu. Web: http://probability.ca/jeff/ Partially supported by NSERC of Canada.

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 herein.

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, 26] and to calculate asymptotically valid Monte Carlo standard errors [5, 12]. While there has been much progress in proving geometric ergodicity for many MCMC samplers [see e.g. 7, 8, 9, 14, 17, 18, 24, 25, 28, 33, 36, 37, 41], doing so typically requires difficult theoretical analysis.

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 $\pi$ be a probability distribution having support $\mathcal{X} \times \mathcal{Y}$, and $\pi_{X|Y}$ and $\pi_{Y|X}$ denote the associated conditional distributions. Suppose $\pi_{Y|X}$ has a density $f_{Y|X}$, and $\pi_{X|Y}$ has density $f_{X|Y}$. There are several potential componentwise MCMC algorithms, each having $\pi$ as its invariant distribution. If it is possible to simulate from $\pi_{X|Y}$ and $\pi_{Y|X}$, then one can implement a deterministic-scan Gibbs sampler (DUGS), which is now described. Suppose 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.

---

Iteration $n+1$ of the deterministic-scan Gibbs sampler (DUGS):

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

---

However, sometimes one or both of these steps will be computationally infeasible, necessitating the use of alternative algorithms. In particular, suppose we continue to simulate directly from $\pi_{Y|X}$, but use a Metropolis-Hastings algorithm for $\pi_{X|Y}$ with proposal density $q(x'|x, y')$. This results in a *conditional Metropolis-Hastings sampler* (CMH), which is now described. 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.

---

Iteration $n+1$ of the conditional Metropolis-Hastings sampler (CMH):

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

---

As is well-known DUGS is a special case of the CMH where the proposal is taken to be the conditional, that is, $q(x'|x,y') = f_{X|Y}(x'|y')$ [30]. Thus, it is natural to suspect that the convergence properties of DUGS and CMH may be related. On the other hand, while geometric ergodicity of the Gibbs sampler has been extensively studied [17, 21, 24, 25], the CMH has received comparatively little attention [10].

If the proposal distribution for $x'$ does not depend on the previous value of $x$, i.e. if $q(x'|x,y') = q(x'|y')$, then in CMH the $X$ values are updated as in an independence sampler [see e.g. 31, 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 organized 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 CMH, including uniform ergodicity in terms of the conditional weight function (Theorems 5 and 7) and uniform return probabilities (Theorem 11), and geometric ergodicity via a comparison to DUGS (Theorem 12). 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, $P : \mathcal{Z} \times \mathcal{F} \to [0, 1]$, 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 $\Phi = \{Z_0, Z_1, \ldots\}$ is the Markov chain with transitions governed by $P$, then for any positive integer $n$, the $n$-step Markov transition kernel is given by $P^n(z, A) = \Pr(Z_{n+j} \in A | Z_j = z)$, which is assumed to be the same for all times $j$.

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

$$\nu P(A) = \int \nu(dz) P(z, A)$$

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

$$Pf(z) = \int f(y) P(z, dy)$$

and

$$\pi(f) = \int f(z) \pi(dz) \ .$$

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

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

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

Uniform ergodicity is equivalent to a so-called minorization condition [see e.g. 20, 30]. That is, $P$ is uniformly ergodic if and only if there exists a positive integer $m \geq 1$, a constant $\varepsilon > 0$ and a probability measure $Q$ on $\mathcal{Z}$ such that for all $z \in \mathcal{Z}$,

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

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

Establishing geometric ergodicity is most commonly done by establishing various Foster-Lyapounov criteria [13, 20, 30], but these will play no role here. Instead we will focus on another characterization of geometric ergodicity which 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

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

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

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

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 $\pi$, that is, if

$$P(z, dz') \, \pi(dz) \; = \; P(z', dz) \, \pi(dz'), \tag{3}$$

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

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

while the *spectral gap* of $P$ is $\text{gap}(P) = 1 - r(P_0)$. If $P$ is reversible with respect to $\pi$ 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 $\pi$, then $P$ is geometrically ergodic if and only if $\text{gap}(P) > 0$, or equivalently $\|P\| < 1$ [26].

## 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 CMH.

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

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

1. *If $P$ and $Q$ have invariant distribution $\pi$ and $Q$ is uniformly ergodic, then so is $P$.*

2. *If $P$ and $Q$ are reversible with respect to $\pi$ and $Q$ is geometrically ergodic, then so is $P$.*

*Proof.* *1.* Note that (5) implies that for all $n$,

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

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

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

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

*2.* 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 restriction of $P$, $Q$ and $R$, respectively, to $L^2_{0,1}(\pi)$. Since $P$ is reversible with respect to $\pi$, and $\|R\| \leq 1$ so $r(R_0) \leq 1$, we have by (4) that

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

Hence,

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

Since $Q$ is geometrically ergodic, $\text{gap}(Q) > 0$, and hence $\text{gap}(P) > 0$. 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 $\sigma$-finite measure spaces, and let $(\mathcal{Z}, \mathcal{F}, \mu)$ be their product space. Let $\pi$ be a probability distribution on $(\mathcal{Z}, \mathcal{F}, \mu)$ which has a density $f(x, y)$ with respect to $\mu$. Then the marginal distributions $\pi_X$ and $\pi_Y$ of $\pi$ have densities given by

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

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 $f_{X|Y}(x|y) = f(x, y)/f_Y(y)$ and $f_{Y|X}(y|x) = f(x, y)/f_X(x)$.

