Bernoulli **28**(2), 2022, 859–885
https://doi.org/10.3150/21-BEJ1369

Convergence rates of two-component MCMC samplers

QIAN QIN[*] and GALIN L. JONES[†]

School of Statistics, University of Minnesota, Minneapolis, USA. E-mail: [*]*qqin@umn.edu;* [†]*galin@umn.edu*

Component-wise MCMC algorithms, including Gibbs and conditional Metropolis-Hastings samplers, are commonly used for sampling from multivariate probability distributions. A long-standing question regarding Gibbs algorithms is whether a deterministic-scan (systematic-scan) sampler converges faster than its random-scan counterpart. We answer this question when the samplers involve two components by establishing an exact quantitative relationship between the L^2 convergence rates of the two samplers. The relationship shows that the deterministic-scan sampler converges faster. We also establish qualitative relations among the convergence rates of two-component Gibbs samplers and some conditional Metropolis-Hastings variants. For instance, it is shown that if some two-component conditional Metropolis-Hastings samplers are geometrically ergodic, then so are the associated Gibbs samplers.

Keywords: Deterministic-scan; geometric ergodicity; Gibbs; Metropolis-within-Gibbs; random-scan

1. Introduction

Markov chain Monte Carlo (MCMC) algorithms are useful for sampling from complicated distributions (Brooks *et al.* [6]). Component-wise MCMC algorithms, such as Gibbs samplers and conditional Metropolis-Hastings (CMH) samplers, sometimes called Metropolis-within-Gibbs, are among the most useful in multivariate settings. We study the convergence rates of two-component Gibbs samplers and the case where the components may be updated using Metropolis-Hastings, paying particular attention to the relationship between the convergence rates of the Markov chains.

Investigating the convergence rates of the underlying Markov chains is important for ensuring a reliable simulation effort (Geyer [18], Flegal, Haran and Jones [15], Jones and Hobert [29], Vats *et al.* [63]). If the Markov chain converges sufficiently fast, then, under moment conditions, a central limit theorem holds (Chan and Geyer [7], Doss *et al.* [11], Hobert *et al.* [24], Jones [28], Robertson *et al.* [50]). Additionally, asymptotically valid Monte Carlo standard errors are available (Dai and Jones [8], Flegal and Jones [16], Jones *et al.* [32], Vats, Flegal and Jones [61,62]).

Let $\Pi(\mathrm{d}x, \mathrm{d}y)$ be a joint probability distribution having support $\mathsf{X} \times \mathsf{Y}$ and let $\Pi_{X|Y}(\mathrm{d}x|y)$, $y \in \mathsf{Y}$, and $\Pi_{Y|X}(\mathrm{d}y|x)$, $x \in \mathsf{X}$, be full conditional distributions. There are many potential component-wise MCMC algorithms having Π as their invariant distribution. When it is possible to simulate from the conditionals, it is natural to use a Gibbs sampler. One version is the deterministic-scan Gibbs (DG) sampler, which is described in Algorithm 1.

An alternative is the random-scan Gibbs (RG) sampler which is described in Algorithm 2.

Two-component Gibbs samplers are surprisingly useful and widely applicable in the analysis of sophisticated Bayesian statistical models. In particular, they arise naturally in data augmentation settings (Hobert [22], Tanner and Wong [57], van Dyk and Meng [60]).

There is abundant study of the convergence properties of Gibbs samplers, both in the general case (see Liu, Wong and Kong [37], Roberts and Polson [42], Liu, Wong and Kong [38]) and for two-component Gibbs samplers in specific statistical settings; see, among many others, Diaconis, Khare and Saloff-Coste [9], Doss and Hobert [10], Ekvall and Jones [13], Hobert and Geyer [23], Johnson

1350-7265

Algorithm 1 Deterministic-scan Gibbs sampler

Input: Current value $(X_n, Y_n) = (x, y)$.
Draw Y_{n+1} from $\Pi_{Y|X}(\cdot|x)$, and call the observed value y'.
Draw X_{n+1} from $\Pi_{X|Y}(\cdot|y')$.
Set $n = n + 1$.

Algorithm 2 Random-scan Gibbs sampler with selection probability $r \in (0, 1)$

Input: Current value $(X_n, Y_n) = (x, y)$.
Draw $U \sim \text{Bernoulli}(r)$, and call the observed value u.
If $u = 1$, draw X_{n+1} from $\Pi_{X|Y}(\cdot|y)$, and set $Y_{n+1} = y$.
If $u = 0$, draw Y_{n+1} from $\Pi_{Y|X}(\cdot|x)$, and set $X_{n+1} = x$.
Set $n = n + 1$.

and Jones [25,26], Jones and Hobert [30], Khare and Hobert [34], Marchev and Hobert [39], Roy [52], Tan and Hobert [55], Wang and Roy [65,66]. However, there is not yet an answer to the following basic question: which converges faster, a deterministic- or random-scan Gibbs sampler?

There exist some qualitative results related to this question (see Johnson, Jones and Neath [27], Tan, Jones and Hobert [56]). For instance, Roberts and Rosenthal's [43] Proposition 3.2 states that a random-scan Gibbs sampler is uniformly ergodic whenever an associated deterministic-scan Gibbs sampler is too. There is also literature devoted to finding the convergence rates of various Gibbs samplers when Π is Gaussian, or approximately Gaussian (see, e.g., Amit [1,2], Amit and Grenander [3], Roberts and Sahu [47]) or in the finite discrete state space setting (Fishman [14]). However, in general, the relationship between the convergence rates of deterministic- and random-scan Gibbs samplers is poorly understood.

A related question is addressed by Andrieu [4], who shows that the DG sampler yields sample means with smaller asymptotic variances than its random-scan counterpart, assuming that, in the RG sampler, the selection probability is $r = 1/2$ (see, also, Greenwood, McKeague and Wefelmeyer [19]). On the other hand, the author remarks that making such a comparison in terms of convergence times is unlikely to bear fruit (Andrieu [4], page 720). This is because there are examples suggesting that, when the Gibbs sampler has a large number of components, there is no definite answer to the question above (Roberts and Rosenthal [46]).

We give an exact solution to the question in the two-component setting. Indeed, we develop a quantitative relationship between the convergence rates of the two types of Gibbs samplers, and show that the deterministic-scan sampler converges faster than its random-scan counterpart no matter the selection probability in the random scan. This result is described now, but the full details are dealt with carefully later. The L^2 convergence rate of a Markov chain is a number in [0, 1], with smaller rates indicating faster convergence. Let $\rho(P_{\text{DG}})$ be the L^2 convergence rate of the DG sampler, and, $\rho(P_{\text{RG}})$, that of the RG sampler. We show that

$$\rho(P_{\text{RG}}) = \frac{1 + \sqrt{1 - 4r(1-r)[1 - \rho(P_{\text{DG}})]}}{2}. \tag{1}$$

There are some easy, but noteworthy, consequences of this result. Notice that (i) $\rho(P_{\text{RG}}) \in [1/2, 1]$ while $\rho(P_{\text{DG}}) \in [0, 1]$; (ii) as either $\rho(P_{\text{DG}})$ or $\rho(P_{\text{RG}})$ increases so does the other; (iii) if $\rho(P_{\text{DG}}) < 1$, then $\rho(P_{\text{RG}}) > \rho(P_{\text{DG}})$, but $\rho(P_{\text{DG}}) = 1$ if and only if $\rho(P_{\text{RG}}) = 1$; and (iv) the optimal selection probability

Algorithm 3 Deterministic-scan CMH sampler

1: *Input:* Current value $(X_n, Y_n) = (x, y)$
2: Draw Y_{n+1} from $\Pi_{Y|X}(\cdot|x)$, and call the observed value y'.
3: Draw a random element Z from $q(\cdot|x, y')$, and call the observed value z. With probability

$$a(z; x, y') = \min\left\{1, \frac{\pi_{X|Y}(z|y')q(x|z, y')}{\pi_{X|Y}(x|y')q(z|x, y')}\right\},$$

set $X_{n+1} = z$; with probability $1 - a(z; x, y')$, set $X_{n+1} = x$.
4: Set $n = n + 1$.

Algorithm 4 Random-scan CMH sampler with selection probability $r \in (0, 1)$

1: *Input:* Current value $(X_n, Y_n) = (x, y)$.
2: Draw $U \sim \text{Bernoulli}(r)$, and call the observed value u.
3: If $u = 1$, draw a random element Z from $q(\cdot|x, y)$, and call the observed value z. With probability

$$a(z; x, y) = \min\left\{1, \frac{\pi_{X|Y}(z|y)q(x|z, y)}{\pi_{X|Y}(x|y)q(z|x, y)}\right\},$$

set $X_{n+1} = z$; with probability $1 - a(z; x, y)$, set $X_{n+1} = x$. Set $Y_{n+1} = y$.
4: If $u = 0$, draw Y_{n+1} from $\Pi_{Y|X}(\cdot|x)$, and set $X_{n+1} = x$.
5: Set $n = n + 1$.

for P_{RG} is $r = 1/2$ in which case

$$\rho(P_{\text{RG}}) = \frac{1 + \sqrt{\rho(P_{\text{DG}})}}{2}.$$

In Section 4 we generalize this discussion and show that the DG sampler converges faster even after taking into account computation time. Indeed, if k_{D} and k_{R} are the number of iterations that can be run by DG and RG samplers, respectively, in unit time, then (1) implies that $\rho(P_{\text{DG}})^{k_{\text{D}}} \le \rho(P_{\text{RG}})^{k_{\text{R}}}$ for any selection probability.

Perhaps the most common type of MCMC sampler in applications are conditional Metropolis-Hastings (CMH) samplers. These Markov chains arise when it is infeasible to sample from at least one of the conditional distributions associated with Π so that at least one Metropolis-Hastings update must be used. Assume that $\Pi_{Y|X}$ and $\Pi_{X|Y}$, respectively, admit density functions $\pi_{Y|X}$ and $\pi_{X|Y}$. Let $q(\cdot|x, y)$, $(x, y) \in \mathsf{X} \times \mathsf{Y}$, be a proposal density function on X. A deterministic-scan CMH (DC) sampler we study is now described in Algorithm 3, but a more general algorithm is considered in Section 5.2.

There is an obvious alternative random-scan CMH (RC) sampler in Algorithm 4.

Despite their utility, compared to Gibbs samplers there has been little investigation of CMH Markov chains (Fort *et al.* [17], Herbei and McKeague [21], Johnson, Jones and Neath [27], Jones, Roberts and Rosenthal [31], Rosenthal and Rosenthal [51], Roberts and Rosenthal [43,44]) but what there is tends not to focus on specific statistical models. For example, Johnson, Jones and Neath's [27] Theorem 3 states that if a deterministic scan component-wise Markov chain is uniformly ergodic, then so is its random-scan counterpart, thus generalizing the result proved for Gibbs samplers by Roberts and Rosenthal [43], which was described previously.

Both versions of Gibbs samplers are special cases of the respective versions of CMH samplers. Thus it is plausible that there should be some relationship among the convergence rates of the Markov

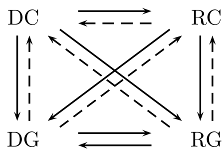

Figure 1. Relationship among two-component Gibbs samplers and their CMH variants in terms of L^2 geometric ergodicity.

chains of Algorithms 1–4, especially if the CMH samplers are "close" to the Gibbs samplers. There are a few results in this direction. For example, there are sufficient conditions which ensure that if the RG Markov chain is geometrically ergodic, then so is the RC Markov chain (Jones, Roberts and Rosenthal [31], Theorem 6). However, these relationships are not well understood in general and the following question has not been addressed satisfactorily: if one of the four basic component-wise samplers is geometrically ergodic, then, in general, which of the remaining three are also geometrically ergodic?

We give an answer to this question by developing qualitative relationships among the convergence rates of the DG, RG, DC, and RC samplers, which are depicted in Figure 1. Here, we consider L^2 geometric ergodicity. A Markov chain is L^2 geometrically ergodic if its L^2 convergence rate is strictly less than 1. Under regularity conditions, L^2 geometric ergodicity is equivalent to the usual notion of geometric ergodicity defined in terms of the total variation distance. (This equivalence will be made precise in Section 3.) In Figure 1, a solid arrow from one sampler to another means that, if the former is L^2 geometrically ergodic, then so is the latter. A dashed arrow means that L^2 geometric ergodicity of the former only implies that of the latter under appropriate conditions on the proposal density $q(\cdot|x, y)$. One of these conditions is condition 5.1 in Section 5.

Figure 1 yields the following. The DG sampler is L^2 geometrically ergodic if and only if the RG sampler is. If the RC sampler is L^2 geometrically ergodic for some proposal density, then so are the DG and RG samplers. If the DC sampler is L^2 geometrically ergodic for some proposal density, then so is the RC sampler with the same proposal density. The relations depicted in Figure 1 hold regardless of the selection probabilities for the random-scan samplers.

The remainder is organized as follows. Section 2 contains some general theoretical background. In Section 3, we lay out some basic properties of the four types of samplers. In Section 4, we derive (1), and discuss its implications. In Section 5, we establish the relations shown in Figure 1 along with additional connections with more general CMH samplers. We give some final remarks in Section 6. Some technical details are relegated to the appendices.

2. Preliminary Markov chain theory

Let $(\mathsf{Z}, \mathcal{F})$ be a measurable space and let P be a Markov transition kernel (Mtk), that is, let $P : \mathsf{Z} \times \mathcal{F} \to [0, 1]$ be such that for each $z \in \mathcal{Z}$, $P(z, \cdot)$ is a probability measure and for each $A \in \mathcal{F}$, $P(\cdot, A)$ is measurable. For a positive integer n, denote the n-step transition kernel associated with P by P^n, so that $P^1 = P$, and

$$P^{n+1}(z, A) = \int_{\mathsf{Z}} P(z', A) P^n(z, \mathrm{d}z')$$

for $z \in \mathsf{Z}$ and $A \in \mathcal{F}$. If ω is a probability measure on $(\mathsf{Z}, \mathcal{F})$ and $A \in \mathcal{F}$, define

$$(\omega P)(A) = \int \omega(dz) P(z, A).$$

Say ω is invariant for P if $\omega P = \omega$. If

$$P(z, dz')\omega(dz) = P(z', dz)\omega(dz'), \tag{2}$$

then P is said to be reversible with respect to ω. Integrating both sides of the equality in (2) shows that ω is invariant for P.

For a measurable function $f : \mathsf{Z} \to \mathbb{R}$ and a probability measure $\mu : \mathcal{F} \to [0, 1]$, define

$$(Pf)(z) = \int f(z')P(z, dz') \quad \text{and} \quad \mu f = \int_{\mathsf{Z}} f(z)\mu(\mathrm{d}z).$$

Assume that ω is invariant for P. Let $L^2(\omega)$ be the set of measurable real functions f that are square integrable with respect to ω and let $L_0^2(\omega)$ be the set of functions $f \in L^2(\omega)$ such that $\omega f = 0$. For $f, g \in L^2(\omega)$, define their inner product to be

$$\langle f, g\rangle_\omega = \int_{\mathsf{Z}} f(z)g(z)\,\omega(\mathrm{d}z)\,,$$

and let $\|f\|_\omega^2 = \langle f, f\rangle_\omega$. Then $(L^2(\omega), \langle\cdot,\cdot\rangle_\omega)$ and $(L_0^2(\omega), \langle\cdot,\cdot\rangle_\omega)$ form two real Hilbert spaces. For any $f \in L_0^2(\omega)$, we have $Pf \in L_0^2(\omega)$. Thus, P can be regarded as a linear operator on $L_0^2(\omega)$. Let

$$\|P\|_\omega = \sup_{f \in L_0^2(\omega),\, \|f\|_\omega = 1} \|Pf\|_\omega\,.$$

By the Cauchy-Schwarz inequality, $\|P\|_\omega \le 1$. When P is reversible with respect to ω, P, as an operator on $L_0^2(\omega)$, is self-adjoint so that $\langle Pf_1, f_2\rangle_\omega = \langle f_1, Pf_2\rangle_\omega$ for $f_1, f_2 \in L_0^2(\omega)$, and

$$\|P\|_\omega = \sup_{f \in L_0^2(\omega),\, \|f\|_\omega = 1} |\langle Pf, f\rangle_\omega|.$$

Moreover, if P is self-adjoint, then for each positive integer n,

$$\|P^n\|_\omega = \|P\|_\omega^n$$

(see, e.g., Helmberg [20], §30 Corollary 8.1, §31 Corollary 2.1). Say P is non-negative definite if it is self-adjoint, and $\langle Pf, f\rangle_\omega \ge 0$ for each $f \in L_0^2(\omega)$.

For two probability measures μ and ν on $(Z, \mathcal{F})$, define their L^2 (or χ^2) distance to be

$$\|\mu - \nu\|_\omega = \sup_{f \in L_0^2(\omega),\, \|f\|_\omega = 1} |\mu f - \nu f|\,.$$

Let $L_*^2(\omega)$ be the set of probability measures μ such that $\mathrm{d}\mu/\mathrm{d}\omega \in L^2(\omega)$. When $\mu, \nu \in L_*^2(\omega)$,

$$\|\mu - \nu\|_\omega = \sup_{f \in L_0^2(\omega),\, \|f\|_\omega = 1} \left\langle \frac{\mathrm{d}\mu}{\mathrm{d}\omega} - \frac{\mathrm{d}\nu}{\mathrm{d}\omega}, f\right\rangle_\omega = \left\| \frac{\mathrm{d}\mu}{\mathrm{d}\omega} - \frac{\mathrm{d}\nu}{\mathrm{d}\omega} \right\|_\omega .$$

The L^2 convergence rate of the Markov chain associated with P, denoted by $\rho(P)$, is defined to be the infimum of $\rho \in [0, 1]$ such that, for each $\mu \in L_*^2(\omega)$, there exists $C_\mu < \infty$ such that, for each positive integer n,

$$\|\mu P^n - \omega\|_\omega < C_\mu \rho^n.$$

When $\rho(P) < 1$, we say that the Markov chain is L^2 geometrically ergodic, or more simply, P is L^2 geometrically ergodic. The following is a direct consequence of Roberts and Rosenthal's [43] Theorem 2.1 and we will use it extensively.

Lemma 2.1. *If P is reversible with respect to ω, then $\rho(P) = \|P\|_\omega$.*

The following comparison lemma will be useful in conjunction with Lemma 2.1.

Lemma 2.2. *Let P_1 and P_2 be Mtks on $(\mathsf{Z}, \mathcal{F})$ having a common stationary distribution ω. Suppose further that $\|P_2\|_\omega < 1$ and there exists $\delta > 0$ such that, for $z \in \mathsf{Z}$ and $A \in \mathcal{F}$, $P_1(z, A) \geq \delta P_2(z, A)$. Then $\|P_1\|_\omega < 1$.*

Proof. Without loss of generality, assume that $\delta < 1$. Let $R(z, A) = (1-\delta)^{-1}(P_1(z, A) - \delta P_2(z, A))$. Then $R(z, A)$ defines an Mtk such that $\omega R = \omega$. By Cauchy-Schwarz, $\|R\|_\omega \leq 1$. By the triangle inequality, $\|P_1\|_\omega \leq \delta\|P_2\|_\omega + (1-\delta)\|R\|_\omega < 1$. □

We can use these lemmas to obtain a generalization of Jones, Roberts and Rosenthal's [31] Proposition 2. This will allow us to treat the selection probabilities in the random-scan algorithms as arbitrary when studying their qualitative convergence rates.

Proposition 2.3. *Let P_1 and P_0 be Mtks on $(\mathsf{Z}, \mathcal{F})$ such that for any $0 < r < 1$ the mixture kernel $P_r = rP_1 + (1-r)P_0$ is reversible with respect to ω. If $\rho(P_{r_0}) < 1$ for some $r_0 \in (0, 1)$, then $\rho(P_r) < 1$ for every $r \in (0, 1)$.*

Proof. For each $z \in \mathsf{Z}$ and $A \in \mathcal{F}$,

$$P_r(z, A) \geq \min\left\{\frac{r}{r_0}, \frac{1-r}{1-r_0}\right\} P_{r_0}(z, A).$$

Since P_r is reversible with respect to ω for all $r \in (0, 1)$, the claim follows from Lemmas 2.1 and 2.2. □

Remark 2.4. While we will not require it, it is straightforward to extend the proof of Proposition 2.3 to the setting where there is an arbitrary, but finite, number of Mtks in the mixture.

We are now in position to begin our study of the algorithms defined in Section 1.

3. Basic properties of two-component samplers

We begin by defining the Markov transition kernels for the four algorithms described in Section 1 along with some related Markov chains that will be useful later. Then we will turn our attention to some basic properties of the operators and total variation norms for these Markov chains.

Suppose $(\mathsf{X} \times \mathsf{Y}, \mathcal{F}_X \times \mathcal{F}_Y)$ is a measurable space with a joint probability distribution $\Pi(\mathrm{d}x, \mathrm{d}y)$. Let $\Pi_X(\mathrm{d}x)$ and $\Pi_Y(\mathrm{d}y)$ be the associated marginal distributions, and, $\Pi_{X|Y}(\mathrm{d}x|y)$ and $\Pi_{Y|X}(\mathrm{d}y|x)$, the full conditional distributions. To avoid trivial cases we make the following standing assumption.

Assumption 3.1. There exist $A_1, A_2 \in \mathcal{F}_X$ and $B_1, B_2 \in \mathcal{F}_Y$ such that $A_1 \cap A_2 = \varnothing$, $B_1 \cap B_2 = \varnothing$, and that $\Pi_X(A_1) > 0$, $\Pi_X(A_2) > 0$, $\Pi_Y(B_1) > 0$, $\Pi_Y(B_2) > 0$.

When Assumption 3.1 is violated, at least one of $\mathcal{F}_X$ and $\mathcal{F}_Y$ contain only sets of measure zero or one, and all the problems we study become essentially trivial.

Letting Π, Π_X, or Π_Y play the role of ω from Section 2, as appropriate, allows us to consider the Mtks defined in the sequel as linear operators on the appropriate Hilbert spaces. Assumption 3.1 ensures $L_0^2(\Pi)$, $L_0^2(\Pi_X)$, and $L_0^2(\Pi_Y)$ contain non-zero elements.

3.1. Markov transition kernels

The Mtk for the DG sampler is

$$P_{\text{DG}}((x,y),(\mathrm{d}x',\mathrm{d}y')) = \Pi_{X|Y}(\mathrm{d}x'|y')\Pi_{Y|X}(\mathrm{d}y'|x).$$

Now P_{DG} has Π as its invariant distribution, but it is not reversible with respect to Π. If δ_x and δ_y are point masses at x and y, respectively, then the Mtk for the RG sampler is

$$P_{\text{RG}}((x,y),(\mathrm{d}x',\mathrm{d}y')) = r\Pi_{X|Y}(\mathrm{d}x'|y)\delta_y(\mathrm{d}y') + (1-r)\Pi_{Y|X}(\mathrm{d}y'|x)\delta_x(\mathrm{d}x').$$

It is well known that P_{RG} is reversible with respect to Π and hence has Π as its invariant distribution. Now let P_{MH} denote the Metropolis-Hastings Mtk (Tierney [58,59]) which is reversible with respect to the full conditional $\Pi_{X|Y}$. Then the Mtk for the DC sampler is

$$P_{\text{DC}}((x,y),(\mathrm{d}x',\mathrm{d}y')) = P_{\text{MH}}(\mathrm{d}x'|x,y')\,\Pi_{Y|X}(\mathrm{d}y'|x).$$

Note that P_{DC} has Π as its invariant distribution, but it is not reversible with respect to Π. The Mtk for the RC sampler is

$$P_{\text{RC}}((x,y),(\mathrm{d}x',\mathrm{d}y')) = rP_{\text{MH}}(\mathrm{d}x'|x,y)\delta_y(\mathrm{d}y') + (1-r)\Pi_{Y|X}(\mathrm{d}y'|x)\delta_x(\mathrm{d}x')$$

and it is again well known that P_{RC} is reversible with respect to Π and hence has Π as its invariant distribution.

It will be convenient to consider marginalized versions of the DG chain, which we now define. The X-marginal DG chain is defined on X, and its Mtk is

$$P_{\text{XDG}}(x,\mathrm{d}x') = \int_{\mathsf{Y}} \Pi_{X|Y}(\mathrm{d}x'|y)\Pi_{Y|X}(\mathrm{d}y|x).$$

Similarly, the Y-marginal DG chain is defined on Y, and has Mtk

$$P_{\text{YDG}}(y,\mathrm{d}y') = \int_{\mathsf{X}} \Pi_{Y|X}(\mathrm{d}y'|x)\Pi_{X|Y}(\mathrm{d}x|y).$$

Note that P_{XDG} and P_{YDG} are reversible with respect to Π_X and Π_Y, respectively (Liu, Wong and Kong [37], Lemma 3.1). Moreover, it is well-known that the convergence properties of the marginal, P_{XDG} and P_{YDG}, chains are essentially those of the original DG chain (Robert [41], Roberts and Rosenthal [45]).