Define a Markov kernel for a $Y$ update by

$$P_{GS:Y}(x, A) \;=\; \int_{\{y:(x,y)\in A\}} f_{Y|X}(y|x)\, \mu_Y(dy)\,,$$

and similarly an $X$ update is described by the Markov kernel

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

We can define the Markov kernel for the deterministic-scan Gibbs sampler (DUGS) by the composition of $X$ and $Y$ updates, i.e. $P_{DUGS} \;=\; P_{GS:Y}\, P_{GS:X}$ corresponding 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 $k_{DUGS}(x', y'|x, y) = f_{Y|X}(y'|x)\, f_{X|Y}(x'|y')$, then we can also write this as

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

Note that $\pi P_{DUGS} = \pi$, i.e. $\pi$ is a stationary distribution for $P_{DUGS}$, although $P_{DUGS}$ is not reversible with respect to $\pi$. 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 below).

**Proposition 2.** *If the $Y$-update of $P_{DUGS}$ is 1-minorisable, in the sense that there is $\epsilon > 0$ and a probability measure $\nu$ such that $P_{GS:Y}(x, A) \geq \epsilon\nu(A)$ for all $x$ and $A$, then $P_{DUGS}$ is 1-minorisable.*

*Proof.* The result follows from noting that

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

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 a Markov kernel $P^*_{DUGS} \;=\; P_{GS:X}\, P_{GS:Y}$, which will play a role in Section 3.2. Notice that with essentially the same argument as in Proposition 2 we have that 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 $Y \sim P_{GS:Y}$ with probability $p$, or updating $X \sim P_{GS:X}$ with probability $1-p$. The RSGS has kernel

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

i.e.

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

It follows that $P_{RSGS}$ is reversible with respect to $\pi$. Furthermore, it is well known [e.g. 10, 26] that if $P_{DUGS}$ is uniformly ergodic, then so is $P_{RSGS}$ (as follows immediately from (2), since we always have $P^{2n}_{RSGS}(z,A) \geq \big(p(1-p)\big)^n P^n_{DUGS}(z,A)$). We also have the following.

**Proposition 3.** *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 $P_{RSGS,p}$ be the RSGS kernel using selection probability $p$, so that if $A \in \mathcal{F}$, then

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

It follows immediately that

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

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

Next, consider the deterministically updated conditional Metropolis-Hastings sampler (CMH) 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 $P_{GS:Y}$ as above. To define the $X$ update, let $q(x'|x,y')$ be a proposal density and set

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

and

$$r(x, y') \;=\; 1 - \int q(x'|x, y')\alpha(x', x, y')\,\mu_X(dx')\,.$$

Then the $X$ update follows the Markov kernel defined by

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

By construction $P_{MH:X}$ is reversible with respect to $\pi$ (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 conditional Metropolis-Hastings sampler by their composition, corresponding to doing first a Gibbs sampler $Y$-move and then a Metropolis-Hastings $X$-move:

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

It then follows that $\pi P_{CMH} = \pi$, but $P_{CMH}$ is not reversible with respect to $\pi$. 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

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

then by construction we have that

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

We will also consider the random scan CMH (RCMH) sampler. For any fixed selection probability $p \in (0, 1)$, RCMH 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 conditional Metropolis-Hastings step for $X$), while leaving the other coordinate unchanged. Hence, its kernel is given by

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

Then $P_{RCMH}$ is reversible with respect to $\pi$. A similar argument to the one 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 $q(x'|x, y') = q(x'|y')$, then the CMH algorithm becomes the *conditional independence sampler* (CIS). In this case, we will continue to use all of the same notation as for CMH above, except omitting the unnecessary $x$ arguments.

### 2.4 Embedded $X$-Chains

When studying geometric ergodicity, Theorem 1 (part 2) does not apply directly to $P_{DUGS}$ and $P_{CMH}$ since they are not reversible with respect to $\pi$. However, each of these samplers do produce marginal $X$-sequences which are reversible with respect to the marginal distribution $\pi_X$ (with density as in (6)). Moreover, as we discuss below, if either of these $X$-sequences are 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

$$k_X(x'|x) = \int_{\mathcal{Y}} f_{X|Y}(x'|y) f_{Y|X}(y|x) \mu_Y(dy)$$

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

$$P_{DUGS}^X(x, A) = \int_A k_X(x'|x) \mu_X(dx'), \qquad A \in \mathcal{F}_X \ .$$

Now $P_{DUGS}$ has $\pi$ as its invariant distribution while $P_{DUGS}^X$ has the marginal distribution $\pi_X$ as its invariant distribution and, in fact, $P_{DUGS}^X$ is reversible with respect to $\pi_X$. Moreover, it is well known that $P_{DUGS}$ and $P_{DUGS}^X$ converge to their respective invariant distributions at the same rate [17, 23, 29]. This has been routinely exploited in the analysis of two-variable Gibbs samplers where $P_{DUGS}^X$ may be much easier to analyze than $P_{DUGS}$.