There also exists an X-marginal version of the DC sampler (but not a Y-marginal version) with Mtk given by

$$P_{\text{XDC}}(x,\mathrm{d}x') = \int_{\mathsf{Y}} P_{\text{MH}}(\mathrm{d}x'|x,y)\Pi_{Y|X}(\mathrm{d}y|x).$$

Jones, Roberts and Rosenthal's [31] Section 2.4 shows that P_{XDC} is reversible with respect to Π_X and enjoys the same qualitative rate of convergence in total variation norm as the parent DC sampler.

3.2. Operator norms

It is clear that P_{DG}, P_{RG}, P_{DC}, and P_{RC} can be regarded as operators defined on $L_0^2(\Pi)$. Among them, P_{RG} and P_{RC} are self-adjoint. It can be checked that P_{RG} is non-negative definite; see Liu, Wong and Kong [38], Lemma 3, and Rudolf and Ullrich [53], Section 3.2. Also, P_{XDG} and P_{XDC} are self-adjoint operators on $L_0^2(\Pi_X)$, while P_{YDG} is a self-adjoint operator on $L_0^2(\Pi_Y)$. Moreover, P_{XDG} and P_{YDG} are non-negative definite (Liu, Wong and Kong [37], Lemma 3.2).

Using Lemma 2.1 and the fact that RG and RC chains are reversible with respect to Π, we have

$$\rho(P_{\text{RG}}) = \|P_{\text{RG}}\|_\Pi \quad \text{and} \quad \rho(P_{\text{RC}}) = \|P_{\text{RC}}\|_\Pi.$$

Similar relations for the deterministic-scan samplers are given in the following lemma, whose proof is given in Appendix A.

Lemma 3.2. *For each positive integer n,*

$$\|P_{DG}^n\|_\Pi^{1/(n-1/2)} = \rho(P_{DG}) = \|P_{XDG}\|_{\Pi_X} = \|P_{YDG}\|_{\Pi_Y}$$

$$\|P_{DC}^n\|_\Pi^{1/(n-1)} \le \rho(P_{DC}) = \|P_{XDC}\|_{\Pi_X} \le \|P_{DC}^n\|_\Pi^{1/n}.$$

($\|P_{DC}^n\|_\Pi^{1/(n-1)}$ *is interpreted as* 0 *when* $n = 1$.)

The surprising exponent $1/(n-1/2)$ in the above lemma has also appeared in related results on alternating projections (Kayalar and Weinert [33], Theorem 2).

Applying Lemmas 2.1 and 3.2 we obtain the following.

Corollary 3.3. $\rho(P_{DG}) = \rho(P_{XDG}) = \rho(P_{YDG})$ *and* $\rho(P_{DC}) = \rho(P_{XDC})$.

For $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$, let

$$\gamma(g,h) = \int_{\mathsf{X}\times\mathsf{Y}} g(x)h(y)\Pi(\mathrm{d}x,\mathrm{d}y)$$

and

$$\bar{\gamma} = \sup\{\gamma(g,h) : g \in L_0^2(\Pi_X), \|g\|_{\Pi_X} = 1, h \in L_0^2(\Pi_Y), \|h\|_{\Pi_Y} = 1\}.$$

We say that $\bar{\gamma} \in [0,1]$ is the maximal correlation between X and Y. The following result can be found in Theorem 3.2 of Liu, Wong and Kong [37]; see also Liu, Wong and Kong [38] and Vidav [64].

Lemma 3.4.

$$\bar{\gamma}^2 = \|P_{XDG}\|_{\Pi_X} = \|P_{YDG}\|_{\Pi_Y}.$$

3.3. Total variation

We consider the connection between L^2 geometric ergodicity and the usual notion of geometric ergodicity defined through the total variation norm, denoted by $\|\cdot\|_{\text{TV}}$. For the four component-wise Markov chains considered here we can use results from Roberts and Tweedie [49] to show that these concepts are equivalent (see also Roberts and Rosenthal [43]). A proof is provided in Appendix B.

Proposition 3.5. *Let P denote the Mtk for any of the DG, RG, DC, and RC Markov chains. Suppose that the σ-algebras $\mathcal{F}_X$ and $\mathcal{F}_Y$ are countably generated, and P is φ-irreducible. Then P is L^2-geometrically ergodic if and only if it is Π-almost everywhere geometrically ergodic in the sense that, for Π-almost every (x, y), there exist $C(x, y)$ and $t < 1$ such that, for all n,*

$$\|P^n((x, y), \cdot) - \Pi(\cdot)\|_{TV} \le C(x, y)t^n.$$

Direct applications of Theorem 2 of Roberts and Tweedie [49] yield analogous results for the marginal chains defined by P_{XDG}, P_{YDG}, and P_{XDC}.

4. Quantitative relationship between $\rho(P_{\text{DG}})$ and $\rho(P_{\text{RG}})$

4.1. Main result

The proof of the following is given in Section 4.2.

Theorem 4.1.

$$\rho(P_{RG}) = \frac{1 + \sqrt{1 - 4r(1-r)[1 - \rho(P_{DG})]}}{2}.$$

We illustrate Theorem 4.1 in two examples.

Example 4.2. When Π is Gaussian, there are explicit formulas for $\rho(P_{\text{DG}})$ and $\rho(P_{\text{RG}})$ (Amit [2], Roberts and Sahu [47]). In particular, when Π is a bivariate Gaussian, and the correlation between X and Y is $\gamma \in [-1, 1]$, it is well-known that $\rho(P_{\text{DG}}) = \gamma^2$ (see, e.g., Diaconis, Khare and Saloff-Coste [9], Section 4.3). Meanwhile,

$$\rho(P_{\text{RG}}) = \frac{1 + \sqrt{1 - 4r(1-r)(1-\gamma^2)}}{2}$$

(Levine and Casella [35], page 193). This is in accordance with the general result in Theorem 4.1.

Example 4.3. When $\mathsf{X} \times \mathsf{Y}$ is a finite set, Π can be written in the form of a probability mass function (pmf), $\pi(\cdot, \cdot)$. For illustration, take $\mathsf{X} = \mathsf{Y} = \{1, 2, 3, 4, 5\}$, and generate the elements of $\pi(i, j)$, $(i, j) \in \mathsf{X} \times \mathsf{Y}$, via a Dirichlet distribution. The convergence rates of DG and RG samplers can then be calculated using the second largest eigenvalue in modulus of their transition matrices. We repeat this experiment 20 times for different values of selection probabilities. The results are displayed in Figure 2.

We now turn our attention to some of the implications of Theorem 4.1. Notice that, given the selection probability $r \in (0, 1)$, $\rho(P_{\text{RG}})$ and $\rho(P_{\text{DG}})$ are monotonic functions of each other. Indeed, in light of Lemmas 3.2 and 3.4, given any selection probability, the convergence rates of the two types of Gibbs chains are completely determined by the maximal correlation between X and Y. $\rho(P_{\text{DG}}) = \bar{\gamma}^2 = 0$ if and only if $\rho(P_{\text{RG}}) = \max\{r, 1-r\}$; $\rho(P_{\text{DG}}) = \bar{\gamma}^2 = 1$ if and only if $\rho(P_{\text{RG}}) = 1$; and when $\rho(P_{\text{DG}}) = \bar{\gamma}^2 \in (0, 1)$, $\rho(P_{\text{RG}}) \in (\max\{r, 1-r\}, 1)$ and $\rho(P_{\text{DG}}) < \rho(P_{\text{RG}})$.

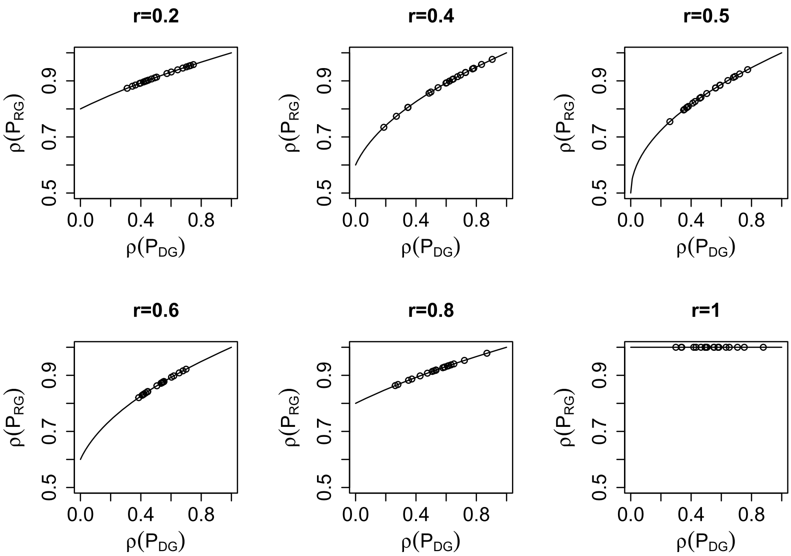

Figure 2. Relationship between $\rho(P_{\mathrm{DG}})$ and $\rho(P_{\mathrm{RG}})$ for discrete target distributions. In each subplot, 20 joint pmfs are randomly generated using Dirichlet distributions with random concentration parameters. Each circle corresponds to a joint pmf. The solid curves depict the relationship given in Theorem 4.1.

Let $k_* > 0$ be such that $\rho(P_{\mathrm{RG}})^{k_*} = \rho(P_{\mathrm{DG}})$ so that, roughly speaking, one iteration of the DG sampler is "worth" k_* iterations of the RG sampler in terms of convergence rate. By Young's inequality,

$$\begin{aligned}\rho(P_{\mathrm{RG}}) &= \frac{1+\sqrt{1-4r(1-r)+4r(1-r)\rho(P_{\mathrm{DG}})}}{2}\\ &\geq [1-4r(1-r)+4r(1-r)\rho(P_{\mathrm{DG}})]^{1/4}\\ &\geq \rho(P_{\mathrm{DG}})^{r(1-r)}.\end{aligned}$$

Therefore, $k_* \geq 1/[r(1-r)]$.

Let t_1 and t_2 be the time it takes to sample from $\Pi_{X|Y}$ and $\Pi_{Y|X}$, respectively. For simplicity, assume that they are constants. Suppose that, within unit time, one can run k_{D} iterations of the DG sampler, and k_{R} iterations of the RG sampler. Then

$$\frac{k_{\mathrm{R}}}{k_{\mathrm{D}}} = \frac{t_1+t_2}{rt_1+(1-r)t_2} \leq \frac{1}{\min\{r, 1-r\}} \leq \frac{1}{r(1-r)}.$$

Since $k_* \geq 1/[r(1-r)]$,

$$\rho(P_{\mathrm{DG}})^{k_{\mathrm{D}}} = \rho(P_{\mathrm{RG}})^{k_* k_{\mathrm{D}}} = \exp\left\{[\log\rho(P_{\mathrm{RG}})]k_* k_{\mathrm{R}}\frac{rt_1+(1-r)t_2}{t_1+t_2}\right\} \leq \rho(P_{\mathrm{RG}})^{k_{\mathrm{R}}}.$$

In this sense, the DG sampler converges faster than its random-scan counterpart.

Remark 4.4. It is natural to consider whether Theorem 4.1 can be extended to the case where there are more than two components. This proves to be challenging. Two-component deterministic-scan Gibbs samplers have reversible marginal chains and hence P_{XDG} and P_{YDG}, as linear operators, are self-adjoint. Consequentially, there are results like Lemma 3.2, which link the convergence rate $\rho(P_{\text{DG}})$ to the norms of P_{XDG} and P_{YDG}. However, deterministic-scan Gibbs samplers with more than two components do not possess reversible marginal chains. Therefore, techniques employed herein cannot be readily applied to compare the convergence rates of DG and RG samplers when the number of components exceed two.

4.2. Proof of Theorem 4.1

Remark 4.5. Before we begin the proof, we note that the result of Theorem 4.1 is related to the theory of two projections (Böttcher and Spitkovsky [5]). When $r = 1/2$, an alternative proof of Theorem 4.1 is available if we apply results on the norm of the sum of two projections (e.g., Duncan and Taylor [12], Theorem 7) along with Lemma 3.2.

By Lemmas 3.2 and 3.4, $\rho(P_{\text{DG}}) = \bar{\gamma}^2$, where $\bar{\gamma} \in [0, 1]$ is the maximal correlation between X and Y. To prove Theorem 4.1, we need to connect $\rho(P_{\text{RG}})$ to $\bar{\gamma}$. We begin with a preliminary result.