Now consider the CMH algorithm, and let its resulting values be $Y_0, X_0, Y_1, X_1, Y_2, X_2, \ldots$. This sequence in turn provides a marginal sequence, $X_0, X_1, \ldots$ which is itself a Markov chain on $\mathcal{X}$, since the $P_{GS:Y}$ update within CMH 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 CMH 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 $P^X_{CMH}(x, A)$, and if

$$h_X(x'|x) \;=\; \int_{\mathcal{Y}} f_{Y|X}(y'|x)\, q(x'|x, y')\, \alpha(x', x, y') \mu_Y(dy') \;,$$

it follows by construction that

$$P^X_{CMH}(x, A) \;\geq\; \int_A h_X(x'|x)\, \mu_X(dx'), \qquad A \in \mathcal{F}_X \;.$$

Note that $P_{CMH}$ and $P^X_{CMH}$ have invariant distributions $\pi$ and $\pi_X$, respectively. Now $P_{CMH}$ is not reversible with respect to $\pi$, but we shall show that $P^X_{CMH}$ is reversible with respect to $\pi_X$. Indeed, first note that by construction

$$P_{MH:X}((x, y), (dx', y))\pi_{X|Y}(dx|y) \;=\; P_{MH:X}((x', y), (dx, y))\pi_{X|Y}(dx'|y) \;.$$

Now we compute

$$\begin{aligned}
P^X_{CMH}(x, dx')\pi_X(dx) &= \pi_X(dx) \int_{\mathcal{Y}} P_{MH:X}((x, y), (dx', y))\, \pi_{Y|X}(dy|x) \\
&= \int_{\mathcal{Y}} P_{MH:X}((x, y), (dx', y))\, \pi(dx, dy) \\
&= \int_{\mathcal{Y}} P_{MH:X}((x, y), (dx', y))\, \pi_{X|Y}(dx|y)\pi_Y(dy) \\
&= \int_{\mathcal{Y}} P_{MH:X}((x', y), (dx, y))\, \pi_{X|Y}(dx'|y)\pi_Y(dy) \\
&= \int_{\mathcal{Y}} P_{MH:X}((x', y), (dx, y))\, \pi(dx', dy) \\
&= \pi_X(dx') \int_{\mathcal{Y}} P_{MH:X}((x', y), (dx, y))\, \pi_{Y|X}(dy|x') \\
&= P^X_{CMH}(x', dx)\pi_X(dx')
\end{aligned}$$

and conclude that $P^X_{CMH}$ is reversible with respect to $\pi_X$.

It is straightforward to see that, in the language of [29], the embedded chain $P^X_{CMH}$ is *de-initialising* for $P_{CMH}$. This implies that if $P^X_{CMH}$ is geometrically (or uniformly) ergodic, then $P_{CMH}$ is geometrically (or uniformly) ergodic [29, Theorem 1]. In fact, it is not too hard to show the converse [10] and conclude that $P^X_{CMH}$ is geometrically (or uniformly) ergodic if and only if $P_{CMH}$ is geometrically (or uniformly) ergodic.

## 3 Ergodicity Properties of CMH

Our goal in this section is to derive ergodicity properties of the conditional Metropolis-Hastings (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 CMH 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

$$w(x', x, y') := \frac{f_{X|Y}(x'|y')}{q(x'|x, y')} \qquad x', x \in \mathcal{X},\ y' \in \mathcal{Y}\,. \tag{8}$$

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

**Lemma 4.**

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

*Proof.* Notice that

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

□

Say that $w$ is *bounded* if

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

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

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

We then have the following.

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

*Proof.* By Lemma 4, we have

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

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

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

and hence

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

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:

**Corollary 6.** *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 5 can be weakened if we strengthen the assumption on the Gibbs sampler.

**Theorem 7.** *Suppose that $w$ is $X$-bounded, and that there exists a non-negative function $g$ on $\mathcal{Z}$, with $\mu\{(x, y) : g(x, y) > 0\} > 0$, such that for all $x$ and $y$,*

$$k_{DUGS}(x', y'|x, y) \geq g(x', y') \; . \tag{9}$$

*Then $P_{CMH}$ is uniformly ergodic.*

*Proof.* By Lemma 4 we have

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

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

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

and using (9) we obtain

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

Letting

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

we have that

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

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

*Remark* 3. Notice that condition (9) implies that $P_{DUGS}$ is 1-minorisable.

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

**Corollary 8.** *If $w$ is $X$-bounded, and condition* (9) *holds, then $P_{RCMH}$ is uniformly ergodic for any selection probability $p \in (0, 1)$.*

### 3.2 A Counter-Example

In this section, we show that Theorem 7 might not hold if $P_{DUGS}$ is just 2-minorisable (as opposed to 1-minorisable). We begin with a lemma about interchanging 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

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

The following lemma shows that we can convert a 1-minorisation for $P^*_{DUGS}$ into a 2-minorisation for $P_{DUGS}$.

**Lemma 9.** *Suppose there exists a non-negative function $g$ on $\mathcal{Z}$, with $\mu\{(x,y) : g(x,y) > 0\} > 0$, such that for all $x$ and $y$,*

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

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

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

*Proof.* We compute that

$$\begin{aligned}
k^2_{DUGS}(x', y'|x, y) &= \int_{\mathcal{X}} \int_{\mathcal{Y}} k_{DUGS}(x', y'|u, v) k_{DUGS}(u, v|x, y) \mu_Y(dv)\mu_X(du) \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{Y|X}(y'|u) f_{X|Y}(x'|y') f_{Y|X}(v|x) f_{X|Y}(u|v) \mu_Y(dv)\mu_X(du) \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X|Y}(x'|y') f_{Y|X}(v|x) \left[ f_{X|Y}(u|v) f_{Y|X}(y'|u) \right] \mu_Y(dv)\mu_X(du) \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X|Y}(x'|y') f_{Y|X}(v|x) k^*_{DUGS}(u, y'|x, v) \mu_Y(dv)\mu_X(du) \\
&\geq \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X|Y}(x'|y') f_{Y|X}(v|x) g(u, y') \mu_Y(dv)\mu_X(du) \\
&= \int_{\mathcal{X}} f_{X|Y}(x'|y') g(u, y') \left[ \int_{\mathcal{Y}} f_{Y|X}(v|x) \mu_Y(dv) \right] \mu_X(du) \\
&= \int_{\mathcal{X}} f_{X|Y}(x'|y') g(u, y') \mu_X(du) \\
&=: h(x', y') \, .
\end{aligned}$$

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

We now proceed to our counter-example.

**Proposition 10.** *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 $\pi$ be the distribution on $(0, \infty)^2$ with density function $f(x,y) = \frac{1}{2} e^{-y} \mathbf{1}_A(x,y)$, where $A$ is the union of the squares $(m, m+1] \times (m-1, m]$ for $m = 1, 2, 3 \ldots$ together with the infinite rectangle $(0, 1] \times (0, \infty)$ (see Figure 1).

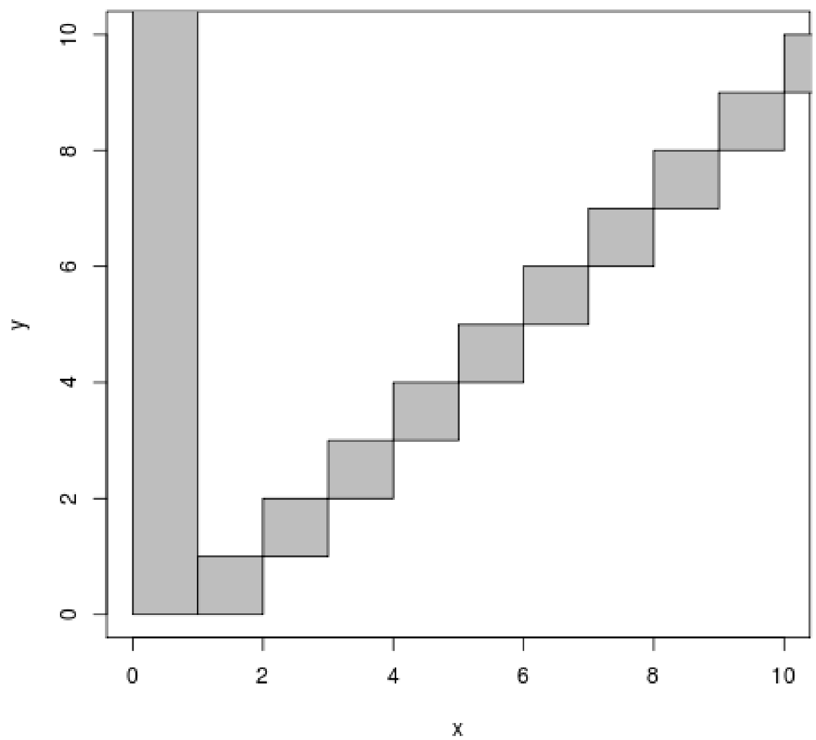


**Figure 1:** The region $A$ used in the proof of Proposition 10.

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

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

so

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

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 the one in the proof of Proposition 2. Specifically, if the $X$-update is

1-minorisable, then so is $P^*_{DUGS}$. Notice that if $m-1 < y \le m$, then

$$f_{X|Y}(x'|y) = \frac{1}{2}\mathbf{1}_{[0,1]\cup(m,m+1]}(x') \ge \frac{1}{2}\mathbf{1}_{[0,1]}(x') \; .$$

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

$$f_{X|Y}(x'|y) \ge \frac{1}{2}\mathbf{1}_{[0,1]}(x') \; .$$

From this, it is easy to see that $P^*_{DUGS}$ is minorised by the measure $2^{-1}\,\text{Uniform}[0,1]\times\text{Exp}(1)$. Hence, by Lemma 9, $P_{DUGS}$ is 2-minorisable and hence is uniformly ergodic.

Finally, we use a capacitance argument (see e.g. [15, 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 $\pi_X$ while $P_{CMH}$ is not reversible with respect to $\pi$, 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 $m \ge 3$ 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

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

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

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

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

Let $\kappa$ be the capacitance of $P^X_{CMH}$. Then

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

Hence, $P^X_{CMH}$ has capacitance zero, and hence has no spectral gap ([15, 39]), and hence fails to be geometrically ergodic [26]. Thus, $P_{CMH}$ also fails to be geometrically ergodic. □

### 3.3 Uniform Return Probabilities (URP)

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 CMH. To examine this question further, we present two examples involving the CIS version of CMH. The first shows that in general $P_{CIS}$ can fail to be even geometrically ergodic. The second shows that a slightly modified example is still uniformly ergodic even though $w$ is neither bounded nor $X$-bounded.