Lemma 4.6.

$$\rho(P_{RG}) \geq \max\{1 - r + r\bar{\gamma}^2, r + (1 - r)\bar{\gamma}^2\}.$$

Proof. Let $g \in L_0^2(\Pi_X)$ be such that $\|g\|_{\Pi_X} = 1$. Let f_g be such that $f_g(x, y) = g(x)$ for each $(x, y) \in \mathsf{X} \times \mathsf{Y}$ so that $f_g \in L_0^2(\Pi)$, and $\|f_g\|_\Pi = 1$. By Cauchy-Schwarz,

$$\begin{aligned} \|P_{\text{RG}}\|_\Pi &\geq \langle P_{\text{RG}} f_g, f_g\rangle_\Pi \\ &= r\int_{\mathsf{X}\times\mathsf{Y}}\left(\int_{\mathsf{X}} g(x')\Pi_{X|Y}(\mathrm{d}x'|y)\right) g(x)\Pi(\mathrm{d}x, \mathrm{d}y) + (1 - r)\langle g, g\rangle_{\Pi_X} \qquad (3) \\ &= r\langle P_{\text{XDG}} g, g\rangle_{\Pi_X} + 1 - r. \end{aligned}$$

Recall that P_{XDG} is non-negative definite. This implies that

$$\|P_{\text{XDG}}\|_{\Pi_X} = \sup\{\langle P_{\text{XDG}} g', g'\rangle_{\Pi_X} : g' \in L_0^2(\Pi_X), \|g'\|_{\Pi_X} = 1\}.$$

(See, e.g., Helmberg [20], §14 Corollary 5.1.) Taking the supremum with respect to g in (3) yields

$$\|P_{\text{RG}}\|_\Pi \geq 1 - r + r\|P_{\text{XDG}}\|_{\Pi_X} = 1 - r + r\bar{\gamma}^2, \qquad (4)$$

where the last equality follows from Lemma 3.4.

By an analogous argument,

$$\|P_{\text{RG}}\|_\Pi \geq r + (1 - r)\|P_{\text{YDG}}\|_{\Pi_Y} = r + (1 - r)\bar{\gamma}^2. \qquad (5)$$

Recall that $\rho(P_{\text{RG}}) = \|P_{\text{RG}}\|_\Pi$. The proof is completed by combining (4) and (5). □

Our proof of Theorem 4.1 hinges on the fact that, for each $f \in L_0^2(\Pi)$ and $(x, y) \in \mathsf{X} \times \mathsf{Y}$, $P_{\text{RG}} f(x, y)$ can be written in the form of $g(x) + h(y)$, where

$$g(x) = (1 - r) \int_{\mathsf{Y}} f(x, y') \Pi_{Y|X}(\mathrm{d}y'|x), \quad h(y) = r \int_{\mathsf{X}} f(x', y) \Pi_{X|Y}(\mathrm{d}x'|y).$$

As we will see, this allows us to restrict our attention to a well-behaved subspace of $L_0^2(\Pi)$ when studying the norm of P_{RG}.

For $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$, let $g \oplus h$ be the function on $\mathsf{X} \times \mathsf{Y}$ such that

$$(g \oplus h)(x, y) = g(x) + h(y)$$

for $(x, y) \in \mathsf{X} \times \mathsf{Y}$ (in a Π-almost everywhere sense). Let

$$H = \{g \oplus h : g \in L_0^2(\Pi_X),\ h \in L_0^2(\Pi_Y)\}.$$

Then H, equipped with the inner product $\langle \cdot, \cdot \rangle_\Pi$, is a subspace of $L_0^2(\Pi)$. For $g \oplus h \in H$,

$$\|g \oplus h\|_\Pi^2 = \|g\|_{\Pi_X}^2 + \|h\|_{\Pi_Y}^2 + 2\gamma(g, h),$$

where $\gamma(g, h)$ is defined in Section 3.1. It follows that

$$(1 - \bar{\gamma})(\|g\|_{\Pi_X}^2 + \|h\|_{\Pi_Y}^2) \le \|g \oplus h\|_\Pi^2 \le (1 + \bar{\gamma})(\|g\|_{\Pi_X}^2 + \|h\|_{\Pi_Y}^2). \tag{6}$$

When $\bar{\gamma} < 1$, $g \oplus h = 0$ if and only if $g = 0$ and $h = 0$. It follows that, whenever $\bar{\gamma} < 1$, for any $f \in H$, the decomposition $f = g \oplus h$ is unique.

To proceed, we present two technical results concerning H. Lemma 4.7 is proved in Appendix C, and Lemma 4.8 is a direct consequence of (6).

Lemma 4.7. *If $\bar{\gamma} < 1$, then H is a Hilbert space.*

Lemma 4.8. *Let $\bar{\gamma} < 1$ and suppose that $\{g_n\}_{n=1}^\infty$ and $\{h_n\}_{n=1}^\infty$ are sequences in $L_0^2(\Pi_X)$ and $L_0^2(\Pi_Y)$, respectively. For $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$, $\lim_{n\to\infty}(g_n \oplus h_n) = g \oplus h$ if and only if $\lim_{n\to\infty} g_n = g$, and $\lim_{n\to\infty} h_n = h$.*

It is easy to check that, for every $f \in L_0^2(\Pi)$, $P_{\text{RG}} f \in H$. Define $P_{\text{RG}}|_H$ to be P_{RG} restricted to H. The norm of $P_{\text{RG}}|_H$ is

$$\|P_{\text{RG}}|_H\|_\Pi = \sup_{f \in H,\, \|f\|_\Pi = 1} \|P_{\text{RG}} f\|_\Pi.$$

We then have the following lemma.

Lemma 4.9.

$$\|P_{RG}\|_\Pi = \|P_{RG}|_H\|_\Pi.$$

Proof. It is clear that

$$\|P_{\text{RG}}\|_\Pi \ge \|P_{\text{RG}}|_H\|_\Pi. \tag{7}$$

Because the range of P_{RG} is in H, for any $f \in L_0^2(\Pi)$ and positive integer n,

$$\|P_{\text{RG}}^n f\|_\Pi = \|P_{\text{RG}}|_H^{n-1} P_{\text{RG}} f\|_\Pi \leq \|P_{\text{RG}}|_H\|_\Pi^{n-1} \|f\|_\Pi .$$

Note that we have used the fact that $\|P_{\text{RG}}\|_\Pi \leq 1$. Since P_{RG} is self-adjoint, for each positive integer n, $\|P_{\text{RG}}^n\|_\Pi = \|P_{\text{RG}}\|_\Pi^n$. It follows that

$$\|P_{\text{RG}}\|_\Pi = \lim_{n\to\infty} \|P_{\text{RG}}^n\|_\Pi^{1/n} \leq \lim_{n\to\infty} \|P_{\text{RG}}|_H\|_\Pi^{(n-1)/n} = \|P_{\text{RG}}|_H\|_\Pi . \tag{8}$$

Combining (7) and (8) yields the desired result. □

We are now ready to prove the theorem.

Proof of Theorem 4.1. When $\bar{\gamma} = 1$, the theorem follows from Lemma 4.6 and the fact that $\rho(P_{\text{RG}}) = \|P_{\text{RG}}\|_\Pi \leq 1$. Assume that $\bar{\gamma} < 1$. We first show that

$$\rho(P_{\text{RG}}) \leq \frac{1 + \sqrt{1 - 4r(1-r)(1-\bar{\gamma}^2)}}{2} . \tag{9}$$

It follows from Lemma 4.9 that $\rho(P_{\text{RG}}) = \|P_{\text{RG}}|_H\|_\Pi$. Note that $P_{\text{RG}}|_H$ is a non-negative definite operator on H. By Lemma D.1 in Appendix D, $\rho(P_{\text{RG}})$ is an approximate eigenvalue of $P_{\text{RG}}|_H$, that is, there exists a sequence of functions $\{g_n \oplus h_n\}_{n=1}^\infty$ in H such that $\|g_n \oplus h_n\|_\Pi = 1$ for each n, and

$$\lim_{n\to\infty} [P_{\text{RG}}(g_n \oplus h_n) - \rho(P_{\text{RG}})(g_n \oplus h_n)] = 0 . \tag{10}$$

For every positive integer n,

$$P_{\text{RG}}(g_n \oplus h_n) = [(1-r)g_n + (1-r)Q_1 h_n] \oplus (r Q_2 g_n + r h_n) ,$$

where $Q_1 : L_0^2(\Pi_Y) \to L_0^2(\Pi_X)$ and $Q_2 : L_0^2(\Pi_X) \to L_0^2(\Pi_Y)$ are bounded linear transformations such that, for $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$,

$$(Q_1 h)(x) = \int_{\mathsf{Y}} h(y)\Pi_{Y|X}(\mathrm{d}y|x) , \quad (Q_2 g)(y) = \int_{\mathsf{X}} g(x)\Pi_{X|Y}(\mathrm{d}x|y) .$$

By Lemma 4.8, (10) implies that

$$\begin{aligned} \lim_{n\to\infty} \{[1 - r - \rho(P_{\text{RG}})]g_n + (1-r)Q_1 h_n\} = 0 , \\ \lim_{n\to\infty} \{[r - \rho(P_{\text{RG}})]h_n + r Q_2 g_n\} = 0 . \end{aligned} \tag{11}$$

Applying Q_1 to the second equality in (11) yields

$$\lim_{n\to\infty} \{[r - \rho(P_{\text{RG}})]Q_1 h_n + r P_{\text{XDG}} g_n\} = 0 .$$

Subtracting (a multiple of) this from (a multiple of) the first equality in (11) gives

$$\lim_{n\to\infty} \{[1 - r - \rho(P_{\text{RG}})][r - \rho(P_{\text{RG}})]g_n - r(1-r)P_{\text{XDG}} g_n\} = 0 . \tag{12}$$

Similarly, applying Q_2 to the first equality in (11) and subtracting it from the second equality in (11) yields

$$\lim_{n\to\infty}\{[1-r-\rho(P_{\text{RG}})][r-\rho(P_{\text{RG}})]h_n - r(1-r)P_{\text{YDG}}h_n\} = 0\,. \tag{13}$$

By Lemma 3.4, $\|P_{\text{XDG}}g_n\|_{\Pi_X} \le \bar{\gamma}^2\|g_n\|_{\Pi_X}$, $\|P_{\text{YDG}}h_n\|_{\Pi_Y} \le \bar{\gamma}^2\|h_n\|_{\Pi_Y}$. It follows from (12) and (13) that

$$\limsup_{n\to\infty}\{[1-r-\rho(P_{\text{RG}})][r-\rho(P_{\text{RG}})] - r(1-r)\bar{\gamma}^2\}\|g_n\|_{\Pi_X} \le 0\,,$$

$$\limsup_{n\to\infty}\{[1-r-\rho(P_{\text{RG}})][r-\rho(P_{\text{RG}})] - r(1-r)\bar{\gamma}^2\}\|h_n\|_{\Pi_Y} \le 0\,.$$

In particular,

$$\limsup_{n\to\infty}\{[1-r-\rho(P_{\text{RG}})][r-\rho(P_{\text{RG}})] - r(1-r)\bar{\gamma}^2\}(\|g_n\|_{\Pi_X} + \|h_n\|_{\Pi_Y}) \le 0\,. \tag{14}$$

By the triangle inequality,

$$\|g_n\|_{\Pi_X} + \|h_n\|_{\Pi_Y} = \|g_n \oplus 0\|_{\Pi} + \|0 \oplus h_n\|_{\Pi} \ge \|g_n \oplus h_n\|_{\Pi} = 1\,.$$

It then follows from (14) that

$$[\rho(P_{\text{RG}}) + r - 1][\rho(P_{\text{RG}}) - r] - r(1-r)\bar{\gamma}^2 \le 0\,. \tag{15}$$

This proves (9).

Next, we show that

$$\rho(P_{\text{RG}}) \ge \frac{1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2}\,. \tag{16}$$

This will complete the proof, since, by Lemmas 3.2 and 3.4, $\rho(P_{\text{DG}}) = \bar{\gamma}^2$. If $\bar{\gamma} = 0$, then (16) follows immediately from Lemma 4.6. Now assume $\bar{\gamma} \in (0, 1)$. Recall that $\rho(P_{\text{RG}}) = \|P_{\text{RG}}\|_{\Pi}$. It suffices to show that

$$\|P_{\text{RG}}\|_{\Pi} \ge \frac{1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2}. \tag{17}$$

Recall that P_{XDG} is non-negative definite, and $\|P_{\text{XDG}}\|_{\Pi_X} = \bar{\gamma}^2$. Hence, $\bar{\gamma}^2$ is an approximate eigenvalue of P_{XDG}. In other words, there exists a sequence of functions $\{\hat{g}_n\}_{n=1}^{\infty}$ in $L_0^2(\Pi_X)$ such that $\|\hat{g}_n\|_{\Pi_X} = 1$ for each n, and

$$\lim_{n\to\infty}(P_{\text{XDG}}\hat{g}_n - \bar{\gamma}^2\hat{g}_n) = 0\,. \tag{18}$$

Let

$$a = \frac{2r-1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2(1-r)\bar{\gamma}^2}\,.$$

Consider the sequence of functions $\{\hat{g}_n \oplus aQ_2\hat{g}_n\}_n$ in $H \subset L_0^2(\Pi)$. It is easy to show that, for each n,

$$\begin{aligned} P_{\text{RG}}(\hat{g}_n \oplus aQ_2\hat{g}_n) &= (1-r)[\hat{g}_n + aP_{\text{XDG}}\hat{g}_n] \oplus r(a+1)Q_2\hat{g}_n \\ &= (1-r)(1+a\bar{\gamma}^2)\hat{g}_n \oplus r(a+1)Q_2\hat{g}_n + (1-r)a(P_{\text{XDG}}\hat{g}_n - \bar{\gamma}^2\hat{g}_n) \oplus 0\,. \end{aligned} \tag{19}$$

It is straightforward to verify that

$$(1-r)(1+a\bar{\gamma}^2) = \frac{r(a+1)}{a} = \frac{1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2}.$$

Hence, (19) can be written as

$$P_{\text{RG}}(\hat{g}_n \oplus aQ_2\hat{g}_n) - \frac{1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2}(\hat{g}_n \oplus aQ_2\hat{g}_n) = (1-r)a(P_{\text{XDG}}\hat{g}_n - \bar{\gamma}^2\hat{g}_n) \oplus 0.$$

By (18), the right-hand-side goes to $0 \in H$ as $n \to \infty$. Moreover, by (6), $\|\hat{g}_n \oplus aQ_2\hat{g}_n\|_\Pi^2 \geq 1-\bar{\gamma} > 0$. It follows that

$$\|P_{\text{RG}}\|_\Pi \geq \limsup_{n\to\infty} \left\| P_{\text{RG}}\left(\frac{\hat{g}_n \oplus aQ_2\hat{g}_n}{\|\hat{g}_n \oplus aQ_2\hat{g}_n\|_\Pi}\right)\right\|_\Pi = \frac{1+\sqrt{1-4r(1-r)(1-\bar{\gamma}^2)}}{2},$$

and (17) holds. □

5. Qualitative relationships among convergence rates

It follows from Theorem 4.1 that the RG sampler is L^2 geometrically ergodic if and only if the associated DG sampler is too; see Roberts and Rosenthal's [43] Proposition 3.2 for what is essentially a proof of the "if" part. Our objective for Section 5.1 is to establish similar relations between other pairs of component-wise samplers introduced in Section 1, and eventually build Figure 1. In Section 5.2 we consider a more general setting where the Gibbs update from P_{DC} is replaced by a second Metropolis-Hastings update. Most of the results from Section 5.1 extend to the setting of Section 5.2.

5.1. Two-component samplers with Gibbs updates

First, by Proposition 2.3, if the RG or RC sampler is L^2 geometric ergodic for some selection probability, then it is L^2 geometrically ergodic for all selection probabilities. This allows us to treat the selection probabilities of the RG and RC sampler as arbitrary in what follows.

It is certainly not true that, in general, L^2 geometric ergodicity of the DG and RG samplers implies that of the DC or RC samplers. The following condition on the proposal density q will be useful.

Condition 5.1.

$$C = \sup_{(x',x,y)\in \mathsf{X}\times\mathsf{X}\times\mathsf{Y}} \frac{\pi_{X|Y}(x'|y)}{q(x'|x,y)} < \infty.$$

Condition 5.1 is analogous to a commonly-used condition for uniform ergodicity for full dimensional Metropolis-Hastings samplers (Liu [36], Mengersen and Tweedie [40], Roberts and Tweedie [48], Smith and Tierney [54]). Indeed, if P_{MH} is the Metropolis-Hastings Mtk which is reversible with respect to $\Pi_{X|Y}$ with proposal density q, then under condition 5.1, for each $y \in \mathsf{Y}$, $x \in \mathsf{X}$, and $A \in \mathcal{F}_X$,

$$P_{\text{MH}}(A|x,y) \geq \int_A \min\left\{\frac{q(x|x',y)}{\pi_{X|Y}(x|y)}, \frac{q(x'|x,y)}{\pi_{X|Y}(x'|y)}\right\}\pi_{X|Y}(x'|y)\,\mathrm{d}x' \geq \frac{1}{C}\Pi_{X|Y}(A|y). \tag{20}$$

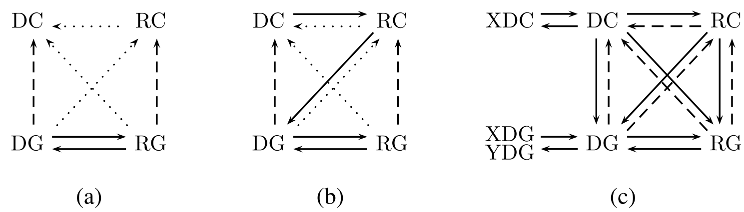

Figure 3. Building the relations among the convergence rates of component-wise samplers.

For a fixed $y \in \mathsf{Y}$, the Markov chain on X defined by P_{MH} has stationary distribution $\Pi_{X|Y}(\cdot|y)$ and (20) implies that this chain is uniformly ergodic. Moreover, condition 5.1 implies that if C can be calculated, then one can use an accept-reject sampler, at least in principle, to sample from $\Pi_{X|Y}$.

Condition 5.1 allows us to draw the dashed arrows between DG and DC and between RG and RC in Figure 3b.

Proposition 5.2. *Suppose condition* 5.1 *holds.*

1. *If* $\rho(P_{DG}) < 1$, *then* $\rho(P_{DC}) < 1$.
2. *If* $\rho(P_{RG}) < 1$, *then* $\rho(P_{RC}) < 1$.

Proof. Consider the first item. By Corollary 3.3, $\rho(P_{\text{XDG}}) = \rho(P_{\text{DG}}) < 1$ and hence by Lemma 3.2 $\|P_{\text{XDG}}\|_\Pi < 1$. Theorem 5 in Jones, Roberts and Rosenthal [31] establishes that, under condition 5.1, $P_{\text{XDC}}(x, A) \ge C^{-1} P_{\text{XDG}}(x, A)$ for $x \in \mathsf{X}$ and $A \in \mathcal{F}_X$. Hence by Lemma 2.2 we have that $\|P_{\text{XDC}}\|_\Pi < 1$. An appeal to Lemma 3.2 yields $\rho(P_{\text{DC}}) < 1$.

The second item follows from Lemma 2.2 since Theorem 6 in Jones, Roberts and Rosenthal [31] establishes that, under condition 5.1, $P_{\text{RC}}((x, y), A) \ge C^{-1} P_{\text{RG}}((x, y), A)$ for each (x, y) and measurable A. □

Next, observe that L^2 geometric ergodicity of the RC sampler does not necessarily imply that of the associated DC sampler. Indeed, a counter example can be constructed as follows. Let $\mathsf{X} = \mathsf{Y} = \{1, 2\}$, and suppose that Π is a uniform distribution on $\mathsf{X} \times \mathsf{Y}$. Then $\Pi_{Y|X}$ and $\Pi_{X|Y}$ are uniform distributions on Y and X, respectively. Let $q(\cdot|x, y)$ be defined with respect to the counting measure, and suppose that, for $y \in \{1, 2\}$, $q(2|1, y) = q(1|2, y) = 1$. In other words, $q(\cdot|x, y)$ always proposes a point in X that is different from x. The resulting RC chain is L^2-geometrically ergodic, but the associated DC chain is periodic, and it is easy to show that $\rho(P_{\text{DC}}) = 1$.

The relations that we have described so far can be summarized in Figure 3a. As in Figure 1, a solid arrow from one sampler to another means that L^2 geometric ergodicity of the former implies that of the latter, while a dashed arrow means that the relation does not hold in general, but does under condition 5.1. A dotted arrow from one sampler to another means that L^2 geometric ergodicity of the former does not imply that of the latter in general, and we have not yet addressed whether it does under condition 5.1.

Lemma 2.2 allows us to establish the following result, which shows that the RC sampler is L^2 geometrically ergodic whenever the DC sampler is. This allows us to draw a solid arrow from the DC sampler to the RC sampler in Figure 3b.

Proposition 5.3. *If* $\rho(P_{DC}) < 1$, *then* $\rho(P_{RC}) < 1$.

Proof. By Lemma 3.2, $\|P_{\mathrm{DC}}^2\|_\Pi \leq \rho(P_{\mathrm{DC}}) < 1$. For $(x, y) \in \mathsf{X} \times \mathsf{Y}$ and $A \in \mathcal{F}_X \times \mathcal{F}_Y$,

$$P_{\mathrm{RC}}^4((x, y), A) \geq r^2(1-r)^2 P_{\mathrm{DC}}^2((x, y), A)\,.$$

By Lemma 2.2, $\|P_{\mathrm{RC}}^4\|_\Pi < 1$. Since P_{RC} is self-adjoint, $\rho(P_{\mathrm{RC}}) = \|P_{\mathrm{RC}}\|_\Pi = \|P_{\mathrm{RC}}^4\|_\Pi^{1/4} < 1$. □

The final relation that we need to establish is given below. It allows us to draw a solid arrow from the RC sampler to the DG sampler.

Proposition 5.4. *If* $\rho(P_{RC}) < 1$, *then* $\rho(P_{DG}) < 1$.

Proof. Consider the contrapositive and recall that, by Lemmas 3.2 and 3.4, $\rho(P_{\mathrm{DG}}) = \bar{\gamma}^2$. Assume that $\bar{\gamma} = 1$. It suffices to show that $\rho(P_{\mathrm{RC}}) = 1$.

Let $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$ be such that $\|g\|_{\Pi_X} = \|h\|_{\Pi_Y} = 1$. Let $f_g \in L_0^2(\Pi)$ be such that $f_g(x, y) = g(x)$, and, $f_h \in L_0^2(\Pi)$, $f_h(x, y) = h(y)$. Recall that $\rho(P_{\mathrm{RC}}) = \|P_{\mathrm{RC}}\|_\Pi$. By Cauchy-Schwarz,

$$\begin{aligned}\rho(P_{\mathrm{RC}}) &\geq \langle P_{\mathrm{RC}} f_h, f_g\rangle_\Pi \\ &= r\int_{\mathsf{X}\times\mathsf{Y}} h(y)g(x)\Pi(\mathrm{d}x, \mathrm{d}y) + (1-r)\int_{\mathsf{X}\times\mathsf{Y}}\int_{\mathsf{Y}} h(y')\,\Pi_{Y|X}(\mathrm{d}y'|x)\,g(x)\,\Pi(\mathrm{d}x, \mathrm{d}y) \\ &= \gamma(g, h)\,.\end{aligned}$$

Taking the supremum with respect to g and h shows that $\rho(P_{\mathrm{RC}}) \geq \bar{\gamma} = 1$. □

Incorporating Propositions 5.3 and 5.4 in Figure 3a yields Figure 3b. From here, one can obtain Figure 1 by following the steps below:

1. When there is a path from one sampler to another consisting of only solid arrows, draw a (direct) solid arrow from the former to the latter, if there isn't one already. This allows us to draw solid arrows from DC to DG, from RC to RG, and and from DC to RG.
2. When there is a dotted arrow from one sampler to another, and there is a second path from the former to the latter consisting of dashed and possibly solid arrows, convert the dotted arrow to a dashed one. This allows us to convert all dotted arrows in Figure 3b to dashed ones.

For example, from Figure 3b we see that if $\rho(P_{\mathrm{DC}}) < 1$, then $\rho(P_{\mathrm{RC}}) < 1$ and if $\rho(P_{\mathrm{RC}}) < 1$, then $\rho(P_{\mathrm{DG}}) < 1$. Hence if $\rho(P_{\mathrm{DC}}) < 1$, then $\rho(P_{\mathrm{DG}}) < 1$ and we can obtain the solid arrow from DC to DG in Figure 1.

Finally, we can integrate Corollary 3.3 into Figure 1, and this yields Figure 3c.

5.2. CMH with two Metropolis-Hastings updates

Consider the setting where there are two components, both of which will be updated via Metropolis-Hastings. Let $q_1(\cdot|x, y)$ be a density on Y and $q_2(\cdot|x, y)$ a density on X. The deterministic-scan conditional Metropolis-Hastings algorithm with two Metropolis-Hastings updates is now described in Algorithm 5.