*Example* 1. Let $\pi = \text{Uniform}([0,1]^2)$ so that $f_{X|Y}(x|y) = f_X(x) = I(0 \leq x \leq 1)$ and $f_{Y|X}(y|x) = f_Y(y) = I(0 \leq y \leq 1)$. Consider CIS with proposal density $q(x'|y') = 2x'$. 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 $\pi = \text{Uniform}([0,1]^2)$, but now let $q(x'|y') = 2\{y' - x'\}$ 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

$w(x',y') = f_{X|Y}(x'|y')/q(x'|y') = 1/(2\{y'-x'\})$. Intuitively, the $x'$ proposals will usually be accepted unless $x$ is very close to $y'$. More precisely, let $S(x) = \{y \in [0,1] : \{y-x\} \geq 1/2\}$. 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 $h_X(x'|x)$ satisfies

$$\begin{aligned} h_X(x'|x) &= \int_{y' \in \mathcal{Y}} q(x'|y')\, \alpha(x',x,y')\, f_{Y|X}(y'|x)\, dy' \\ &\geq \int_{y' \in S(x)} q(x'|y')\, \min\left(1,\, \frac{w(x',y')}{w(x,y')}\right)\, f_{Y|X}(y'|x)\, dy' \\ &\geq \int_{y' \in S(x)} (2\{y'-x'\})(1/2)(1)\, dy' \\ &= \int_{y' \in S(x)} \{y'-x'\}\, dy' \, . \end{aligned}$$

Now, $S(x)$ is union of two disjoint intervals (or perhaps just one interval, if $x = 0$) within $[0,1]$, of total length $1/2$. Also, the mapping $y' \mapsto \{y'-x'\}$ is some re-arrangement of the identity mapping on $[0,1]$. So, since $\int_{y' \in S(x)} \{y'-x'\}\, dy'$ is an integral of some re-arrangement of the identity over some set of total length $1/2$, we must have $\int_{y' \in S(x)} \{y'-x'\}\, dy' \geq \int_0^{1/2} r\, dr = 1/8$. Hence, $h_X(x'|x) \geq 1/8$. Thus, for $A \in \mathcal{F}_X$,

$$P^X_{CIS}(x,A) \;\geq\; \int_A h_X(x'|x)\, \mu_X(dx') \;\geq\; \frac{1}{8}\, \mu_X(A) \, .$$

So, $P^X_{CIS}$ is 1-minorisable, so $P^X_{CIS}$ is uniformly ergodic, so $P_{CIS}$ 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.

**Theorem 11.** *Suppose that a CIS algorithm satisfies the following conditions:*

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

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

*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 $h_X(x'|x)$ now satisfies

$$\begin{aligned}
h_X(x'|x) &= \int_{y'\in\mathcal{Y}} q(x'|y')\,\alpha(x', x, y')\, f_{Y|X}(y'|x)\,\mu_Y(dy') \\
&\geq \int_{y'\in S(x)} q(x'|y')\,\min\left(1,\ \frac{w(x', y')}{w(x, y')}\right) f_{Y|X}(y'|x)\,\mu_Y(dy') \\
&\geq \int_{y'\in S(x)} q(x'|y')\,\min\left(1,\ \frac{f_{X|Y}(x'|y')}{q(x'|y')}\frac{1}{c}\right) f_{Y|X}(y'|x)\,\mu_Y(dy') \\
&\geq \int_{y'\in S(x)} \min\left(q(x'|y'),\ f_{X|Y}(x'|y')\frac{1}{c}\right) f_{Y|X}(y'|x)\,\mu_Y(dy') \\
&\geq \int_{y'\in S(x)} \min\left(1, \frac{1}{c}\right) g(x')\, f_{Y|X}(y'|x)\,\mu_Y(dy') \\
&\geq \min\left(1, \frac{1}{c}\right) g(x')\,\delta\ .
\end{aligned}$$

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

$$P^X_{CIS}(x, A) \ \geq\ \int_A h_X(x'|x)\,\mu_X(dx')\ \geq\ \int_A \min\left(1, \frac{1}{c}\right) g(x')\,\delta\,\mu_X(dx')\,.$$

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 CMH

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 (part 2), 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 12.** *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

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

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

$$P_{CMH}^X(x,A) \geq \delta\, P_{DUGS}^X(x,A), \qquad x \in \mathcal{X}, \quad A \in \mathcal{F}_X\ .$$

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

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

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

Then it is easily seen that there exists a constant $c > 0$ such that $q(x|y) \geq c f_{X|Y}(x|y)$. Hence, Theorem 12 shows that the conditional independence sampler is geometrically ergodic. □

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

**Theorem 13.** *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 4,

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

Hence,

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

Since both $P_{RSGS}$ and $P_{RCMH}$ are reversible with respect to $\pi$, the first claim now follows from Theorem 1. That the result holds for any selection probability then follows from Proposition 3. □

# 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 12) 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 [27], 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 $\sigma$-finite measure space for $i = 1, 2, \ldots, d$ $(d \geq 2)$, and let $(\mathcal{X}, \mathcal{F}, \mu)$ be the corresponding product space. Let $\pi$ be a target probability distribution on $(\mathcal{X}, \mathcal{F}, \mu)$, having density $f$ with respect to $\mu$. For $x \in \mathcal{X}$ and $1 \leq i \leq d$, set $x_{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$, $x_{[i]} = (x_1, \ldots, x_i)$ and $x^{[i]} = (x_i, \ldots, x_d)$. 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 $X_i | X_{(i)}$. Then the usual deterministic-scan Gibbs sampler (DUGS) has kernel

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

where

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

Now consider the situation where some coordinates $i$ are updated from the full-conditional Gibbs update $f_i(x'_i|x'_{[i-1]}, x^{[i+1]})$ as above, while other coordinates $i$ are updated from a Metropolis-Hastings update with proposal density $q_i(x'_i|x'_{[i-1]}, x_i, x^{[i+1]})$ and corresponding acceptance probability