Denote the Mtk of the two Metropolis-Hastings steps by $P_1(\mathrm{d}y'|x, y)$ and $P_2(\mathrm{d}x'|x, y')$ so that the transition kernel, $\tilde{P}_{\mathrm{DC}}$, of the CMH algorithm is formed by composing the two Metropolis-Hastings

Algorithm 5 Deterministic-scan CMH with two Metropolis-Hastings updates

1: *Input:* Current value $(X_n, Y_n) = (x, y)$
2: Draw a random element W from $q_1(\cdot|x, y)$, and call the observed value w. With probability

$$a_1(w; x, y) = \min\left\{1, \frac{\pi_{Y|X}(w|x)q_1(y|x, w)}{\pi_{Y|X}(y|x)q_1(w|x, y)}\right\},$$

set $Y_{n+1} = w$; with probability $1 - a_1(w; x, y)$, set $Y_{n+1} = y$. Denote the observed value of Y_{n+1} by y'.
3: Draw a random element Z from $q_2(\cdot|x, y')$, and call the observed value z. With probability

$$a_2(z; x, y') = \min\left\{1, \frac{\pi_{X|Y}(z|y')q_2(x|z, y')}{\pi_{X|Y}(x|y')q_2(z|x, y')}\right\},$$

set $X_{n+1} = z$; with probability $1 - a_2(z; x, y')$, set $X_{n+1} = x$.
4: Set $n = n + 1$.

updates:

$$\tilde{P}_{\text{DC}}((x, y), (\mathrm{d}x', \mathrm{d}y')) = P_2(\mathrm{d}x'|x, y')P_1(\mathrm{d}y'|x, y).$$

We will refer to the algorithm as the $\widetilde{\text{DC}}$ sampler. Of course, there is a random-scan version with transition kernel denoted $\tilde{P}_{\text{RC}}$ which is formed by mixing the two Metropolis-Hastings updates:

$$\tilde{P}_{\text{RC}}((x, y), (\mathrm{d}x', \mathrm{d}y')) = rP_2(\mathrm{d}x'|x, y)\delta_y(\mathrm{d}y') + (1 - r)P_1(\mathrm{d}y'|x, y)\delta_x(\mathrm{d}x').$$

We will refer to the algorithm as the $\widetilde{\text{RC}}$ sampler. Notice that P_{DG} and P_{DC} are special cases of $\tilde{P}_{\text{DC}}$ where the proposal density is chosen to be the appropriate full conditional. Similarly, P_{RG} and P_{RC} are both special cases of $\tilde{P}_{\text{RC}}$. Thus there may be connections between the qualitative convergence rates of these Markov chains. By Proposition 2.3, we may treat all selection probabilities as arbitrary. We begin with an extension of Proposition 5.4.

Proposition 5.5. *If* $\rho(\tilde{P}_{RC}) < 1$, *then* $\rho(P_{DG}) < 1$.

Proof. The proof is essentially the same as that of Proposition 5.4 with $\tilde{P}_{\text{RC}}$ playing the role of P_{RC}. The only difference is that one needs to make use of the fact that $P_1(\mathrm{d}y'|x, y)$ leaves $\Pi_{Y|X}(\mathrm{d}y|x)$ invariant for each $x \in \mathsf{X}$: By Cauchy-Schwarz,

$$\begin{aligned}
\rho(\tilde{P}_{\text{RC}}) &\geq \langle \tilde{P}_{\text{RC}} f_h, f_g \rangle_\Pi \\
&= r\int_{\mathsf{X}\times\mathsf{Y}} h(y)g(x)\Pi(\mathrm{d}x, \mathrm{d}y) + (1 - r)\int_{\mathsf{X}\times\mathsf{Y}}\int_{\mathsf{Y}} h(y')\, P_1(\mathrm{d}y'|x, y)\, g(x)\, \Pi_{Y|X}(\mathrm{d}y|x)\Pi_X(\mathrm{d}x) \\
&= r\int_{\mathsf{X}\times\mathsf{Y}} h(y)g(x)\Pi(\mathrm{d}x, \mathrm{d}y) + (1 - r)\int_{\mathsf{X}\times\mathsf{Y}} h(y')\, \Pi_{Y|X}(\mathrm{d}y'|x)\, g(x)\, \Pi_X(\mathrm{d}x) \\
&= \gamma(g, h).
\end{aligned}$$

□

We will have need of the following condition on the proposal density q_1 at several points. Notice the analogy to condition 5.1.

Condition 5.6.

$$C_1 = \sup_{(x,y,y')\in \mathsf{X}\times\mathsf{Y}\times\mathsf{Y}} \frac{\pi_{Y|X}(y'|x)}{q_1(y'|x,y)} < \infty.$$

Proposition 5.7. *Suppose that condition 5.6 holds. If $\rho(P_{DC}) < 1$, then $\rho(\tilde{P}_{DC}) < 1$.*

Proof. Observe that by condition 5.6 we have

$$q_1(y'|x,y)a_1(y';x,y) = \pi_{Y|X}(y'|x)\min\left\{\frac{q_1(y'|x,y)}{\pi_{Y|X}(y'|x)}, \frac{q_1(y|x,y')}{\pi_{Y|X}(y|x)}\right\} \geq \frac{1}{C_1}\pi_{Y|X}(y'|x),$$

so that

$$P_1(dy'|x,y) \geq \frac{1}{C_1}\Pi_{Y|X}(dy'|x),$$

and hence,

$$\tilde{P}_{\mathrm{DC}}((x,y),(\mathrm{d}x',\mathrm{d}y')) \geq \frac{1}{C_1}P_{\mathrm{DC}}((x,y),(\mathrm{d}x',\mathrm{d}y')).$$

It follows that, for $(x,y)\in\mathsf{X}\times\mathsf{Y}$ and $A\in\mathcal{F}_X\times\mathcal{F}_Y$,

$$\tilde{P}^2_{\mathrm{DC}}((x,y),A) \geq \frac{1}{C_1^2}P^2_{\mathrm{DC}}((x,y),A).$$

By Lemma 3.2, $\|P^2_{\mathrm{DC}}\|_\Pi \leq \rho(P_{\mathrm{DC}}) < 1$. Then Lemma 2.2 implies that $\|\tilde{P}^2_{\mathrm{DC}}\|_\Pi < 1$. For $\mu \in L^2_*(\Pi)$ and positive integer n,

$$\|\mu\tilde{P}^n_{\mathrm{DC}} - \Pi\|_\Pi = \|(\mu-\Pi)\tilde{P}^n_{\mathrm{DC}}\|_\Pi = \sup_{f\in L^2_0(\Pi),\|f\|_\Pi=1}\left\langle \frac{\mathrm{d}\mu}{\mathrm{d}\Pi} - 1, \tilde{P}^n_{\mathrm{DC}}f\right\rangle_\Pi.$$

Using Cauchy-Schwarz and treating even and odd n separately, we see that

$$\sup_{f\in L^2_0(\Pi),\|f\|_\Pi=1}\left\langle \frac{\mathrm{d}\mu}{\mathrm{d}\Pi} - 1, \tilde{P}^n_{\mathrm{DC}}f\right\rangle_\Pi \leq \|\mu-\Pi\|_\Pi\|\tilde{P}^2_{\mathrm{DC}}\|_\Pi^{\lfloor n/2\rfloor} \leq \|\mu-\Pi\|_\Pi\|\tilde{P}^2_{\mathrm{DC}}\|_\Pi^{(n-1)/2},$$

where $\lfloor n/2\rfloor$ is the largest integer that does not exceed $n/2$. Therefore, $\rho(\tilde{P}_{\mathrm{DC}}) \leq \|\tilde{P}^2_{\mathrm{DC}}\|_\Pi^{1/2} < 1$. □

Proposition 5.8. *Suppose that condition 5.6 holds. If $\rho(P_{RC}) < 1$, then $\rho(\tilde{P}_{RC}) < 1$.*

Proof. Since both P_{RC} and $\tilde{P}_{\mathrm{RC}}$ are reversible, $\rho(P_{\mathrm{RC}}) = \|P_{\mathrm{RC}}\|_\Pi$ and $\rho(\tilde{P}_{\mathrm{RC}}) = \|\tilde{P}_{\mathrm{RC}}\|_\Pi$. Recall that under condition 5.6,

$$P_1(dy'|x,y) \geq \frac{1}{C_1}\Pi_{Y|X}(dy'|x),$$

and hence,

$$\tilde{P}_{\mathrm{RC}}((x,y),(\mathrm{d}x',\mathrm{d}y')) \geq rP_2(\mathrm{d}x'|x,y)\delta_y(\mathrm{d}y') + \frac{1-r}{C_1}\Pi_{Y|X}(dy'|x)\delta_x(\mathrm{d}x')$$

$$\geq \frac{1}{C_1} P_{\mathrm{RC}}((x, y), (\mathrm{d}x', \mathrm{d}y')).$$

Note that, without loss of generality, we have assumed that P_{RC} and $\tilde{P}_{\mathrm{RC}}$ have the same selection probability. The desired result now follows from Lemma 2.2. □

One can combine the results above with those in the previous subsection to obtain other relations. For example, combining Proposition 5.8 with Proposition 5.3 gives the following result.

Corollary 5.9. *Suppose that condition 5.6 holds. If $\rho(P_{DC}) < 1$, then $\rho(\tilde{P}_{RC}) < 1$.*

To make progress on developing further qualitative convergence relationships, we will need to include condition 5.1 so that we can appeal to the results of the previous section. Notice that the proposal density q_2 from Algorithm 5 corresponds to the proposal density q from Algorithm 3 so that condition 5.1 can be interpreted as a condition on q_2.

Proposition 5.10. *Suppose that conditions 5.1 and 5.6 hold. Then*

1. *if $\rho(P_{DG}) < 1$, then $\rho(\tilde{P}_{DC}) < 1$ and $\rho(\tilde{P}_{RC}) < 1$;*
2. *if $\rho(P_{RG}) < 1$, then $\rho(\tilde{P}_{RC}) < 1$ and $\rho(\tilde{P}_{DC}) < 1$; and*
3. *if $\rho(\tilde{P}_{RC}) < 1$, then $\rho(\tilde{P}_{DC}) < 1$.*

Proof. We consider only the first item as the others are similar. From Figure 3c we have that, under condition 5.1, if $\rho(P_{\mathrm{DG}}) < 1$, then $\rho(P_{\mathrm{DC}}) < 1$ and $\rho(P_{\mathrm{RC}}) < 1$. Combining this with Propositions 5.7 and 5.8 yields the claim. □

Remark 5.11. The relations in Proposition 5.10 do not necessarily hold without conditions such as 5.1 and 5.6. For instance, in the previous subsection we have shown that $\rho(P_{\mathrm{RC}}) < 1$ does not imply $\rho(P_{\mathrm{DC}}) < 1$ in general. Since P_{RC} and P_{DC} are respectively special cases of $\tilde{P}_{\mathrm{RC}}$ and $\tilde{P}_{\mathrm{DC}}$, $\rho(\tilde{P}_{\mathrm{RC}}) < 1$ does not imply $\rho(\tilde{P}_{\mathrm{DC}}) < 1$ in general.

We depict the known qualitative convergence relations among DG, RG, $\widetilde{\mathrm{DC}}$, and $\widetilde{\mathrm{RC}}$ in Figure 4. A dashed arrow with double arrowheads from one to another means that L^2 geometric ergodicity of the former implies that of the latter under conditions 5.1 and 5.6, but not in general. Figure 4 illustrates that the complexity of $\widetilde{\mathrm{DC}}$ (in particular, the lack of a reversible marginal Markov chain) means that it is an open question as to whether the L^2 geometric ergodicity of $\widetilde{\mathrm{DC}}$ implies the L^2 geometric ergodicity of any of the rest.

Finally note that combining Figures 3c and 4 does not characterize all known qualitative convergence relationships. For example, combining the results in the figures would suggest that under conditions 5.1 and 5.6 if $\rho(P_{\mathrm{DC}}) < 1$, then $\rho(\tilde{P}_{\mathrm{DC}}) < 1$, but Proposition 5.7 shows that only condition 5.6 is required.

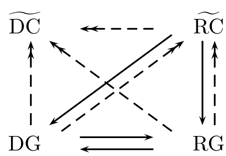

Figure 4. Known qualitative convergence relationships among the Gibbs samplers and their CMH variants with two Metropolis-Hastings updates.

6. Final remarks

We have focused on convergence relationships between deterministic-scan and random-scan MCMC algorithms when there are two component-wise updates. At the heart of these relationships is the explicit quantitative relationship developed between Gibbs samplers in Theorem 4.1. This result is intuitively appealing since a random-scan Gibbs sampler may update the same component consecutively and thus one might expect convergence to be slower than the deterministic-scan version. When there are more than two components it is intuitively less obvious that the random-scan Gibbs sampler will converge substantially more slowly than the deterministic-scan version since it is much less likely to update the same component consecutively. Indeed, Roberts and Rosenthal [46] provide some examples where the relationship between the convergence rates becomes more complicated when the number of components is large. However, as we explained in Remark 4.4 there are technical hurdles to investigating this rigorously. As we saw in Section 5.2, the situation is even more complicated, even in the two-component setting, when considering deterministic-scan and random-scan versions of CMH Markov chains. There is ample room for future work along these lines.