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

In fact, if $q_i(x'_i|x'_{[i-1]}, x_i, x^{[i+1]}) = f_i(x'_i|x'_{[i-1]}, x^{[i+1]})$, 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 $g_i(w_i|z) = q_i(w_i|z_{[i-1]}, z_i, z^{[i+1]})\, \alpha_i(z_{[i-1]}, z_i, z^{[i+1]}, w_i)$. Thus, $g_i$ represents the absolutely continuous sub-kernel 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 $g_i(w_i|z) = f_i(w_i|z_{[i-1]}, z^{[i+1]})$, the full conditional density of coordinate $i$. Thus if we let

$$k_{CMH}(x'|x) \ = \ g_1(x'_1|x) \ g_2(x'_2|x'_1, x^{[2]}) \ \cdots \ g_d(x'_d|x'_{[d-1]}, x_d)\,,$$

then

$$P_{CMH}(x, A) \;\geq\; \int_A k_{CMH}(x'|x)\,\mu(dx'), \qquad A \in \mathcal{F}\;.$$

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

Notice that if each $g_i$ is a Gibbs update, i.e. $g_i(x'_i|x'_{[i-1]}, x^{[i]}) = f_i(x'_i|x'_{[i-1]}, x^{[i+1]})$, then $P_{CMH}$ is just the deterministic scan Gibbs sampler. That is, $P_{DUGS}$ is a special case of $P_{CMH}$ [30], 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

$$w_i(x'_{[i-1]}, x'_i, x_i, x^{[i+1]}) = \frac{f_i(x'_i|x'_{[i-1]}, x^{[i+1]})}{q_i(x'_i|x'_{[i-1]}, x_i, x^{[i+1]})}\;.$$

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 $C : \mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \to (0, \infty)$ 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 4.

**Lemma 14.**

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

*Proof.* Notice that for $i = 1, \ldots, d$

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

□

In light of the above lemma, the proofs of the following two theorems are similar to the proofs of Theorems 5 and 7. The corollaries follow as before.

**Theorem 15.** *If each $w_i$ is bounded and $P_{DUGS}$ is uniformly ergodic, then $P_{CMH}$ is uniformly ergodic.*

*Proof.* By Lemma 14 we have

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

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

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

and hence

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

The result now follows from Theorem 1. □

**Corollary 16.** *If each $w_i$ is bounded and $P_{DUGS}$ is uniformly ergodic, then $P_{RCMH}$ is uniformly ergodic for any selection probabilities.*

**Theorem 17.** *If each $w_i$ is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$-bounded, and there exists a non-negative function $g$ on $\mathcal{X}$, with $\mu\{x \in \mathcal{X} : g(x) > 0\} > 0$, such that*

$$
k_{DUGS}(x'|x) \; \geq \; g(x'), \qquad x \in \mathcal{X} , \tag{10}
$$

*then $P_{CMH}$ is uniformly ergodic.*

*Proof.* By Lemma 14 we have

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

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

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

Then using (10) we have

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

Letting

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

we have that if $A \in \mathcal{F}$, then

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

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

**Corollary 18.** *If each $w_i$ is $(\mathcal{X}_i \times \cdots \times \mathcal{X}_d)$-bounded, and condition* (10) *holds, then $P_{RCMH}$ is uniformly ergodic for any selection probabilities.*

Furthermore, Proposition 3 extends easily to the general case.

**Proposition 19.** *If $P_{RSGS}$ is geometrically ergodic for some selection probability, then it is geometrically ergodic for all selection probabilities.*

Just as with Theorem 13, we can also give sufficient conditions for geometric ergodicity of $P_{RCMH}$ in terms of the geometric ergodicity of $P_{RSGS}$.

**Theorem 20.** *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, 32].

To fix ideas, consider a one-dimensional diffusion satisfying $dX_t = dB_t + \alpha(X_t)\, dt$ for $0 \le t \le 1$, where $\alpha : \mathbb{R} \to \mathbb{R}$ is a $C^1$ function. Suppose 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 $\mathbf{P}_\theta$ be the law of the diffusion starting at $X_0$, conditional on $\theta$, and let $\mathbf{W}$ be the law of Brownian motion starting at $X_0$. Then by Girsanov's Formula [see e.g. 34], the density of $\mathbf{P}_\theta$ with respect to $\mathbf{W}$ satisfies (writing $X_{[0,1]}$ for $\{X_t\}_{0 \le t \le 1}$) that

$$G_\theta(X_{[0,1]}) \ := \ \frac{d\mathbf{P}_\theta}{d\mathbf{W}}(X_{[0,1]}) \ = \ \exp[A(X_1) - A(X_0) - \int_0^1 \phi_\theta(X_s) ds] \, , \tag{11}$$

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

Furthermore, if $\widetilde{\mathbf{P}}$ is the law of the diffusion conditional on the observed values of $X_0$ and $X_1$, and $\widetilde{\mathbf{W}}$ is the law of Brownian motion conditional on the same observed values of $X_0$ and $X_1$ (i.e., of the corresponding Brownian bridge process), then $\frac{d\widetilde{\mathbf{P}}}{d\widetilde{\mathbf{W}}}$ is still proportional to the same density $G$ from (11).