Appendix A: Proof of Lemma 3.2

We will prove

$$\|P_{\text{DG}}^n\|_{\Pi}^{1/(n-1/2)} = \rho(P_{\text{DG}}) = \|P_{\text{XDG}}\|_{\Pi_X} = \|P_{\text{YDG}}\|_{\Pi_Y}\,.$$

The proof for the other equation is similar.

(i) $\|P_{\text{XDG}}\|_{\Pi_X} = \|P_{\text{YDG}}\|_{\Pi_Y}$. This is given in Liu, Wong and Kong's [37] Theorem 3.2.

(ii) $\|P_{\text{DG}}^n\|_{\Pi}^{1/(n-1/2)} = \|P_{\text{XDG}}\|_{\Pi_X}$. Firstly, since P_{XDG} is self-adjoint, for each positive integer n, $\|P_{\text{XDG}}^n\|_{\Pi_X} = \|P_{\text{XDG}}\|_{\Pi_X}^n$ and similarly for P_{YDG}.

We begin by showing that $\|P_{\text{DG}}^n\|_{\Pi}^{1/(n-1/2)} \le \|P_{\text{XDG}}\|_{\Pi_X}$. Let $f \in L_0^2(\Pi)$ be such that $\|f\|_{\Pi} = 1$, and let

$$h_f(y) = \int_{\mathsf{X}} f(x,y)\Pi_{X|Y}(\mathrm{d}x|y)\,,\ y \in \mathsf{Y}\,,$$

$$g_f(x) = \int_{\mathsf{Y}} h_f(y)\Pi_{Y|X}(\mathrm{d}y|x)\,,\ x \in \mathsf{X}\,.$$

Then $h_f \in L_0^2(\Pi_Y)$, and $g_f \in L_0^2(\Pi_X)$. Note that

$$\begin{aligned}\langle g_f, g_f\rangle_{\Pi_X} &= \int_{\mathsf{X}}\int_{\mathsf{Y}} h_f(y)\Pi_{Y|X}(\mathrm{d}y|x)\int_{\mathsf{Y}} h_f(y')\Pi_{Y|X}(\mathrm{d}y'|x)\,\Pi_X(\mathrm{d}x)\\ &= \int_{\mathsf{Y}\times\mathsf{X}\times\mathsf{Y}} h_f(y)\,h_f(y')\Pi_{Y|X}(\mathrm{d}y'|x)\Pi_{X|Y}(\mathrm{d}x|y)\,\Pi_Y(\mathrm{d}y)\\ &= \langle h_f, P_{\text{YDG}}h_f\rangle_{\Pi_Y}\,.\end{aligned}$$

Moreover, by the Cauchy-Schwarz inequality, $\|h_f\|_{\Pi_Y} \le 1$. It follows that

$$\|g_f\|_{\Pi_X}^2 = \langle h_f, P_{\text{YDG}}h_f\rangle_{\Pi_Y} \le \|P_{\text{YDG}}\|_{\Pi_Y} = \|P_{\text{XDG}}\|_{\Pi_X}\,.$$

It is easy to verify that, for each positive integer n and $(x, y) \in \mathsf{X}\times\mathsf{Y}$, $P_{\text{DG}}^n f(x, y) = P_{\text{XDG}}^{n-1} g_f(x)$. Therefore,

$$\|P_{\text{DG}}^n f\|_\Pi = \|P_{\text{XDG}}^{n-1} g_f\|_{\Pi_X} \le \|P_{\text{XDG}}\|_{\Pi_X}^{n-1} \|g_f\|_{\Pi_X} \le \|P_{\text{XDG}}\|_{\Pi_X}^{n-1/2}.$$

Taking the supremum with respect to f yields the desired inequality.

We now show that $\|P_{\text{YDG}}\|_{\Pi_Y} \le \|P_{\text{DG}}^n\|_\Pi^{1/(n-1/2)}$ and it will follow immediately that $\|P_{\text{XDG}}\|_{\Pi_X} \le \|P_{\text{DG}}^n\|_\Pi^{1/(n-1/2)}$. Let $h \in L_0^2(\Pi_Y)$ be such that $\|h\|_{\Pi_Y} = 1$. Let $f_h \in L_0^2(\Pi)$ be such that $f_h(x, y) = h(y)$ for $(x, y) \in \mathsf{X}\times\mathsf{Y}$. Then $\|f_h\|_\Pi = 1$. Lastly, let $Q_1 h \in L_0^2(\Pi_X)$ be such that $(Q_1 h)(x) = \int_{\mathsf{Y}} h(y)\Pi_{Y|X}(\mathrm{d}y|x)$. Note that, for $(x, y) \in \mathsf{X}\times\mathsf{Y}$, $(Q_1 h)(x) = (P_{\text{DG}} f_h)(x, y)$. A careful calculation shows that

$$\begin{aligned}
\langle h, P_{\text{YDG}}^{2n-1} h\rangle_{\Pi_Y} &= \int_{\mathsf{Y}} h(y) \int_{\mathsf{X}} (P_{\text{XDG}}^{2n-2} Q_1 h)(x)\, \Pi_{X|Y}(\mathrm{d}x|y)\, \Pi_Y(\mathrm{d}y) \\
&= \int_{\mathsf{Y}\times\mathsf{X}} h(y)\, (P_{\text{XDG}}^{2n-2} Q_1 h)(x)\, \Pi_{Y|X}(\mathrm{d}y|x)\, \Pi_X(\mathrm{d}x) \\
&= \langle P_{\text{XDG}}^{n-1} Q_1 h, P_{\text{XDG}}^{n-1} Q_1 h\rangle_{\Pi_X} \\
&= \langle P_{\text{DG}}^n f_h, P_{\text{DG}}^n f_h\rangle_\Pi \\
&\le \|P_{\text{DG}}^n\|_\Pi^2.
\end{aligned}$$

Since P_{YDG}^{2n-1} is non-negative definite,

$$\|P_{\text{YDG}}\|_{\Pi_Y}^{2n-1} = \|P_{\text{YDG}}^{2n-1}\|_{\Pi_Y} = \sup\{\langle P_{\text{YDG}}^{2n-1} h', h'\rangle_{\Pi_Y} : h' \in L_0^2(\Pi_Y), \|h'\|_{\Pi_Y} = 1\}.$$

This shows that $\|P_{\text{YDG}}\|_{\Pi_Y}^{2n-1} \le \|P_{\text{DG}}^n\|_\Pi^2$.

(iii) $\rho(P_{\text{DG}}) = \|P_{\text{XDG}}\|_{\Pi_X}$. By Lemma 2.1, $\|P_{\text{XDG}}\|_{\Pi_X} = \rho(P_{\text{XDG}})$, the L^2 convergence rate of the X-marginal DG chain.

We now show that $\rho(P_{\text{XDG}}) \le \rho(P_{\text{DG}})$. Let $g \in L_0^2(\Pi_X)$ be such that $\|g\|_{\Pi_X} = 1$. Let $f_g \in L_0^2(\Pi)$ be such that $f_g(x, y) = g(x)$. Then $\|f_g\|_\Pi = 1$. For any $\mu \in L_*^2(\Pi_X)$ and positive integer n,

$$|\mu P_{\text{XDG}}^n g - \Pi_X g| = |\tilde{\mu} P_{\text{DG}}^n f_g - \Pi f_g| \le \|\tilde{\mu} P_{\text{DG}}^n - \Pi\|_\Pi,$$

where $\tilde{\mu}$ is any measure in $L_*^2(\Pi)$ such that $\int_{\mathsf{Y}} \tilde{\mu}(\cdot, \mathrm{d}y) = \mu(\cdot)$. Taking the supremum with respect to g shows that

$$\|\mu P_{\text{XDG}}^n - \Pi_X\|_{\Pi_X} \le \|\tilde{\mu} P_{\text{DG}}^n - \Pi\|_\Pi.$$

This implies that $\rho(P_{\text{XDG}}) \le \rho(P_{\text{DG}})$.

Finally, we show that $\rho(P_{\text{DG}}) \le \rho(P_{\text{XDG}})$. Let $\tilde{\mu} \in L_*^2(\Pi)$, and define $f \in L_0^2(\Pi)$ and $g_f \in L_0^2(\Pi_X)$ as in (ii). Then, for a positive integer n,

$$|\tilde{\mu} P_{\text{DG}}^n f - \Pi f| = |\mu P_{\text{XDG}}^{n-1} g_f - \Pi_X g_f| \le \|\mu P_{\text{XDG}}^{n-1} - \Pi_X\|_{\Pi_X},$$

where $\mu(\cdot) = \int_{\mathsf{Y}} \tilde{\mu}(\cdot, \mathrm{d}y)$. Taking the supremum with respect to f shows that

$$\|\tilde{\mu} P_{\text{DG}}^n - \Pi\|_\Pi \le \|\mu P_{\text{XDG}}^{n-1} - \Pi_X\|_{\Pi_X},$$

which implies that $\rho(P_{\text{DG}}) \le \rho(P_{\text{XDG}})$.

Appendix B: Proof of Proposition 3.5

We will prove the result for P_{DC} and P_{RC}. The proofs for P_{DG} and P_{RG} are similar. We will make use of results in Roberts and Tweedie [49]. These results require $\mathcal{F}_X$ and $\mathcal{F}_X \times \mathcal{F}_Y$ to be countably generated. We have assumed that $\mathcal{F}_X$ and $\mathcal{F}_Y$ are countably generated. This implies that $\mathcal{F}_X \times \mathcal{F}_Y$ is also countably generated. Indeed, if $\mathcal{F}_X$ is generated by $\{A_i\}_{i=1}^\infty$ and $\mathcal{F}_Y$ is generated by $\{B_j\}_{j=1}^\infty$, then $\mathcal{F}_X \times \mathcal{F}_Y$ can be generated by sets of the forms $A_i \times \mathsf{Y}$ and $\mathsf{X} \times B_j$.

Consider P_{RC}. Then the claim follows immediately due to its reversibility with respect to Π (Roberts and Tweedie [49], Theorem 2).

Now suppose P_{DC} is L^2 geometrically ergodic. Then it is Π-a.e. geometrically ergodic (Roberts and Tweedie [49], Theorem 1). Conversely, suppose that the P_{DC} is Π-a.e. geometrically ergodic. This implies P_{XDC} is Π_X-a.e. geometrically ergodic. It is also straightforward to check that P_{XDC} is φ^*-irreducible, with $\varphi^*(\cdot) = \int_{\mathsf{Y}} \varphi(\cdot, \mathrm{d}y)$. Since P_{XDC} is reversible with respect to Π_X, it is also L^2 geometrically ergodic. By Corollary 3.3, P_{DC} must be L^2 geometrically ergodic as well.

Appendix C: Proof of Lemma 4.7

It suffices to show that H is closed (see, e.g., Helmberg [20], §6). Consider a sequence of functions in H, $\{g_n \oplus h_n\}_{n=1}^\infty$, such that

$$\lim_{n\to\infty} (g_n \oplus h_n) = f \in L_0^2(\Pi).$$

The sequence $\{g_n \oplus h_n\}$ is Cauchy, that is,

$$\lim_{n\to\infty} \sup_{m\geq n} \|g_n \oplus h_n - (g_m \oplus h_m)\|_\Pi = 0.$$

By (6),

$$\|g_n \oplus h_n - (g_m \oplus h_m)\|_\Pi^2 \geq (1-\bar{\gamma})(\|g_n - g_m\|_{\Pi_X}^2 + \|h_n - h_m\|_{\Pi_Y}^2).$$

Since $\bar{\gamma} < 1$, $\{g_n\}$ and $\{h_n\}$ are Cauchy as well. By the completeness of $L_0^2(\Pi_X)$ and $L_0^2(\Pi_Y)$, there exist $g \in L_0^2(\Pi_X)$ and $h \in L_0^2(\Pi_Y)$ such that

$$\lim_{n\to\infty} g_n = g, \quad \lim_{n\to\infty} h_n = h.$$

Again by (6),

$$\|g_n \oplus h_n - (g \oplus h)\|_\Pi^2 \leq (1+\bar{\gamma})(\|g_n - g\|_{\Pi_X}^2 + \|h_n - h\|_{\Pi_Y}^2).$$

This implies that

$$\lim_{n\to\infty} (g_n \oplus h_n) = g \oplus h.$$

Hence, $f = g \oplus h \in H$, meaning that H is closed.