Assume now that $\alpha(x) = \sum_{i=1}^m p_i(x)\theta_i = p^T\theta$, 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 $\theta \sim$MVN$(0, \Sigma_0)$ on the vector $\theta$, for some strictly positive-definite symmetric $m \times m$ covariance matrix $\Sigma_0$. Then conditional on $X_0$ and $X_1$, and letting $X_{miss} = \{X_s : 0 < s < 1\}$ be the missing (unobserved) part of the diffusion's sample path, the joint posterior density of the pair $(\theta, X_{miss})$ is proportional to

$$e^{-\theta^T \Sigma_0^{-1} \theta/2} \ G_\theta(X_{[0,1]}) = \exp\left[ -\frac{1}{2} \left( \theta^T \Sigma_0^{-1} \theta + \int_0^1 \sum_{i=1}^m \sum_{j=1}^m p_i(X_s)\, p_j(X_s)\, \theta_i\, \theta_j \right.\right.$$
$$\left.\left. + \int_0^1 \sum_{i=1}^m p_i'(X_s)\, \theta_i\, ds \right) \right]$$

We can write this joint posterior density as being proportional to

$$\exp\left[-\frac{1}{2}\theta^T V^{-1}\theta - r^T\theta\right] , \tag{12}$$

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

$$V^{-1} = \Sigma_0^{-1} + \int_0^1 p(X_s)(p(X_s))^T ds \ . \tag{13}$$

Then, since

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

(using that $V^T = V$, and that $r^T\theta = \theta^T r$ is a scalar), equation (12) in turn implies that the conditional distribution $\theta|X_{miss}$ is given by:

$$\theta \,|\, X_{miss} \ \sim \ \mathrm{MVN}(-Vr,\, V)\,. \tag{14}$$

Now, suppose we wish to sample the pair $(\theta, X_{miss})$ from its posterior density (12). We first consider using a deterministic-scan Gibbs sampler (DUGS), in which we alternately sample $\theta|X_{miss}$ and then $X_{miss}|\theta$.

**Lemma 21.** *Assume the $p_i$ and $p_i'$ functions are all bounded, i.e.*

$$\max_{1\le i\le m}\ \sup_{x\in\mathbb{R}}\ \max(|p_i(x)|, |p_i'(x)|) \ < \ \infty\,. \tag{15}$$

*Then the deterministic-scan Gibbs sampler (DUGS) for the pair $(\theta, X_{miss})$ is 1-minorisable.*

*Proof.* In light of Proposition 2, it suffices to show that the $\theta$ updates, as carried out through (14), 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 $\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{M}$ (where $\mathbf{M}$ denotes the space of positive definite $m \times m$ matrices). Therefore by the standard compactness argument, if $A$ is any compact set in $\mathbf{R}^m \times \mathbf{M}$ then for all $\theta \in \mathbf{R}^m$

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

thus providing a minorisation measure. It remains therefore to show that given all possible diffusion trajectories, the mean $(-Vr)$ and variance $(V)$ in (14) are uniformly contained in bounded regions, with the determinant of the variance bounded away from zero. Note that (15) and the definition of $V$ imply immediately that $V$ is uniformly bounded proving the first part of this. Moreover, showing that $\det(V)$ is uniformly bounded away from zero is equivalent to a uniform upper bound on $\det(V^{-1})$. However this also follows trivially from (14). Thus it follows that the $\theta$ update is 1-minorisable. $\square$

The above lemma shows that DUGS for the pair $(\theta, X_{miss})$ is uniformly ergodic. However, in practice it is entirely infeasible to sample the entire path $X_{miss}$ from its correct conditional distribution given $\theta$. Thus, to sample the pair $(\theta, X_{miss})$ from the posterior density (12), we instead consider using a conditional independence sampler (CIS). Here $\theta$ plays the role of $Y$, and $X_{miss}$ plays the role of $X$. We shall alternately update $\theta$ from its full conditional distribution conditional on the current value of $X_{miss}$ (which is easy to implement in practice, since $\theta | X_{miss}$ follows a Gaussian distribution), and then update $X_{miss}$ using a conditional Metropolis-Hastings update step with proposal distribution $q(X_{miss}|\theta)$ given by the corresponding Brownian bridge, i.e. with $q(X_{miss}|\theta) = \widetilde{\mathbf{W}}$ (which can be implemented in practice by e.g. discretising the time interval $[0, 1]$ and then using the Gaussian conditional distributions of 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

$$w(x_{miss}, \theta) \;=\; \frac{f_{X_{miss}|\theta}(x_{miss}|\theta)}{q(x_{miss}|\theta)} \;=\; \frac{d\widetilde{\mathbf{P}}}{d\widetilde{\mathbf{W}}}(X_{[0,1]}) \;=\; h(\theta) G_\theta(X_{[0,1]})\,.$$

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 22.** *For the above CIS algorithm, assuming* (15)*, the weights are $X$-bounded, i.e.* $\sup_x w(x, \theta) < \infty$ *for each fixed $\theta$.*

*Proof.* From (11), we can write

$$\begin{aligned} w(x_{miss}, \theta) \;&=\; h(\theta) G_\theta(X_{[0,1]}) \;=\; h(\theta) \exp[A(X_1) - A(X_0) - \int_0^1 \phi_\theta(X_s) ds] \\ &\le h(\theta) \exp[A(X_1) - A(X_0)] \exp\{-\inf_x \phi_\theta\} \qquad (16) \end{aligned}$$

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

$$\phi_\theta = \frac{1}{2}\left[\theta^T\left(\int p(X_s)(p(X_s))^T ds\right)\theta + \left(\int (p'(X_s))^T ds\right)\theta\right].$$

Hence, by the boundedness of $p_i$ and $p_i'$ from (15), 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 23.** *Assuming* (15)*, the above CIS algorithm on* $(X_{miss}, \theta)$*, conditional on the observed values* $X_0$ *and* $X_1$*, is uniformly ergodic.*

*Proof.* This follows immediately from Theorem 7, in light of Lemmas 21 and 22 above. □

## 5.1 Generalisation to more data

In practice, fitting a diffusion model, we would almost certainly possess multiple data, $X_{obs} = (X_{t_0}, X_{t_1}, X_{t_2}, \ldots, X_{t_N})$, observed at times $t_0, t_1, t_2, \ldots, t_N$, leading in turn to missing diffusion segments $X_{miss,i} = \{X_t : t_{i-1} < t < t_i\}$ for $1 \leq i \leq N$. 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 23 easily generalises.

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

1. Given $X_{obs}$ and $\{X_{miss,i}\}_{1\leq i\leq N}$, simulate $\theta$ from its full conditional as given in (14).

2. Sequentially for $i = 1, 2, \ldots, N$, propose an update of $X_{miss,i}$ conditional on $X_{obs}$ and $\theta$ from 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 $\theta$, the $\{X_{miss,i}\}_{1\leq i\leq N}$ 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 23.

**Theorem 24.** *Assuming* (15)*, the above CIS algorithm on* $(X_{miss}, \theta)$*, conditional on the observed values* $X_{t_1}, X_{t_2}, \ldots, X_{t_N}$*, is uniformly ergodic.*

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

## References

[1] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors. *Handbook of Markov Chain Monte Carlo.* CRC Press, Boca Raton, FL, 2011.

[2] Kung Sik Chan and Charles J. Geyer. Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758, 1994.

[3] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability*, 3:696–730, 1993.

[4] Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993, 2001.

[5] James M. Flegal, Murali Haran, and Galin L. Jones. Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260, 2008.

[6] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[7] James P. Hobert and Charles J. Geyer. Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67:414–430, 1998.

[8] S. F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications*, 85:341–361, 2000.

[9] Alicia A. Johnson and Galin L. Jones. Gibbs sampling for a Bayesian hierarchical version of the general linear mixed model. *Electronic Journal of Statistics*, 4:313–333, 2010.

[10] Alicia A. Johnson, Galin L. Jones, and Ronald C. Neath. Component-wise Markov chain Monte Carlo: Uniform and geometric ergodicity under mixing and composition. *Statistical Science* (to appear), 2013.

[11] Galin L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

[12] Galin L. Jones, Murali Haran, Brian S. Caffo, and Ronald Neath. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547, 2006.

[13] Galin L. Jones and James P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334, 2001.

[14] Galin L. Jones and James P. Hobert. Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32:784–817, 2004.

[15] G. F. Lawler and A. D. Sokal. Bounds on the $L^2$ spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American Mathematical Society*, 309:557–580, 1988.

[16] Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6:113–119, 1996.

[17] Jun S. Liu, Wing Hung Wong, and Augustine Kong. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81:27–40, 1994.

[18] Dobrin Marchev and James P. Hobert. Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's $t$ model. *Journal of the American Statistical Association*, 99:228–238, 2004.

[19] Kerrie L. Mengersen and Richard L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121, 1996.

[20] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[21] Omiros Papaspiliopoulos and Gareth Roberts. Stability of the Gibbs sampler for Bayesian hierarchical models. *The Annals of Statistics*, 36:95–117, 2008.

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

[23] Christian P. Robert. Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science*, 10:231–253, 1995.

[24] G. O. Roberts and S. K. Sahu. Updating schemes, correlation structure, blocking and parametrization for the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 59:291–317, 1997.

[25] Gareth O. Roberts and Nicholas G. Polson. On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 56:377–384, 1994.

[26] Gareth O. Roberts and Jeffrey S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25, 1997.

[27] Gareth O. Roberts and Jeffrey S. Rosenthal. Two convergence properties of hybrid samplers. *The Annals of Applied Probability*, 8:397–407, 1998.

[28] Gareth O. Roberts and Jeffrey S. Rosenthal. Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society,* Series B, 61:643–660, 1999.

[29] Gareth O. Roberts and Jeffrey S. Rosenthal. Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, 28:489–504, 2001.

[30] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[31] Gareth O. Roberts and Jeffrey S. Rosenthal. Quantitative non-geometric convergence bounds for independence samplers. *Methodology and Computing in Applied Probability*, 13:391–403, 2011.

[32] Gareth O. Roberts and Osnat Stramer. Bayesian inference for incomplete observations of diffusion processes. *Biometrika*, 88:603–621, 2001.

[33] Gareth O. Roberts and Richard L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.

[34] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: vol. 1, Foundations*. Wiley, New York, second edition, 1994.

[35] Jeffrey S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566, 1995.

[36] Jeffrey S. Rosenthal. Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6:269–275, 1996.

[37] Vivekananda Roy and James P. Hobert. Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 69:607–623, 2007.

[38] Mark J. Schervish and Bradley P. Carlin. On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics*, pages 111–127, 1992.

[39] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[40] Richard L. Smith and Luke Tierney. Exact transition probabilities for the independence Metropolis sampler. Technical report, University of North Caolina, 1996.

[41] Aixin Tan and James P. Hobert. Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics*, 18:861–878, 2009.

[42] Luke Tierney. Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762, 1994.

[43] Luke Tierney. A note on Metropolis–Hastings kernels for general state spaces. *The Annals of Applied Probability*, 8:1–9, 1998.