Appendix D: A lemma concerning Theorem 4.1

The following lemma is the result of several elementary facts in functional analysis. See, e.g., Helmberg [20], §23, 24.

Lemma D.1. *Let H' be a real or complex Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Let P be a bounded non-negative definite operator on H'. Then $\|P\|$ is an approximate eigenvalue of P, i.e., there exists a sequence $\{f_n\}_{n=1}^{\infty}$ in H' such that $\|f_n\| = 1$ for each n, and $\lim_{n\to\infty} \|Pf_n - \|P\| f_n\| = 0$.*

Proof. Since P is non-negative definite,

$$\|P\| = \sup_{f \in H', \|f\|=1} \langle Pf, f \rangle .$$

It follows that there exists a sequence $\{f_n\}_n$ in H' such that $\|f_n\| = 1$ for each n, and $\lim_{n\to\infty} \langle Pf_n, f_n \rangle = \|P\|$. Note that

$$\langle Pf_n, f_n \rangle \le \|Pf_n\| \le \|P\| .$$

This implies that $\|Pf_n\| \to \|P\|$ as $n \to \infty$. It follows that

$$\lim_{n\to\infty} \|Pf_n - \|P\| f_n\|^2 = \lim_{n\to\infty} (\|Pf_n\|^2 + \|P\|^2 - 2\|P\| \langle Pf_n, f_n \rangle) = 0 . \qquad \square$$

Acknowledgment

We thank the Editor, the Associate Editor, and two anonymous reviewers for their constructive feedback.

Funding

The first author was partially supported by the National Science Foundation.
The second author was partially supported by the National Science Foundation.

References

[1] Amit, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–99. MR1128938 https://doi.org/10.1016/0047-259X(91)90033-X
[2] Amit, Y. (1996). Convergence properties of the Gibbs sampler for perturbations of Gaussians. *Ann. Statist.* **24** 122–140. MR1389883 https://doi.org/10.1214/aos/1033066202
[3] Amit, Y. and Grenander, U. (1991). Comparing sweep strategies for stochastic relaxation. *J. Multivariate Anal.* **37** 197–222. MR1114473 https://doi.org/10.1016/0047-259X(91)90080-L
[4] Andrieu, C. (2016). On random- and systematic-scan samplers. *Biometrika* **103** 719–726. MR3551794 https://doi.org/10.1093/biomet/asw019
[5] Böttcher, A. and Spitkovsky, I.M. (2010). A gentle guide to the basics of two projections theory. *Linear Algebra Appl.* **432** 1412–1459. MR2580440 https://doi.org/10.1016/j.laa.2009.11.002
[6] Brooks, S., Gelman, A., Jones, G.L. and Meng, X.-L., eds. (2011). *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Boca Raton, FL: CRC Press. MR2742422 https://doi.org/10.1201/b10905
[7] Chan, K.S. and Geyer, C.J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1747–1758.
[8] Dai, N. and Jones, G.L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *J. Multivariate Anal.* **159** 184–199. MR3668555 https://doi.org/10.1016/j.jmva.2017.05.009

[9] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist*. *Sci*. **23** 151–178. With comments and a rejoinder by the authors. MR2446500 https://doi.org/10.1214/07-STS252

[10] Doss, H. and Hobert, J.P. (2010). Estimation of Bayes factors in a class of hierarchical random effects models using a geometrically ergodic MCMC algorithm. *J*. *Comput*. *Graph*. *Statist*. **19** 295–312. With supplementary material available online. MR2675092 https://doi.org/10.1198/jcgs.2010.09182

[11] Doss, C.R., Flegal, J.M., Jones, G.L. and Neath, R.C. (2014). Markov chain Monte Carlo estimation of quantiles. *Electron*. *J*. *Stat*. **8** 2448–2478. MR3285872 https://doi.org/10.1214/14-EJS957

[12] Duncan, J. and Taylor, P.J. (1976). Norm inequalities for C^*-algebras. *Proc*. *Roy*. *Soc*. *Edinburgh Sect*. *A* **75** 119–129. MR0454647 https://doi.org/10.1017/S0308210500017832

[13] Ekvall, K.O. and Jones, G.L. (2021). Convergence analysis of a collapsed Gibbs sampler for Bayesian vector autoregressions. *Electron*. *J*. *Stat*. **15** 691–721. MR4202497 https://doi.org/10.1214/21-EJS1800

[14] Fishman, G.S. (1996). Coordinate selection rules for Gibbs sampling. *Ann*. *Appl*. *Probab*. **6** 444–465. MR1398053 https://doi.org/10.1214/aoap/1034968139

[15] Flegal, J.M., Haran, M. and Jones, G.L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist*. *Sci*. **23** 250–260. MR2516823 https://doi.org/10.1214/08-STS257

[16] Flegal, J.M. and Jones, G.L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann*. *Statist*. **38** 1034–1070. MR2604704 https://doi.org/10.1214/09-AOS735

[17] Fort, G., Moulines, E., Roberts, G.O. and Rosenthal, J.S. (2003). On the geometric ergodicity of hybrid samplers. *J*. *Appl*. *Probab*. **40** 123–146. MR1953771 https://doi.org/10.1239/jap/1044476831

[18] Geyer, C.J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist*. *Sci*. **7** 473–511.

[19] Greenwood, P.E., McKeague, I.W. and Wefelmeyer, W. (1998). Information bounds for Gibbs samplers. *Ann*. *Statist*. **26** 2128–2156. MR1700224 https://doi.org/10.1214/aos/1024691464

[20] Helmberg, G. (2014). *Introduction to Spectral Theory in Hilbert Space*. Elsevier. MR0243367

[21] Herbei, R. and McKeague, I.W. (2009). Hybrid samplers for ill-posed inverse problems. *Scand*. *J*. *Stat*. **36** 839–853. MR2573311 https://doi.org/10.1111/j.1467-9469.2009.00649.x

[22] Hobert, J.P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.). *Chapman & Hall/CRC Handb*. *Mod*. *Stat*. *Methods* 253–293. Boca Raton, FL: CRC Press. MR2858452

[23] Hobert, J.P. and Geyer, C.J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J*. *Multivariate Anal*. **67** 414–430. MR1659196 https://doi.org/10.1006/jmva.1998.1778

[24] Hobert, J.P., Jones, G.L., Presnell, B. and Rosenthal, J.S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* **89** 731–743. MR1946508 https://doi.org/10.1093/biomet/89.4.731

[25] Johnson, A.A. and Jones, G.L. (2008). Comment: Gibbs sampling, exponential families, and orthogonal polynomials. *Statist*. *Sci*. **23** 183–186.

[26] Johnson, A.A. and Jones, G.L. (2015). Geometric ergodicity of random scan Gibbs samplers for hierarchical one-way random effects models. *J*. *Multivariate Anal*. **140** 325–342. MR3372572 https://doi.org/10.1016/j.jmva.2015.06.002

[27] 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. MR3135537 https://doi.org/10.1214/13-STS423

[28] Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab*. *Surv*. **1** 299–320. MR2068475 https://doi.org/10.1214/154957804100000051

[29] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist*. *Sci*. **16** 312–334. MR1888447 https://doi.org/10.1214/ss/1015346317

[30] Jones, G.L. and Hobert, J.P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann*. *Statist*. **32** 784–817. MR2060178 https://doi.org/10.1214/009053604000000184

[31] Jones, G.L., Roberts, G.O. and Rosenthal, J.S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Adv*. *in Appl*. *Probab*. **46** 422–445. MR3215540 https://doi.org/10.1239/aap/1401369701

[32] Jones, G.L., Haran, M., Caffo, B.S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478 https://doi.org/10.1198/016214506000000492

[33] Kayalar, S. and Weinert, H.L. (1988). Error bounds for the method of alternating projections. *Math. Control Signals Systems* **1** 43–59. MR0923275 https://doi.org/10.1007/BF02551235

[34] Khare, K. and Hobert, J.P. (2013). Geometric ergodicity of the Bayesian lasso. *Electron. J. Stat.* **7** 2150–2163. MR3104915 https://doi.org/10.1214/13-EJS841

[35] Levine, R.A. and Casella, G. (2008). Comment: On random scan Gibbs samplers. *Statist. Sci.* **23** 192–195.

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

[37] 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. MR1279653 https://doi.org/10.1093/biomet/81.1.27

[38] Liu, J.S., Wong, W.H. and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* **57** 157–169. MR1325382

[39] Marchev, D. and Hobert, J.P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *J. Amer. Statist. Assoc.* **99** 228–238. MR2054301 https://doi.org/10.1198/016214504000000223

[40] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201

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

[42] Roberts, G.O. and Polson, N.G. (1994). On the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **56** 377–384. MR1281941

[43] Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981

[44] Roberts, G.O. and Rosenthal, J.S. (1998). Two convergence properties of hybrid samplers. *Ann. Appl. Probab.* **8** 397–407. MR1624941 https://doi.org/10.1214/aoap/1028903533

[45] Roberts, G.O. and Rosenthal, J.S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413 https://doi.org/10.1111/1467-9469.00250

[46] Roberts, G.O. and Rosenthal, J.S. (2016). Surprising convergence properties of some simple Gibbs samplers under various scans. *International Journal of Statistics and Probability* **5** 51–60.

[47] Roberts, G.O. and Sahu, S.K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 https://doi.org/10.1111/1467-9868.00070

[48] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158 https://doi.org/10.1093/biomet/83.1.95

[49] Roberts, G.O. and Tweedie, R.L. (2001). Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *J. Appl. Probab.* **38** 37–41. MR1915532 https://doi.org/10.1239/jap/1085496589

[50] Robertson, N., Flegal, J.M., Vats, D. and Jones, G.L. (2021). Assessing and visualizing simultaneous simulation error. *J. Comput. Graph. Statist.* **30** 324–334. MR4270507 https://doi.org/10.1080/10618600.2020.1824871

[51] Rosenthal, J.S. and Rosenthal, P. (2015). Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.* **20** no. 91, 10. MR3434208 https://doi.org/10.1214/ECP.v20-4528

[52] Roy, V. (2012). Convergence rates for MCMC algorithms for a robust Bayesian binary regression model. *Electron. J. Stat.* **6** 2463–2485. MR3020272 https://doi.org/10.1214/12-EJS756

[53] Rudolf, D. and Ullrich, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Probab.* **18** no. 49, 8. MR3078012 https://doi.org/10.1214/ECP.v18-2507

[54] Smith, R.L. and Tierney, L. (1996). Exact transition probabilities for the independence Metropolis sampler. Preprint.

[55] Tan, A. and Hobert, J.P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. MR2598033 https://doi.org/10.1198/jcgs.2009.08153

[56] Tan, A., Jones, G.L. and Hobert, J.P. (2013). On the geometric ergodicity of two-variable Gibbs samplers. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton. Inst. Math. Stat. (IMS) Collect.* **10** 25–42. Beachwood, OH: IMS. MR3586937

[57] Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. With discussion and with a reply by the authors. MR0898357

[58] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. With discussion and a rejoinder by the author. MR1329166 https://doi.org/10.1214/aos/1176325750

[59] Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031

[60] van Dyk, D.A. and Meng, X.-L. (2001). The art of data augmentation. *J. Comput. Graph. Statist.* **10** 1–111. With discussions, and a rejoinder by the authors. MR1936358 https://doi.org/10.1198/10618600152418584

[61] Vats, D., Flegal, J.M. and Jones, G.L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli* **24** 1860–1909. MR3757517 https://doi.org/10.3150/16-BEJ914

[62] Vats, D., Flegal, J.M. and Jones, G.L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306 https://doi.org/10.1093/biomet/asz002

[63] Vats, D., Robertson, N., Flegal, J.M. and Jones, G.L. (2020). Analyzing Markov chain Monte Carlo output. *Wiley Interdiscip. Rev.: Comput. Stat.* **12** e1501, 12. MR4114459 https://doi.org/10.1002/wics.1501

[64] Vidav, I. (1977). The norm of the sum of two projections. *Proc. Amer. Math. Soc.* **65** 297–298. MR0442703 https://doi.org/10.2307/2041910

[65] Wang, X. and Roy, V. (2018). Convergence analysis of the block Gibbs sampler for Bayesian probit linear mixed models with improper priors. *Electron. J. Stat.* **12** 4412–4439. MR3892344 https://doi.org/10.1214/18-ejs1506

[66] Wang, X. and Roy, V. (2018). Geometric ergodicity of Pólya-gamma Gibbs sampler for Bayesian logistic regression with a flat prior. *Electron. J. Stat.* **12** 3295–3311. MR3861283 https://doi.org/10.1214/18-EJS1481

Received June 2020 and revised January 2021